Basic Molecular Biology Techniques

5– Nucleic Acid Analysis Methods
- 5.1 DNA Blotting
- 5.2 RNA Blotting

6– Gene Probe Derivation

7– Labelling DNA Gene Probe Molecules
- 7.1 End Labelling of DNA Molecules
- 7.2 Random Primer Labelling
- 7.3 Nick Translation

8– The Polymerase Chain Reaction

5– NUCLEIC ACID ANALYSIS METHODS

There are numerous methods for analysing DNA and RNA; however, many of them are solution based or more recently include the use of chip-based array systems. Indeed, the lab-on-a-chip approach is developing rapidly and it is possible to envisage many detection and analysis methods being developed in this format in the future.[7] However, traditional methods are still employed in many laboratories and much is still made of producing a hard copy of digested and separated single stranded DNA fragments attached to a matrix such as nylon for analysis with an appropriate labelled probe.

5.1– DNA Blotting

Electrophoresis of DNA restriction fragments allows separation based on size to be carried out; however, it provides no indication as to the presence of a specific, desired fragment among the complex sample (Figure 1.5). This can be achieved by transferring the DNA from the intact gel on to a piece of nitrocellulose or nylon membrane placed in contact with it.[8] This provides a more permanent record of the sample since DNA begins to diffuse out of a gel that is left for a few hours. First the gel is soaked in alkali to render the DNA single stranded. It is then transferred to the membrane so that the DNA becomes bound to it in exactly the same pattern as that originally on the gel. This transfer, named a Southern blot after its inventor Ed Southern, can be performed electrophoretically or by drawing large volumes of buffer through both gel and membrane, thus transferring DNA from one to the other by capillary action. The point of this operation is that the membrane can now be treated with a labelled DNA molecule, for example a gene probe. This single-stranded DNA probe will hybridise under the right conditions to complementary fragments immobilized on the membrane. The conditions of hybridisation, including the temperature and salt concentration, are critical for this process to take place effectively. This is usually referred to as the stringency of the hybridisation and it is particular for each individual gene probe and for each sample of DNA. A series of washing steps with buffer are then carried out to remove any unbound probe and the membrane is developed, after which the precise location of the probe and its target may be visualised. It is also possible to analyse DNA from different species or

organisms by blotting the DNA and then using a gene probe representing a protein or enzyme from one of the organisms. In this way, it is possible to search for related genes in different species. This technique is generally termed Zoo blotting(A **zoo blot** or **garden blot** is a type of Southern blot that demonstrates the similarity between specific, usually protein-coding, DNA sequences of different species. A zoo blot compares animal species while a garden blot compares plant species. The purpose of the zoo blot is to detect the conservation of the gene(s) of interest throughout the evolution of different species).

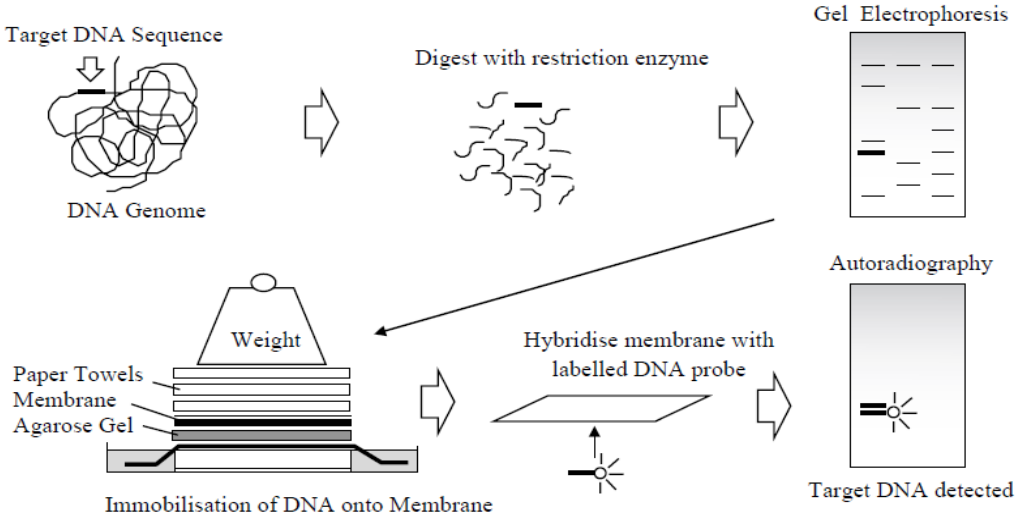

Figure 1.5 The steps involved in the production of a Southern blot and the subsequent detection of a specific DNA sequence following hybridisation with a complementary labelled gene probe.

