Demystified . . .

The polymerase chain reaction

K R N Baumforth, P N Nelson, J E Digby, J D O'Neil, P G Murray

The polymerase chain reaction (PCR) is a technique that enables the amplification of specific sequences of nucleic acids. Although invented by Kary Mullis[1] and described originally by Saiki *et al*,[2] the principle had been described over a decade earlier by Khorana and colleagues,[3 4] and has received widespread application in many diverse fields such as forensic science, histopathology, and prenatal diagnostics. This review will attempt to explain the essentials of the basic PCR technique, including approaches to sample preparation and optimisation of the procedure. Analysis of PCR products will also be examined in detail and finally commonly used variations on the basic PCR reaction will be discussed. *In situ* PCR will not be considered because this was the subject of a previous review in this series.[5] Where appropriate, reference will be made to specific applications but these will not be considered in detail.

Basic principle of the PCR
In the PCR (fig 1), two primers (short single stranded DNAs) are used that are complementary to opposite strands of the DNA sequence to be amplified. After heat mediated denaturation of the template DNA, the primers bind to their respective sequences (annealing) on the template DNA and a DNA polymerase synthesises a complementary strand in the 5' to 3' direction (extension). Each round of denaturation, annealing, and extension is known as a cycle. Theoretically, with each cycle the amount of the template DNA sequence amplified doubles. Therefore, after 10 cycles the target sequence within the template DNA is multiplied by a factor of one thousand and after 20 cycles by a factor of more than 1 million (fig 2B). However, an "amplification plateau" will eventually be reached when additional cycles will not lead to any further increase in amplified product. This amplification plateau results from the exhaustion of reagents such as the dNTPs and primers. The optimum temperature at which each of these steps (denaturation, annealing, and extension) proceeds is different and therefore the reaction is best performed in a thermal cycler, which automatically makes the temperature changes required.

DNA extraction
A number of protocols have been devised that enable the extraction of DNA from a variety of sources including fresh tissues, peripheral blood, or other material such as hair or nail samples. The PCR can tolerate poor quality template DNA and, therefore, can also be applied to DNA extracted from archival material such as paraffin wax embedded tissues or haematoxylin and eosin stained sections.

In general, the extraction of DNA from fresh tissues[6] is achieved by incubating the sample with proteinase K[7] or simply by boiling in water.[8] Proteinase K incubations are time consuming but result in high yield, good quality DNA. The procedure involves incubating the tissue in buffer containing proteinase K for times ranging from minutes[9] up to 24 hours or more.[10] After extraction, the proteinase K is inactivated by heat treatment at 95°C for 10 minutes and the DNA purified in a number of phenol chloroform organic extraction steps. The purified DNA is then resuspended in an appropriate storage medium such as Tris-

Biomedical Research Laboratories, School of Health Sciences, University of Wolverhampton, Wolverhampton, WV1 1DJ, UK
K R N Baumforth
P N Nelson
J E Digby
P G Murray

School of Applied Sciences, University of Wolverhampton
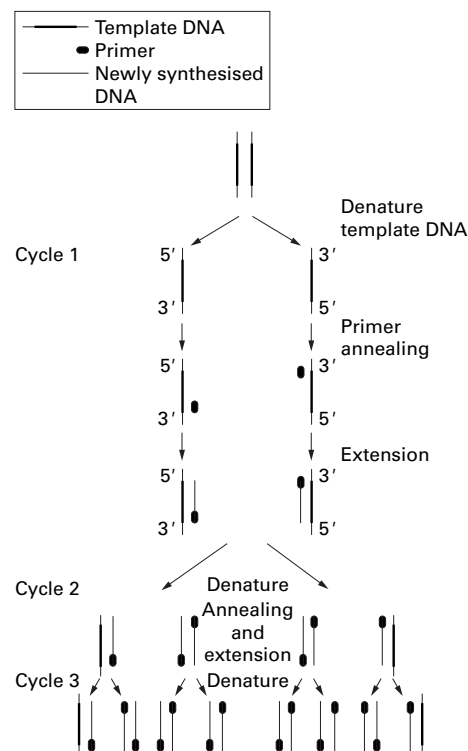J D O'Neil

Correspondence to:
Dr Murray.

Figure 1 Basic principle of the PCR. In the PCR, after initial denaturation of the template DNA two short, single stranded DNAs, known as primers, anneal to complementary strands of the DNA template and a DNA polymerase (not shown) synthesises a complementary strand in the 5' to 3' direction (extension). Each round of denaturation, annealing, and extension is known as a cycle. With each cycle the amount of the template DNA sequence doubles. Although only three cycles are shown here, most PCR reactions occur over 20–35 cycles.
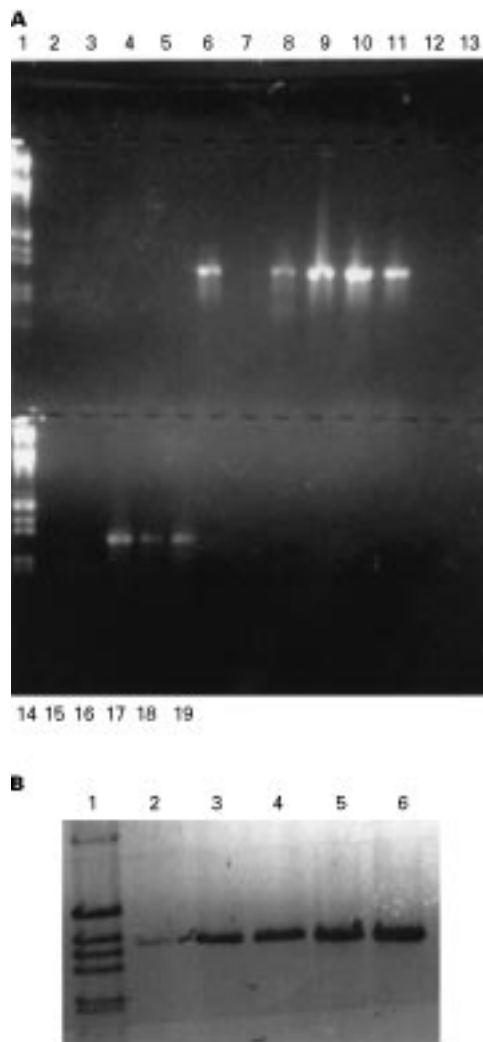
Figure 2   Optimisation of the PCR. (A) Effect of magnesium ion and template concentrations on the PCR. Ethidium bromide stained agarose gel viewed under UV light showing the effect of varying concentrations of magnesium ions and template DNA. Lanes 1 and 14, DNA ladder; lanes 2–6, increasing concentrations of magnesium ions with 1 ng of template DNA; lanes 7–11 and 15–19, the same increasing concentrations of magnesium but with template DNA concentrations of 0.1 ng and 0.01 ng, respectively; lanes 12 and 13, "no DNA" controls. (B) Effect of increasing number of PCR cycles. The effect of increasing the number of PCR cycles on the amount of amplified product is shown. The amount of PCR product increases with an increasing number of cycles up to a point referred to as the "amplification plateau". Once this point is reached, increasing the number of cycles further will not significantly affect the quantity of PCR product generated. Lane 1, DNA ladder; lanes 2–6, amount of amplified product obtained after 24, 26, 28, 30, and 32 cycles, respectively.