5.2– RNA Blotting

The same basic process of nucleic acid blotting can be used to transfer RNA from gels on to similar membranes. This allows the identification of specific mRNA sequences of a defined length by hybridization to a labelled gene probe and is known as Northern blotting.[9] With this technique it is not only possible to detect specific mRNA molecules but it may also be used to quantify the relative amounts of the specific mRNA. It is usual to separate the mRNA transcripts by gel electrophoresis under denaturing conditions since this improves resolution and allows a more accurate estimation of the sizes of the transcripts. The format of the blotting may be altered from transfer from a gel to direct application to slots on a specific blotting apparatus containing the nylon membrane. This is termed slot or dot blotting and provides a convenient means of measuring the abundance of specific mRNA transcripts without the need for gel electrophoresis; it does not, however, provide information regarding the size of the fragments.

A further method of RNA analysis that overcomes the problems of RNA blotting is termed the ribonuclease protection assay. Here the RNA from a sample is extracted and then mixed with a probe representing the sequence of interest in solution. The probe and the appropriate RNA fragment hybridize to form a double-stranded sequence. RNase is then added, which cleaves any single-stranded RNA present but leaves the double-stranded RNA intact. The intact RNA can then be separated by electrophoresis and an indication of the size of the fragment generated. The efficient removal of the background of RNA and the improved

sensitivity make the ribonuclease protection assay a popular choice for the analysis of specific RNA molecules.

An important step in the field of RNA analysis was the development of RNAi (RNA interference), which inhibits gene expression. Here double-stranded DNA promotes the degradation of mRNA. Double-stranded RNA in the cell is cleaved by a dicer enzyme, resulting in the formation of small 21–25 bp interfering RNAs (siRNA). The siRNA are complementary to a target RNA strand. Small RNAi proteins are guided by the siRNA to the appropriate mRNA, where the target is then cleaved and is unable to be translated. Many areas are now benefiting from the adoption of this technique in the molecular biology and biotechnology fields.

6 GENE PROBE DERIVATION

The availability of a gene probe is essential in many molecular biology techniques, yet in many cases is one of the most difficult steps (Figure 1.6). The information needed to produce a gene probe may come from many sources, but with the development and sophistication of genetic databases this is usually one of the first stages.[11] There are a number of genetic databases throughout the world and it is possible to search these over the internet and identify particular sequences relating to a specific gene or protein. In some cases it is possible to use related proteins from the same gene family to gain information on the most useful DNA sequence. Similar proteins or DNA sequences but from different species may also provide a starting point with which to produce a so-called heterologous gene probe. Although in some cases probes have already been produced and cloned, it is possible, armed with a DNA sequence from a DNA database, to synthesise chemically a single-stranded oligonucleotide probe. This is usually undertaken by computer-controlled gene synthesisers which link dNTPs together based on a desired sequence. It is essential to carry out certain checks before probe production to determine that the probe is unique, is not able to self-anneal or is self-complementary, all of which may compromise its use.

Where little DNA information is available to prepare a gene probe, it is possible in some cases to use the knowledge gained from analysis of the corresponding protein. Thus it is possible to isolate and purify proteins and sequence part of the N-terminal end of the protein. From our knowledge of the genetic code, it is possible to predict the various DNA sequences that could code for the protein and then synthesise appropriate oligonucleotide sequences chemically. Due to the degeneracy of the genetic code, most amino acids are coded for by more than one codon, hence there will be more than one possible nucleotide sequence which could code for a given polypeptide.