EDTA (ethylene diaminetetra acetic acid) buffer or simply in sterile distilled water. Storage in phosphate buffers or buffers containing high concentrations of EDTA should be avoided because of their effect on magnesium ion concentration (see later). More rapid DNA extraction is achieved by boiling the tissue in sterile distilled water for ~ 15 minutes. This approach provides DNA of sufficiently good quality for most PCR applications, although yield and quality of DNA are usually lower than for proteinase K extractions.

DNA can also be extracted successfully from archival samples including fixed, paraffin wax embedded tissue samples, stained sections, or cytological smears using either approach described above. In general, however, the extraction of DNA from archival material is less efficient than from fresh tissues,[11] and despite longer incubation times (for example, up to five days with proteinase K), both the quantity and quality of DNA obtained is usually lower than that obtained from fresh tissue.[12]

Primer design and optimisation

Successful specific amplification of the desired target sequence in the PCR is dependent upon both the design and optimal use of the primer pair. A consensus of opinion[13–17] has resulted in a simple set of rules for the design of PCR primers, namely:

- Primers should range from 15–30 bases in length
- Base composition should be 50–60% guanine + cytosine
- Long runs with more than three or four of the same base should be avoided
- Primers should not have a secondary structure (for example, hairpin loops)
- Ideally, primers should not contain sequences that are complementary to each other. This will avoid the annealing of primers to each other (primer dimer formation)
- Palindromic sequences should be avoided
- Primer melting temperatures (T_m) between 55–80°C are preferred.

The primers need only be long enough to be complementary to a unique flanking region of the DNA sequence to be amplified. The chance of a primer n bases long being present only once can be calculated using the formula 4^n. Hence, a primer 17 bases long occurs only once in 4^{17} bases (that is, 1.7×10^{10} bases). This level of specificity is much greater than that of a monoclonal antibody binding to an antigenic determinant.[16] Ideally, the primer pair should show no homology with unwanted sequences within the sample and, for this reason, primer sequences should be checked for any complementarity to other known DNA sequences by using a sequence database. In addition, each primer should have no complementarity either to itself or the other primer. It is also possible to design primers that contain 5' extensions or mismatches to allow the incorporation of restriction sites, ATG start codons, or promoter sequences into the amplified DNA sequence.

The annealing temperature used in the PCR is dependent on both the length and composition of the primers. Ideally, the annealing temperature should be between 1°C and 5°C lower than the lowest T_m value.[12] Too low an annealing temperature results in non-specific annealing and therefore non-specific amplification. Too high an annealing temperature leads to a reduced yield of product. Primer T_m can be calculated using the formula below, where n is equivalent to the length of the primer:

$$T_m = 64.9 + 0.41(\%\text{guanine} + \%\text{cytosine}) - 600/n°C$$

Table 1 Typical standard PCR components

Reagent	Recommended concentrations
Tris HCl, pH 8.3 (25°C)	10–20 mM
Magnesium chloride	1–3 mM
Potassium chloride	50 mM
Gelatin or bovine serum albumin	100 µg/ml
Primers	0.2–1 µM
Deoxynucleotide triphosphates	50–200 µM
Taq polymerase	0.5–2.5 units
Template DNA	1 ng–1 µg

Setting up and running the PCR

In the standard PCR, the template DNA is added to a reaction buffer containing deoxynucleotide triphosphates (dNTPs; which provide the energy and nucleosides necessary for DNA synthesis), a pair of primers, Taq polymerase (the DNA polymerase that catalyses the reaction), together with salts, protein and/or detergents that help to stabilise the other components (table 1). Typical reaction volumes are 50 or 100 µl. The reaction is usually carried out in 0.5 ml microcentrifuge or PCR tubes that contain the reagents outlined above. These are then placed in a thermal cycler for the required number of cycles.

PCR commences with an extended initial denaturation step (usually five minutes at 94°C), which ensures that there is complete strand separation of the template DNA. The reaction mixture is then cycled usually between 20–35 times. A typical temperature programme would be:

- Denaturation at 94°C for 20–30 seconds
- Primer annealing at $\cong$ 55°C for 20–60 seconds
- Extension at 72°C for 30–60 seconds (one minute/kilobase).

Cycling concludes with a final extension step at 72°C, usually for five minutes, ensuring that all template DNA is in double stranded conformation.

In the protocol outlined above it would appear that each cycle takes about one to three minutes; however, extra time is required between steps to heat or cool the tubes and therefore 20–35 PCR cycles typically take between two and three hours. Modern thermal cyclers switch rapidly between temperatures and therefore the use of a mineral oil layer to prevent evaporation of the reaction mixture during cycling might not always be necessary. The PCR reaction works best when amplifying target sequences between 100 and 1000 base pairs (bp) in length, although several kilobase pairs may be amplified with modification to the reaction conditions.

Reaction components

DNA POLYMERASE

The most commonly used DNA polymerase is Taq polymerase, which is isolated from a bacterium found in hot springs known as *Thermus aquaticus*, and is available from a wide range of suppliers. Taq polymerase works optimally at 72°C and over the pH range 7.0–7.5, adding ~ 100 nucleotides/second to the primer under these conditions.[14] Taq polymerase is also heat stable, allowing the enzyme to withstand repeated denaturation cycles. One drawback of Taq polymerase is its lack of 3'→5' exonuclease (proofreading) activity,[18 19] which can lead to the misincorporation of nucleotides (~ 1 in 9000 bases). Enzyme activity is also sensitive to the concentrations of magnesium and other monovalent ions (see below).

DEOXYNUCLEOTIDE TRIPHOSPHATES

dNTPs can be obtained either as freeze dried or neutralised aqueous solutions. They are also available as labelled nucleotides (radioactive or non-radioactive) and this can be useful in the subsequent hybridisation or sequencing of PCR products (see later). They are heat resistant and have a half life of more than 40 cycles of PCR.[20] Usually, each dNTP is used at a concentration between 50 µM and 200 µM—higher concentrations encourage misincorporation by the DNA polymerase. Concentrations of 50 µM and 200 µM of each dNTP are sufficient to synthesise 6.5 µg and 25 µg of DNA, respectively.

REACTION BUFFER

The buffer most often used in the PCR is 10 mM Tris buffer, with a pH range between 8.5 and 9.0 at 25°C. Because the pH of Tris buffers decreases by 0.3 units for each 10°C rise in temperature, a buffer made to pH 8.8 at 25°C will have a pH value of 7.4 at 72°C, this value being optimal for the activity of Taq polymerase at this temperature. The appropriate concentration of magnesium ions in the reaction buffer is also important for maximal Taq polymerase activity and, because the dNTPs bind magnesium ions, the reaction mixture must contain an excess of magnesium ions. As a rule of thumb, the magnesium concentration in the reaction mixture is generally 0.5–2.5 mM greater than the concentration of dNTPs. The concentration of magnesium ions also influences the efficiency of primer to template annealing (fig 2A). As a result, it is possible to modify the magnesium concentration rather than the annealing temperature to regulate primer specificity.

Other salts, for example, KCl or NaCl may help facilitate primer annealing, but concentrations in excess of 50 mM will inhibit Taq polymerase activity. Phosphate salts should be avoided because they might precipitate magnesium ions at the high temperatures used in the PCR.[14] Detergents such as Tween 20, Triton X-100, or Nonidet P-40,[13 14] and/or extra protein (for example, gelatin or bovine serum albumin) may also be added to the reaction buffer. The addition of these reagents helps to prevent precipitation of the hydrophobic Taq polymerase in aqueous solutions.

PRIMERS

In general, the primer concentration should not be greater than 1 µM. High primer concentrations promote mispriming, the formation of primer dimers, or the generation of non-specific products, thereby reducing the yield of desired product.[13] The optimal concentration of primer pairs must be ascertained for a given PCR system by titration (for example, from 0.1 µM to 1 µM). As a general rule, lower

concentrations of primers should be used with increasing concentrations of template or increasing template complexity—for example, when using human genomic DNA. Most primers are designed to be exactly complementary to the target sequence. However, in some cases, the target sequence may not be known precisely—for example, when predicting nucleic acid sequences from amino acid sequences or from the sequence of a related gene. In such cases, primers can be used that do not exactly match the template DNA. The position of these mismatches within the primer is critical to successful PCR; whereas limited mismatches at the 5' end do not significantly affect the efficiency of amplification, those at the 3' end will prevent extension by the DNA polymerase. In other situations, multiple primers can be used that take into account potential or known base changes (fig 3).

TEMPLATE DNA

Clearly, for successful PCR, the sequence to be amplified must be intact. Poor quality DNA, such as that obtained from paraffin wax embedded sections, will often contain short DNA fragments. In such cases, primers should be designed to amplify shorter regions within the template DNA where possible. Low concentrations of DNA are also required for optimal PCR because both the primers and dNTPs should be in excess; an overabundance of template DNA will favour annealing of the two strands of the template sequence, rather than their annealing to the primer pair, and will also increase the chances of forming non-specific products. Typically, a single copy gene can be amplified sufficiently in 30 cycles from less than 0.2 µg of DNA. Higher copy number genes will therefore require smaller quantities of template DNA for optimal amplification.