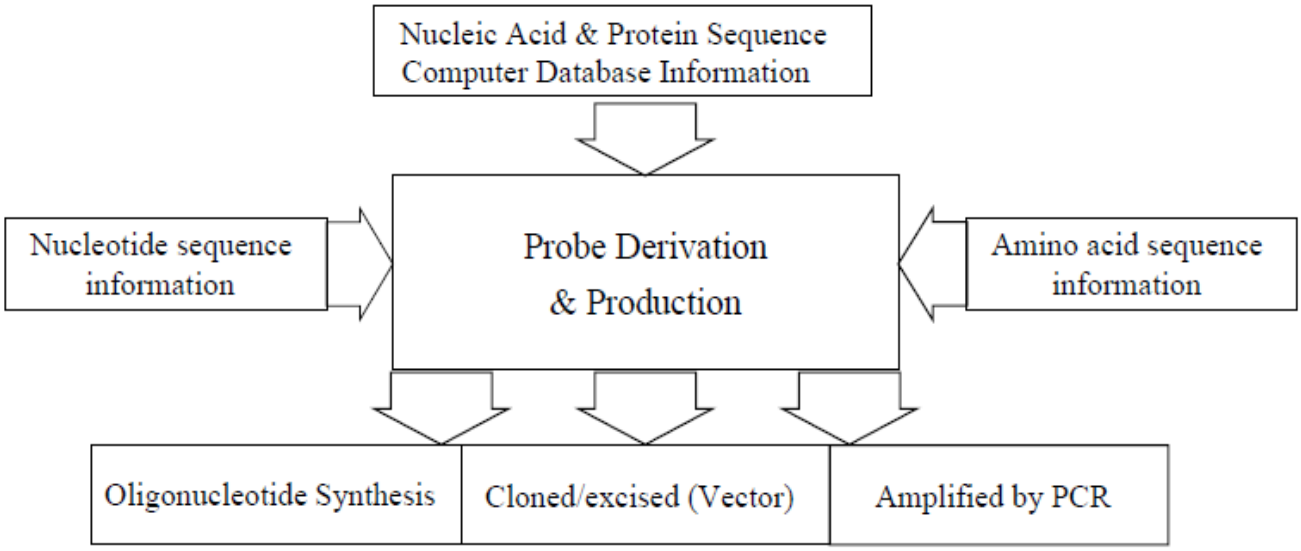

Figure 1.6 Alternative strategies for designing and producing a gene probe.

The longer the polypeptide, the greater is the number of possible oligonucleotides which must be synthesised. Fortunately, there is no need to synthesise a sequence longer than about 20 bases, since this should hybridise efficiently with any complementary sequences and should be specific for one gene. Ideally, a section of the protein should be chosen which contains as many tryptophan and methionine residues as possible, since these have unique codons and there will therefore be fewer possible base sequences which could code for that part of the protein. The synthetic oligonucleotides can then be used as probes in a number of molecular biology methods.

7– LABELLING DNA GENE PROBE MOLECULES

An essential feature of a gene probe is that it can be visualised by some means. In this way, a gene probe that hybridises to a complementary sequence may be detected and identify that desired sequence from a complex mixture. There are two main ways of labelling gene probes; traditionally it has been carried out using radioactive labels, but gaining in popularity are non-radioactive labels. Perhaps the most often used radioactive label is phosphorus-32 (^{32}P), although for certain techniques sulfur-35 (^{35}S) and tritium (^{3}H) are used. These may be detected by the process of autoradiography, where the labelled probe molecule, bound to sample DNA, located for example on a nylon membrane, is placed in contact with an X-ray-sensitive film. Following exposure, the film is developed and fixed just as a black and white negative and reveals the precise location of the labelled probe and therefore the DNA to which it has hybridised.

Non-radioactive labels are increasingly being used to label DNA gene probes. Until recently, radioactive labels were more sensitive than their non-radioactive counterparts. However, recent developments have led to similar sensitivities, which, when combined with their improved safety, have led to their greater acceptance.