Validation of DNA and use of controls

Before the PCR, the test DNA should be validated by using a PCR system that detects a
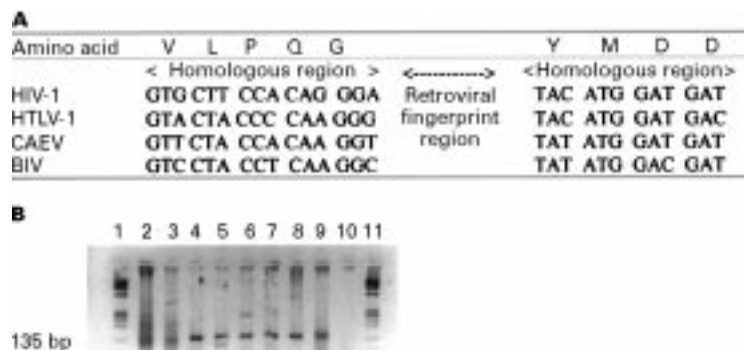
Figure 3   Use of multiple primers to allow for nucleotide variations in related sequences. Primers can be used in the PCR which allow for a limited number of mismatches. (A) Two regions of homology within the reverse transcriptase region of all human and animal retroviruses (shown here as examples are HIV-1, human T cell lymphotropic virus 1 (HTLV-1), caprine arthritis encephalitis virus (CAEV), and bovine immunodeficiency virus (BIV) sequences) shown either side of a unique "fingerprint" region. However, within these regions of homology there are minor base changes. Therefore, multiple forward: GT(ACGT)(CT)T(ACGT)CC(ACT)CA(AG)GG and reverse: (GA)TC(GA)TCCAT (GA)TA primers are designed to accommodate these nucleotide variations (where those bases in parenthesis are differences in primer sequence). (B) Amplification of this region from a variety of human and animal retroviruses using the primer sequences above yields the expected 135 bp product. Lane 1, DNA ladder; lanes 2–5, amplified sequences from HIV, HTLV-I, BIV, and CAEV, respectively; lanes 6–9, amplified regions from human endogenous retroviruses; lane 10, "no DNA" control; lane 11, a second DNA ladder.
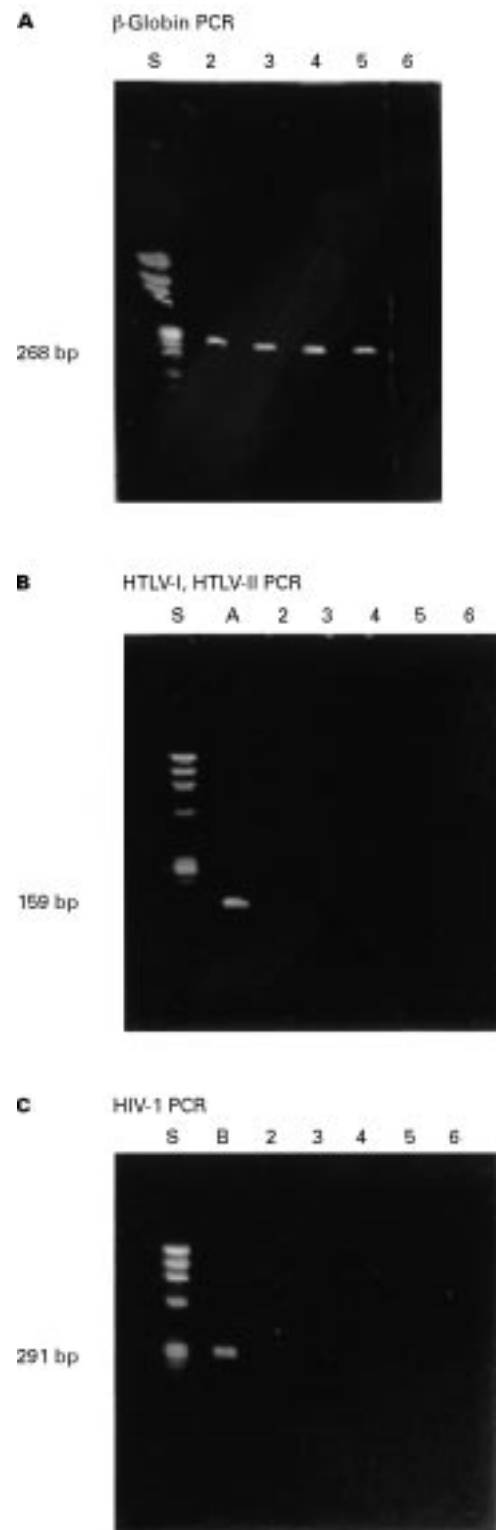
Figure 4   Validation of PCR by amplification of "housekeeping" genes. (A) Lanes 2–5, ethidium bromide stained agarose gel showing β-globin PCR of DNA extracted from peripheral blood lymphocytes of four patients. Successful extraction of DNA from all samples is demonstrated by the presence of a 268 bp product. Lane "S", DNA ladder; lane 6, "no DNA" control. (B) And (C) PCR amplification of human T cell lymphotropic virus (HTLV)-I/II and HIV-1 DNA, respectively, reveals that all patients are negative for these exogenous retroviruses. Lane 5, DNA ladder; lanes A and B, positive controls for HTLV-I/II and HIV-1, respectively.
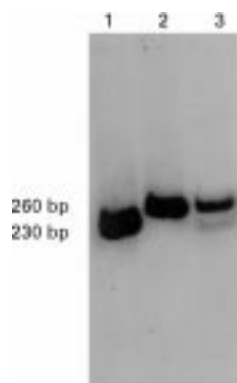
Figure 5   Differentiation of products by size using the same primer pair. It is possible to use PCR to detect additions or deletions within genes. This figure shows a Southern blot of PCR amplified products from Epstein-Barr virus (EBV) encoded latent membrane protein 1 (LMP1) DNA, followed by hybridisation to a labelled probe and autoradiography. Lanes 1 and 2, PCR products from control EBV infected cell lines AG876 and B95.8, respectively. PCR using the same primer pair yields a 230 bp product from AG876 and a 260 bp product from B95.8. This is because of the presence of a 30 bp deletion within the AG876 LMP1 gene. Lane 3, DNA from a Hodgkin's disease tumour shows both 230 bp and 260 bp products, indicating that two virus strains are present in this tumour. Photograph courtesy of CM Constandinou.

"housekeeping" gene such as β-globin (fig 4), β-actin, or glyceraldehyde-3-phosphate dehydrogenase (GAPDH). In this way, a negative PCR signal using the test primers can be considered valid, because amplification of the housekeeping sequences indicates that DNA extraction has been successful and that the DNA obtained is of sufficiently good quality.[21] All PCR runs should also include both positive and negative controls. A positive control should consist of a sample known to contain the target sequence. Negative controls consist of PCR tubes that contain all the necessary reagents except for the DNA polymerase or primers. Alternatively, negative controls can be used in which the template DNA is omitted ("no DNA" control).[22]

Avoiding contamination in the PCR

Contamination of the sample with an exogenous sequence at any stage from sample preparation to the reaction itself can result in the amplification of this contaminant sequence and a false positive result. Because PCR is so sensitive, contamination with even extremely small quantities of exogenous material can result in amplification. The contamination might be of laboratory origin (for example, previously amplified sequences) or of external origin (for example, operator DNA). To minimise the risk of contamination it is advisable to separate the PCR set up stage from the PCR analysis itself. Ideally, these two stages should be performed in separate rooms, although in some instances it may be sufficient to use separate pipettes, tips, tubes, and reagents for the pre-PCR and post-PCR stages. All reagents and consumables should be autoclaved before use.[23] Operator contamination can also be reduced by wearing suitable protective clothing.

If PCR contamination remains a problem, then UV mediated DNA crosslinking[24] might help to eliminate it. Before the addition of the template DNA, the reaction mixture can be irradiated at 254 nm. This serves to nick and crosslink any contaminating sequences and therefore make them unamplifiable. Caution should be exercised, however, because both Taq polymerase and primers can be affected by long exposures to UV. The routine exposure of PCR buffer and dNTPs to UV irradiation is not detrimental.[7]

Detection of PCR products

Once the PCR is complete the next step is to detect the presence of specific product. In general, this is achieved by agarose gel electrophoresis, which is often followed by Southern blotting and hybridisation.

AGAROSE GEL ELECTROPHORESIS

The PCR product should be of a defined length and, therefore, of a defined size. On occasions where it is only necessary to check the size of the amplified product, the simplest method is to load an aliquot of the product on to an agarose gel (typically, 1% wt/vol agarose), usually containing ethidium bromide, which is then subjected to electrophoresis. Bands on the gel corresponding to different sized DNA fragments should then be visible under UV transillumination (fig 2A). The size of the product is then estimated by comparing the product band(s) with appropriate molecular weight markers, known as a DNA "ladder" (fig 2A). DNA ladders are mixtures of DNA fragments that are loaded next to the PCR products in the agarose gel. They can either be obtained as evenly spaced DNA fragments with increments ranging between 10 bp and 1000 bp or they can be produced by digesting a piece of DNA to completion with one or more restriction enzymes (for example, lambda DNA digested with EcoRI, Hind III, or both). The primers themselves and other small products may also appear as diffuse bands close to the leading edge of the gel. Other additional bands on the gel may be the result of non-specific priming or the presence of a single stranded product.

SOUTHERN BLOTTING AND HYBRIDISATION

This was the original method of blotting invented by Ed Southern.[25][26] After electrophoresis, the agarose gel containing the PCR products is placed on a wick made from filter paper, which is in contact with a reservoir of "transfer" buffer (usually sodium citrate and sodium chloride (SSC)). The gel is then sandwiched between the wick and a hybridisation membrane. Additional filter paper layers, a pad of paper tissue, glass plate, and a weight (to keep all the layers tightly compressed) are then placed on top of the membrane. The transfer buffer is drawn upwards by capillary action through the gel, removing the DNA and immobilising it on the membrane. Nylon membranes have replaced nitrocellulose membranes as they have a greater binding capacity and are less fragile. The nucleic acid can be denatured either before or after transfer, this ensures the presence of single stranded DNA that is amenable to probing. The next step involves fixing the nucleic acids to the membrane and this is achieved by heat treatment at 80°C, or by UV crosslinking. The membrane is then placed in a solution containing labelled (radioactive or non-radioactive) nucleic acids (RNA, single stranded DNA, or oligonucleotides), known as the probe, which are complementary to sequences on the membrane. Hybridisation occurs between the labelled probe and complementary sequences on the membrane. Hybridisation conditions are chosen that maximise specific binding but minimise background caused by non-specific binding. The position of the bound probe can then be visualised—for example, by chemiluminescence or autoradiography (fig 5). Hybridisation should enable precise identification of those bands containing the target sequence of interest.

DNA sequencing by the PCR

DNA sequencing using the PCR (fig 6) follows a procedure that is very similar to standard PCR, except that the dNTPs are labelled with either a radioactive or fluorescent tag. In the case of radioactive labelling, four reaction tubes are prepared, each of which contains
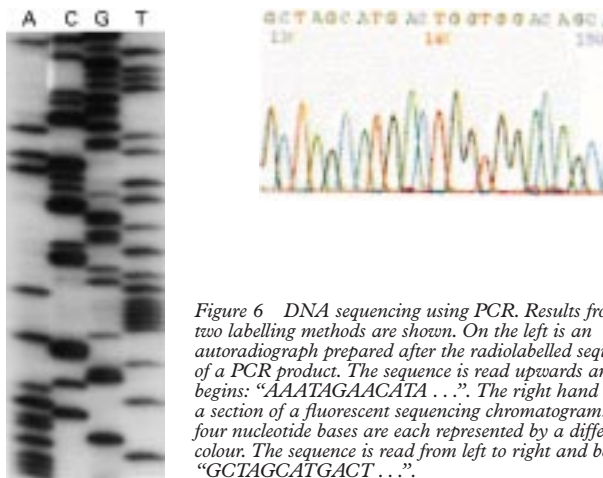
Figure 6   DNA sequencing using PCR. Results from the two labelling methods are shown. On the left is an autoradiograph prepared after the radiolabelled sequencing of a PCR product. The sequence is read upwards and begins: "AAATAGAACATA . . .". The right hand image is a section of a fluorescent sequencing chromatogram. The four nucleotide bases are each represented by a different colour. The sequence is read from left to right and begins: "GCTAGCATGACT . . .".

template, primer, Taq polymerase, all four dNTPs (one of which is labelled), and one of four dideoxy (dd)NTPs (a different one in each tube). Fluorescent labelling can be carried out in a single tube because each of the four dyes fluoresces at a different wavelength.[27] Two alternative methods can be used in fluorescent sequencing, namely:

(1) Dye termination. Any primer can be used together with four fluorescently labelled ddNTPs, it is the easiest, most robust method, and is able to cope with single or double stranded DNA. However, the specificity of the Taq polymerase results in uneven incorporation of the labelled bases.

(2) Dye primer. In this method, the primer used for chain extension is prelabelled with the fluorescent dyes. However, dye primers are expensive to synthesise and are only available for common primers, such as T7, T3, M13, etc. Therefore, before sequencing, the template has to be cloned into one of these cloning vectors. However, the method does overcome the uneven incorporation of labelled bases by Taq polymerase.

Irrespective of the labelling method used, PCR sequencing begins when the DNA template strands are separated and a single primer is allowed to anneal. The extension reaction is carried out in the presence of the dNTPs and either one or all of the ddNTPs. The ddNTPs act as extension terminators. The DNA sequences produced by the PCR reaction will share a common 5' end but have different 3' ends, depending upon which ddNTP halted extension. The various lengths of DNA are then separated by size using polyacrylamide gel electrophoresis (PAGE) and detected, by autoradiography or laser excited fluorescence, depending upon the labelling method used (fig 6). Both the autoradiograph or fluorescence data can be read in an automated fashion and stored on a computer to aid manipulation. Of the two methods outlined above, fluorescent sequencing is more easily automated and does not rely on the preparation of an autoradiograph.[28]

Variations on the standard PCR

MULTIPLEX PCR

Multiplex PCR makes use of two or more primer sets within the same reaction mix. This system is useful for the simultaneous detection of a number of different sequences—for example, two pathogenic viruses from a single DNA sample. However, the system requires careful optimisation of the PCR conditions to ensure that one PCR reaction is not dominant over the other. Furthermore, the products obtained in either reaction should also be of different sizes to enable their easy visualisation on agarose gels.

REVERSE TRANSCRIPTASE PCR (RT-PCR)

RT-PCR is the most sensitive and versatile method developed so far for the analysis of gene expression in cells and tissues. It relies on the initial conversion of RNA to complementary DNA (cDNA) using a reverse transcriptase (RT) enzyme. Most reverse transcriptases used so far have been isolated from viruses—for example, the avian myeloblastosis virus (AMV) and the Moloney murine leukemia virus (MMLV). Both enzymes work well, but AMV reverse transcriptase has the advantage that the optimum temperature for reverse transcription is 42°C, which is of benefit if the RNA template has a high degree of secondary structure.