The labelling systems are termed either direct or indirect. Direct labelling allows an enzyme reporter such as alkaline phosphatase to be coupled directly to the DNA. Although this may alter the characteristics of the DNA gene probe, they offer the advantage of rapid analysis since no intermediate steps are needed. However, indirect labelling is at present more popular. This relies on the incorporation of a nucleotide which has a label attached. At present, three of the main labels in use are biotin, fluorescein and digoxygenin. These molecules are covalently linked to nucleotides using a carbon spacer arm of 7, 14 or 21 atoms. Specific binding proteins may then be used as a bridge between the nucleotide and a reporter protein such as an enzyme. For example, biotin incorporated into a DNA fragment is recognised with a very high affinity by the protein streptavidin. This may be either coupled or conjugated to a reporter enzyme molecule such as alkaline phosphatase. This is able to convert a colourless substrate, p-nitrophenol phosphate (PNPP), into a yellow compound, p-nitrophenol (PNP), and also offers a means of signal amplification. Alternatively labels such as digoxygenin incorporated into DNA sequences may be detected by monoclonal antibodies, again conjugated to reporter molecules including alkaline phosphatase. Thus, rather than the detection system relying on autoradiography, which is necessary for radiolabels, a series of reactions resulting in either a colour or a light or chemiluminescent reaction takes place. This has important practical implications since autoradiography may take 1–3 days, whereas colour and chemiluminescent reactions take minutes.

7.1 End Labelling of DNA Molecules

The simplest form of labelling DNA is by 5' or 3' end labelling; 5' end labelling involves a phosphate transfer or exchange reaction where the 5' phosphate of the DNA to be used as the probe is removed and in its place a labelled phosphate, usually ^{32}P, is added. This is usually carried out by using two enzymes; the first, alkaline phosphatase, is used to remove the existing phosphate group from the DNA. Following removal of the released phosphate from the DNA, a second enzyme, polynucleotide kinase, is added, which catalyses the transfer of a phosphate group (^{32}P-labelled) to the 5' end of the DNA. The newly labelled probe is then purified, usually by chromatography through a Sephadex column, and may be used directly (Figure 1.7).

Using the other end of the DNA molecule, the 3' end, is slightly less complex. Here a new dNTP which is labelled (*e.g.* [^{32}P]adATP or biotinlabelled dNTP) is added to the 3' end of the DNA by the enzyme terminal transferase. Although this is a simpler reaction, a potential problem exists because a new nucleotide is added to the existing sequence and so the complete sequence of the DNA is altered, which may affect its hybridisation to its target sequence. End labelling methods also suffer from the fact that only one label is added to the DNA so they are of a lower activity in comparison with methods that incorporate label along the length of the DNA (Figure 1.8).