The single stranded cDNA produced by the RT reaction is then amplified during the first cycle of the standard PCR by Taq polymerase to yield double stranded cDNA, which is then amplified in further cycles. As well as being used to detect specific RNA molecules, RT-PCR can be used to obtain cDNA for sequence determination and subsequent cloning without having to resort to constructing and screening a cDNA library. There are various strategies for obtaining the first strand cDNA using the reverse transcriptase reaction (fig 7A); either by a downstream (antisense) PCR primer annealed to the RNA, by the use of random hexamers, or by an oligo d(T) primer at the poly(A) tail of mRNA.[29] Each of these approaches has its drawbacks. For example, oligo d(T) primers that anneal to the extreme 3' end of the transcript might not efficiently synthesise cDNA from long mRNAs or those with secondary structure. The use of an antisense primer limits the subsequent PCR to a single product and will, therefore, necessitate the use of paired sense and antisense primers, whereas random hexamers will permit the cDNA template to be used for a number of independent PCRs.

Kits are now available that allow the reverse transcriptase and subsequent PCR reactions to be carried out in the same buffer and hence the same tube, thus reducing the possibility of contamination. Because of the sensitivity of RT-PCR, total cellular nucleic acid can be used, without the need to isolate poly(A) mRNA. However, cellular DNA will also be present in such preparations and amplification of this endogenous DNA is avoided by previous treatment of the sample with deoxyribonuclease.
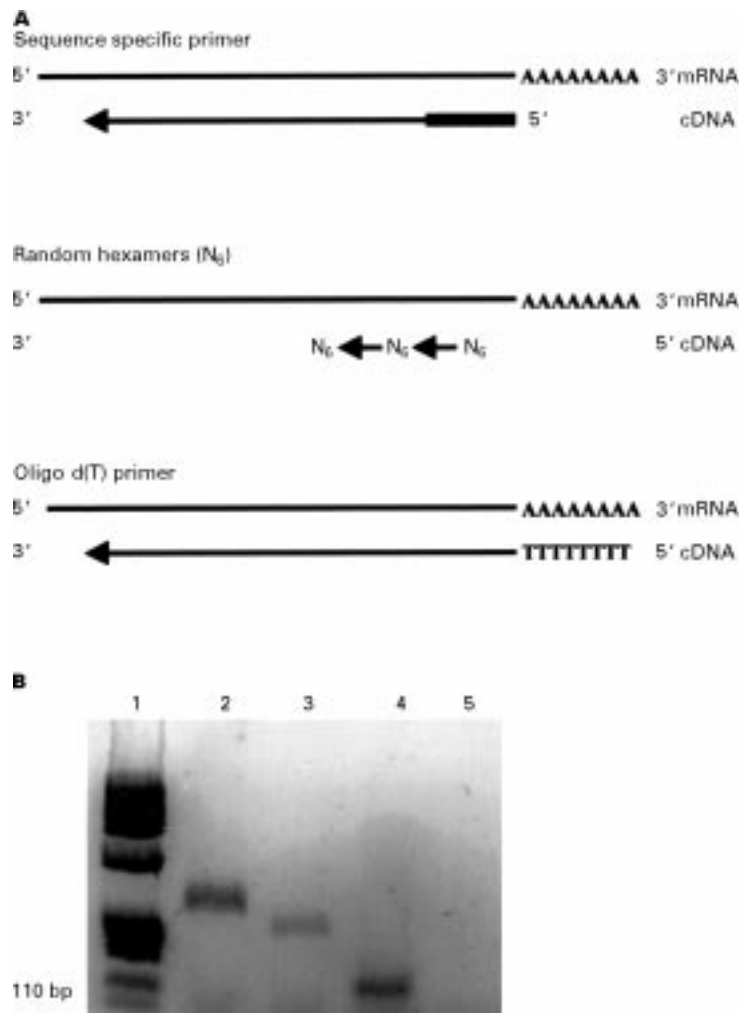
Figure 7 Reverse transcriptase PCR (RT-PCR). (A) The RT-PCR can be primed in three ways, namely: (1) by a downstream sequence specific PCR primer annealed to the RNA; (2) by the use of random hexamers; or (3) by the use of an oligo d(T) primer, which binds to the poly(A) tail of the mRNA. (B) RT-PCR using random hexamers to prime the RT reaction and subsequent PCR amplification using three primer sets: for histidyl tRNA synthetase (lane 4) and two endogenous retrovirus sequences (lanes 2 and 3). The band in lane 4 confirms the presence of a 110 bp cDNA product from this "housekeeping" gene. Lane 1, DNA ladder; lane 5, negative control.

For RT-PCR, the cDNA produced can also be validated by amplifying housekeeping genes that span single or multiple introns (non-coding DNA). For example, PCR amplification of histidyl-tRNA synthetase,[30] which spans a single intron, provides a 110 bp product for cDNA and a 360 bp product for genomic DNA (fig 7B). Consequently, this system permits both verification of the cDNA template and identification of contaminating DNA.[31] In addition to the controls used in a standard PCR, RT-PCR should also include a control in which the RT enzyme is omitted. This also ensures that the RNA preparation does not contain residual contaminating genomic DNA.

QUANTITATIVE RT-PCR
It is possible to quantify the abundance of specific mRNA molecules using multiplex and "mimic" PCR.

Multiplex PCR can also been used semi-quantitatively, to identify changes in concentrations of individual mRNA molecules,[32] relative to a housekeeping gene such as GADPH. This system relies on the premise that changes in mRNA concentrations are analysed in relation to an ostensibly stable housekeeping gene. However, a disadvantage of this approach is that the level of transcription of some house-keeping genes may also be altered.

To combat these problems, mimic systems have been developed by various commercial organisations (for example, Clonetech Laboratories, Palo Alto, California, USA). In this approach, serial dilutions of a competitive DNA fragment (the mimic) are added to constant amounts of cDNA. During PCR amplification, competition occurs between template and the mimic for a given primer set and because a known quantity of mimic is added, the concentrations of a particular cDNA and hence mRNA can be determined.[33]

The use of a synthesised internal (mutant) standard RNA that is reverse transcribed along with the sample (wild-type RNA) overcomes the problem of cDNA validation.[34] The mutant standard RNA is almost identical to the wild-type RNA, differing from it by one or two nucleotide changes that either introduce or remove a restriction endonuclease site. The amplified DNA generated by PCR can then be digested using the appropriate restriction endonuclease and the mutant standard can be distinguished after size fractionation by gel electrophoresis.

Construction of a mutant RNA standard can be achieved using cDNA that has been synthesised using random hexamers as primers. The cDNA is then amplified using a sense primer that contains a T7 polymerase promoter at the 5' end and nucleotide changes that introduce, or remove, a restriction endonuclease site at the 3' end. T7 polymerase can then be used to synthesise the mutant RNA standard from the amplified mutant cDNA template.

Quantification of the wild-type RNA is achieved by carrying out a range of serial dilutions of the mutant standard that are reverse transcribed along with the same concentration of wild-type RNA. The resultant cDNA is then amplified, and after restriction digestion, fractionated on an agarose gel (fig 8A). The relative abundance of wild-type and mutant RNA can then be assessed by densitometry. To determine the point at which the densities of the wild-type RNA and mutant RNA standard bands are equal, the log of the ratio of the mutant standard to the wild-type band densities can be plotted against the log of the mutant standard concentration (fig 8B).

NESTED PCR (nPCR)
The sensitivity of PCR can be affected adversely by the presence of poor quality or low copy template nucleic acid. Both the sensitivity and specificity of PCR can be increased by using nPCR or nRT-PCR. In nPCR, two separate amplifications are used. The first uses a set of primers that yields a large product, which is then used as a template for the second amplifi-
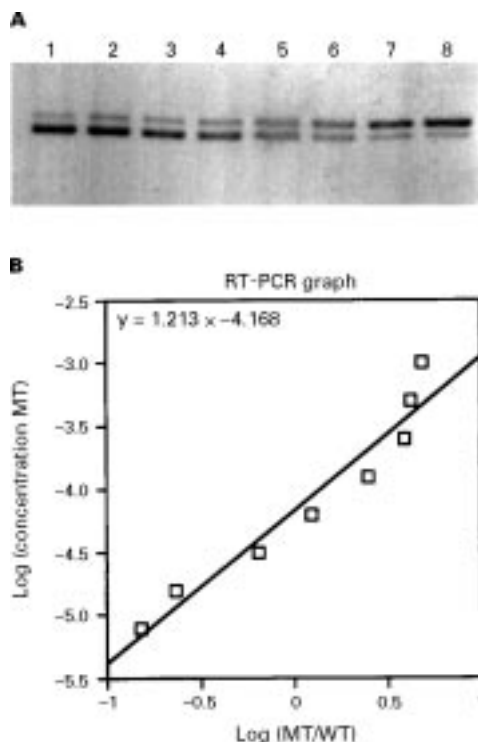
Figure 8   Quantitative reverse transcriptase PCR (RT-PCR). (A) Two products of an RT-PCR reaction, the upper band is the wild-type product and the lower band the mutant product. Lanes 1–8, increasing dilution of the mutant, from 1/1000 in lane 1 to 1/132 000 in lane 8 (that is, a twofold dilution of the mutant in each successive lane). (B) The log of the ratio of the mutant standard (MT) to the wild-type (WT) band density is plotted against the log of the mutant standard concentration to determine the point at which they are equal (x = 0). The calibration curve should be rectilinear with a slope close to 1.0 to ensure that the mutant and wild-type RNAs are amplified with equal efficiency. From the graph: the mutant concentration is equivalent to the wild-type concentration at a value of $10^{-4.168}$.

cation. The second set of primers anneals to sequences within the initial product, producing a second smaller product (fig 9). Nested PCR increases the specificity of the reaction because formation of the final product depends upon the binding of two separate sets of primers, which might preclude the need for verification of the PCR product by blotting, restriction digests, or sequencing.[35] Sensitivity is also increased because two sets of amplifications, both in the order of 25 cycles, are used.[29] As an example, PCR of the human cytomegalovirus mtr II early gene from bone marrow transplant patient buffy coat DNA was readily detected using the nPCR system but not in a single round PCR system.[31]

A variation of nPCR, known as "hot" (hot as in radiolabelled) nPCR, combines the technique described above with high resolution polyacrylamide gel electrophoresis (PAGE).[36] In hot nPCR, an end labelled primer is used in the second amplification step. End labelling of the primer is performed before the reaction, by incubation of the primer with polynucleotide kinase and $[\gamma\text{-}^{35}S]ATP$. After the second amplification step, all products from the nPCR are separated by high resolution PAGE, followed by autoradiography.
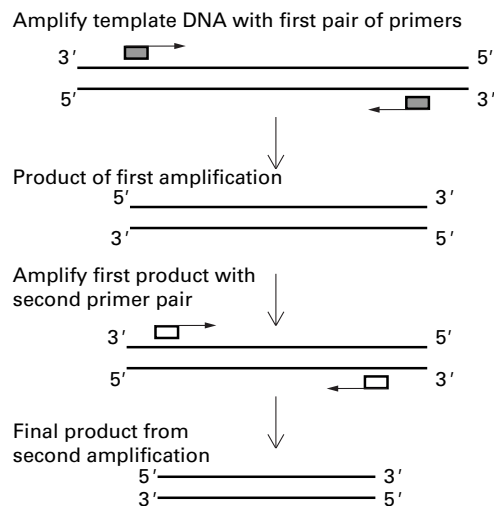

Figure 9   Principle of nested PCR. In nested PCR, two separate amplifications are used. The first uses a pair of primers yielding a large product, which is then used as a template for the second amplification. The second pair of primers anneals to sequences within the initial product, producing a second smaller product.

INVERSE PCR (IPCR)

In standard PCR, a known sequence is amplified between the primer annealing sites. However, it may be desirable to amplify sequences on either or both sides of the known region. This can be achieved using IPCR (sometimes referred to as "inside out" PCR), developed independently by three laboratories.[37–39]

The initial stage of IPCR involves digestion of the sample DNA with a restriction enzyme. The known sequence should not contain restriction sites for the chosen enzyme, so that the restriction enzyme only cuts DNA outside the known sequence. This is followed by circularisation (ligation) of the restriction fragments. A second restriction enzyme is then used to linearise these circular DNA fragments. This time, however, the second enzyme cuts within the known sequence (preferably without interfering with the primer annealing sites), producing a linear DNA fragment in which two halves of the known sequence are situated at either end (fig 10). The unknown sequence (now sandwiched between known sequences) is then amplified using the standard PCR approach.