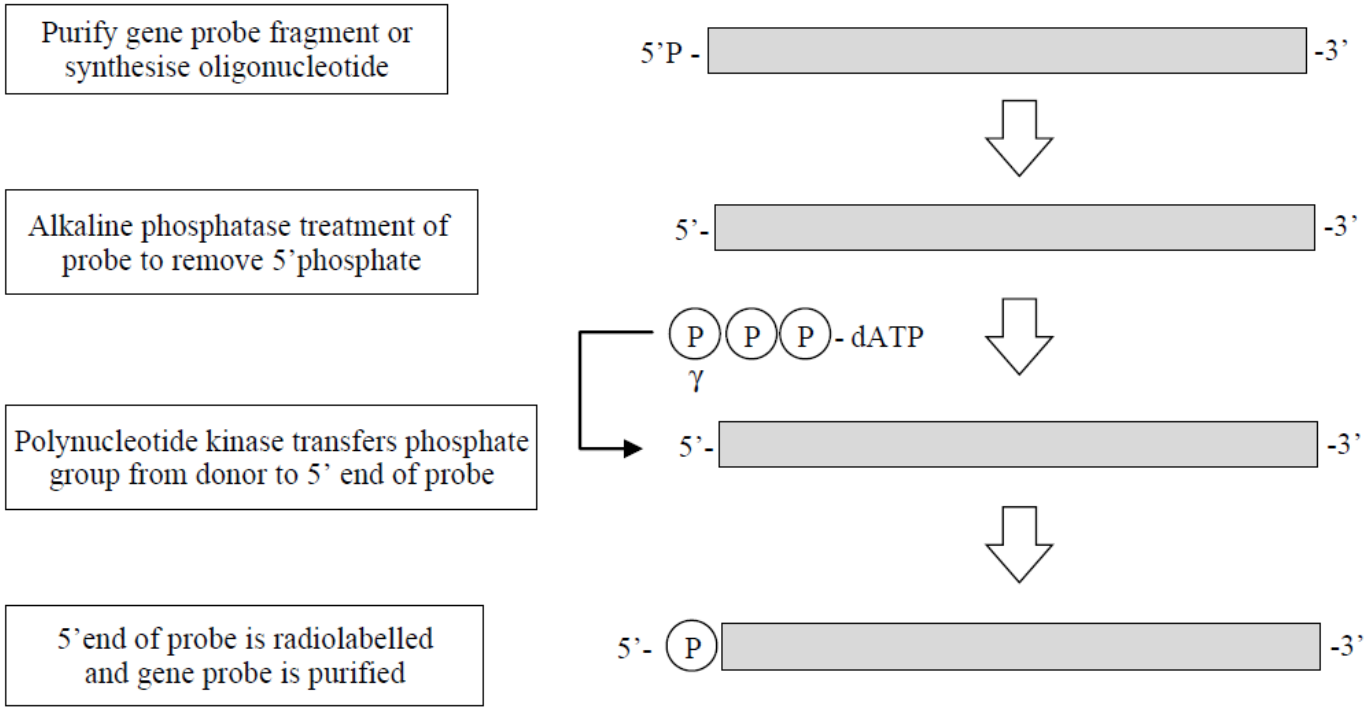

Figure 1.7 The steps involved in the production of a 5′-labelled gene probe.

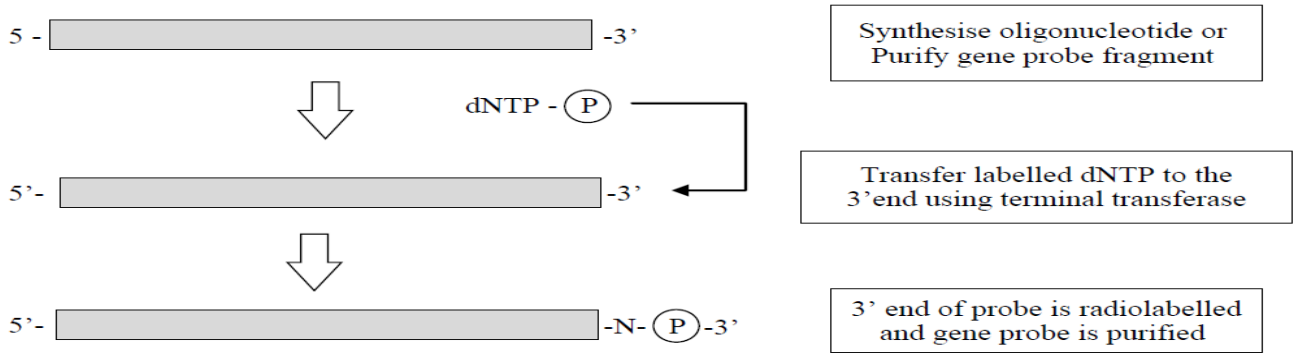

Figure 1.8 The steps involved in the production of 3′-labelled gene probe.

7.2 Random Primer Labelling

In random primer labelling the DNA to be labelled is first denatured and then placed under renaturing conditions in the presence of a mixture of many different random sequences of hexamers or hexanucleotides. These hexamers will, by chance, bind to the DNA sample wherever they encounter a complementary sequence and so the DNA will rapidly acquire an approximately random sprinkling of hexanucleotides annealed to it. Each of the hexamers can act as a primer for the synthesis of a fresh strand of DNA catalysed by DNA polymerase since it has an exposed 3'-hydroxyl group. The Klenow fragment of DNA polymerase is used for random primer labelling because it lacks a 5'–3' exonucleas activity. This is prepared by cleavage of DNA polymerase with subtilisin, giving a large enzyme fragment which has no 5' to 3' exonuclease activity, but which still acts as a 5' to 3' polymerase. Thus, when the Klenow enzyme is mixed with the annealed DNA

sample in the presence of dNTPs, including at least one which is labelled, many short stretches of labelled DNA will be generated (Figure 1.9). In a similar way to random primer labelling, the polymerase chain reaction may also be used to incorporate radioactive or non-radioactive labels.

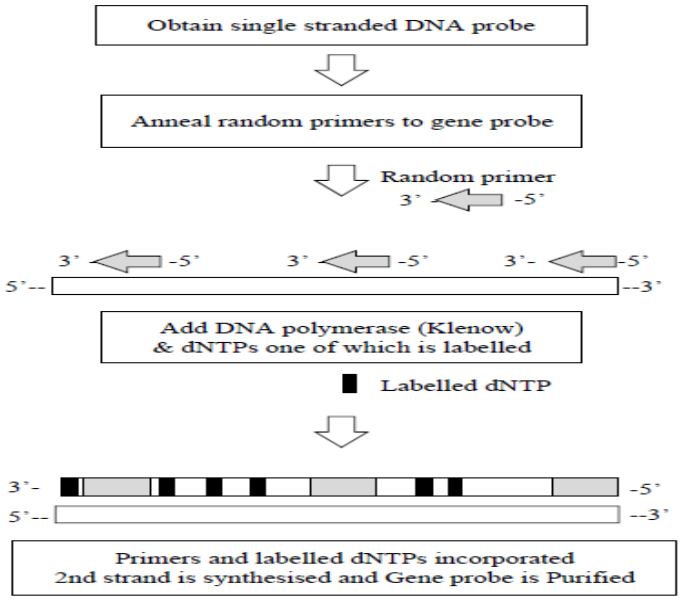

Figure 1.9 The steps involved in the production of a gene probe produced by the random hexamer method.

7.3 Nick Translation

A traditional method of labelling DNA is by the process of nick translation. Low concentrations of DNase I are used to make occasional single-strand nicks in the double-stranded DNA that is to be used as the gene probe. DNA polymerase then fills in the nicks, using an appropriate deoxyribonucleoside triphosphate (dNTP), at the same time making a new nick to the 3' side of the previous one. In this way, the nick is translated along the DNA. If labelled dNTPs are added to the reaction mixture, they will be used to fill in the nicks and so the DNA can be labelled to a very high specific activity (Figure 1.10).

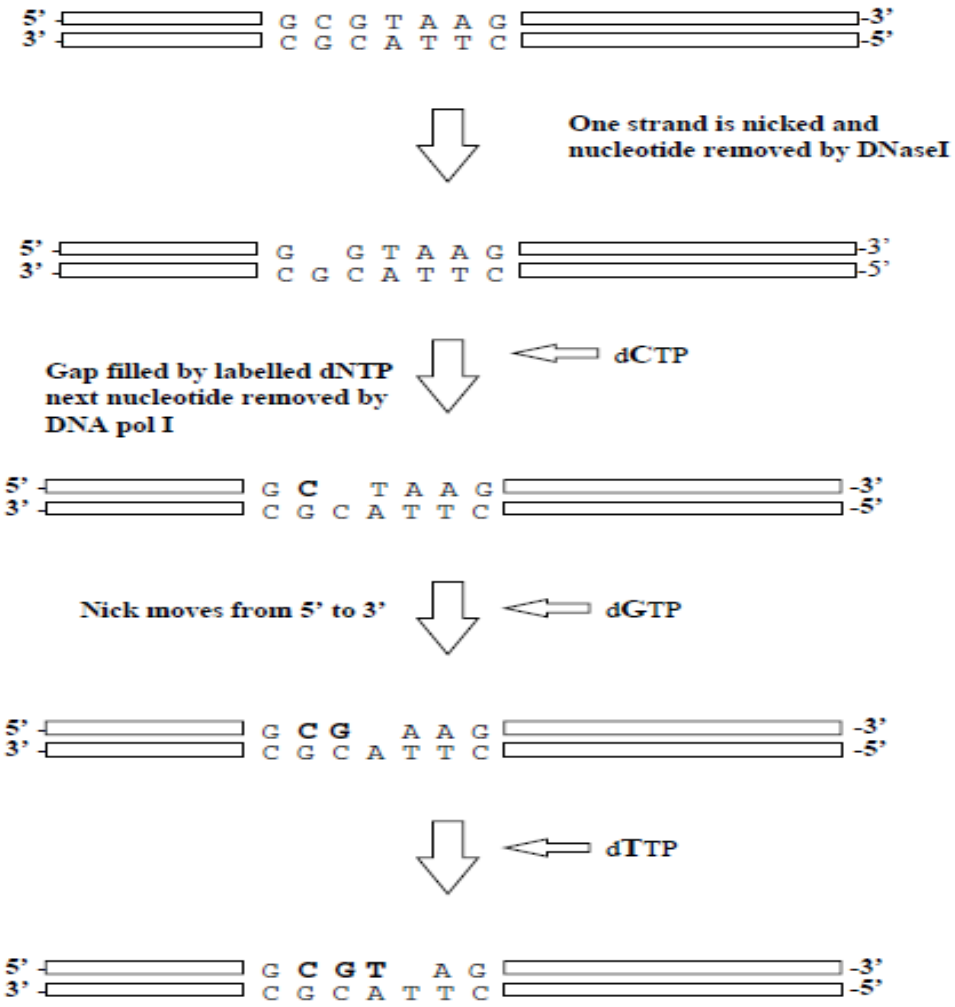

Figure 1.10 The steps involved in the production of a gene probe by the nick translation method.

8– THE POLYMERASE CHAIN REACTION

There have been a number of key developments in molecular biology techniques. However, one that has had the most impact in recent years has been the polymerase chain reaction (PCR). One of the reasons for the adoption of the PCR is the elegant simplicity of the reaction and relative ease of the practical manipulation steps. Frequently this is one of the first techniques to be used when analysing DNA and RNA and in its quantitation it has opened up the analysis of cellular and molecular processes to those outside the field of molecular biology.

The PCR is used to amplify a precise fragment of DNA from a complex mixture of starting material, usually termed the template DNA, and in many cases requires little DNA purification. It does require the knowledge of some DNA sequence information that flanks the fragment of DNA to be amplified (target DNA). From this information, two oligonucleotide primers may be chemically synthesised, each complementary to a stretch of DNA to the 3′ side of the target DNA, one oligonucleotide for each of the two DNA strands. The result is an amplification of a specific DNA fragment which obviates the need for more time-consuming cloning procedures. The technique of the PCR is described in detail in lecture 4. Further developments in molecular biology and biotechnology have allowed numerous

genomes to be analysed and genes identified. It is not surprising that this has been aided by the developments in the area of bioinformatics and whole genome analysis.[13] DNA databases and other nucleic acid sequence and protein analysis software may all be accessed over the internet given the relevant software and authority (Table 1.3).

Table 1.3 Nucleic acid and protein database resources available over the Internet.

Database or Resource		*URL (uniform resource locator)*
General DNA Sequence Databases		
EMBL	European genetic database	http://www.ebi.ac.uk
GenBank	US genetic database	http://ncbi.nlm.nih.gov
DDBJ	Japanese genetic database	http://ddbj.nig.ac.jp
Protein Sequence Databases		
Swiss-Prot	European protein sequence database	http://expasy.hcuge.ch/sprot/sprot-top.html
TREMBL	European protein sequence database	http://www.ebi.ac.uk/pub/databases/trembl
PIR	US protein information resource	http://www-nbrf.gerogetown.edu/pir
Protein Structure Databases		
PDB	Brookhaven protein database	http://www.pdb.bnl|gov
NRL-3D	Protein structure database	http://www.gdb.org/Dan/proteins/nrl3d.html
Genome Project Databases		
Human Mapping Database Johns Hopkins, USA		http://gdbwww.gdb.org
dbEST (cDNA and partial sequences)		http://www.ncbi.nih.gov
Genethon Genetic maps based on repeat markers		http://www.genethon.fr
Whitehead Institute (YAC and physical maps)		http://www-genome.wi.mit.edu

This is now relatively straightforward with web browsers that provide a user-friendly graphical interface for sequence manipulation. Consequently, the new expanding and exciting areas of bioscience research are those that analyse genome and DNA sequence databases, (genomics) and also their protein counterparts (proteomics). This is sometimes referred to as in silico research, and there is no doubt that for basic and biotechnological research it is as important to have Internet and database access as it is to have equipment and reagents for laboratory molecular biology.