Use of PCR in mutation analysis

PCR analysis of gene mutations is now commonplace, allowing both the nature and location of the mutation to be determined. A few of the important techniques are outlined briefly below. It should be noted that the techniques described are not an exhaustive list of those available. Others, including allele specific oligonucleotide hybridisation (ASO), density gradient gel electrophoresis (DGGE), and the protein truncation test (PTT), are described elsewhere.[40 41]

SINGLE STRANDED CONFORMATION POLYMORPHISM (SSCP) ANALYSIS

This technique is used commonly because of its simplicity and sensitivity.[42] Wild-type DNA
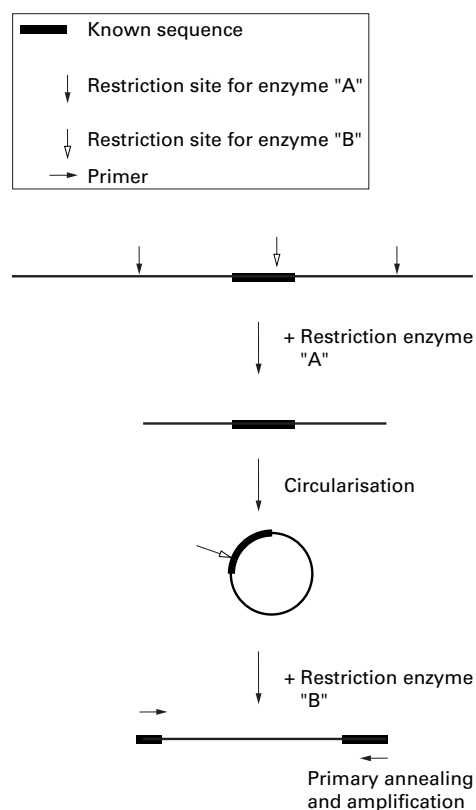
Figure 10   Inverse PCR. Inverse PCR enables the amplification of unknown sequences of DNA. The initial stage involves digestion of the sample DNA with a restriction enzyme (in this case shown as restriction enzyme "A". Restriction enzyme "A" cuts DNA at specific sites (restriction sites). The known sequence should not contain restriction sites for enzyme "A", so that only DNA outside the known sequence is cut. This is followed by circularisation (ligation) of the restriction fragments. A second restriction enzyme (shown here as enzyme "B") is then used to linearise these circular DNA fragments. This time, however, enzyme "B" cuts within the known sequence (without interfering with the primer annealing sites), producing a linear DNA fragment in which two halves of the known sequence are situated at either end. The unknown sequence (now sandwiched between known sequences) is then amplified using the standard PCR approach.

and test DNA undergo PCR and the products are denatured. The resulting single stranded products are analysed by electrophoresis on a non-denaturing gel. Single stranded DNA molecules adopt a three dimensional structure (conformation), which is dependent upon their nucleotide sequence. This in turn determines their migration properties through a gel. In other words, the presence of any mutations in the test DNA will alter its migration through the gel when compared with control wild-type DNA. The resulting bands are visualised by either incorporating a radiolabel during the PCR or by staining the gel for nucleic acids. The sensitivity of the technique in detecting mutations is ~ 70–90% for PCR fragments less than 200 bp but decreases for those above 400 bp. Once a difference between wild-type and test DNA has been established, the precise nature of the mutation present can be determined by sequencing. Therefore, SSCP analysis is a useful screen for mutations where sequencing of multiple DNA fragments would be time consuming.

HETERODUPLEX ANALYSIS

If both wild-type and mutant DNA are present in the same PCR reaction, then heteroduplexes may form. These are double stranded DNA fragments comprising one strand of the wild-type DNA and one strand of the mutant DNA. When analysed on an acrylamide gel the heteroduplexes exhibit different migrational behaviour when compared with homoduplexes composed of either wild-type or mutant DNA. This is an extremely simple technique with a similar sensitivity to SSCP.

CHEMICAL MISMATCH CLEAVAGE (CMC) AND RNase CLEAVAGE

In this technique, heteroduplexes are formed between radiolabelled wild-type DNA and mutant DNA by denaturing and reannealing of PCR products. The mismatches are then modified chemically, using osmium tetroxide for thymines and hydroxylamine for mismatched cytosines. The other mismatched bases are detected by labelling the antisense strand. Chemically modified sites can be cleaved with piperidine. Products are then analysed on a denaturing acrylamide gel and detected by autoradiography. The sensitivity of the technique approaches 100% for sequences up to 600 bp. However, the process is labour intensive and uses hazardous reagents.

RNase A cleavage uses a similar principle but has a lower sensitivity (40–60%). The heteroduplex is formed between a wild-type RNA probe and mutant (test) DNA. Mismatches formed as a result of sequence mutations are cleaved using RNase A and analysed as for CMC.

Summary

Since its initial description over twenty years ago the PCR has become one of the most valuable and flexible tools available to biomedical research. Subsequently, refinements and modifications to the basic approach, many of which have been desribed in this review, have enabled the application of the PCR to many areas of diagnostic medicine and have ensured its rapid acceptance as a routine test in many pathology disciplines. The growing importance of molecular approaches to the diagnosis of disease, particularly in histopathology, will continue to secure an ever expanding role for the PCR in diagnostic pathology.

1 Mullis K, Faloona F, Scharf S, *et al.* Specific enzymatic amplification of DNA in vitro. The polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1986;**51**:263–73.
2 Saiki RK, Scharf S, Faloona FA, *et al.* Enzymatic amplification of β globin sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* 1985;**230**:1350–4.
3 Kleppe K, Ohstuka E, Kleppe R, *et al.* Repair replication of short synthetic DNAs as catalysed by DNA polymerases. *J Mol Biol* 1971;**56**:341.
4 Panet A, Khorana HG. The linkage of deoxyribopolynucleotide templates to cellulose and its use in their replication. *J Biol Chem* 1974;**249**:5213–21.
5 Uhlmann V, Silva I, Luttich, *et al.* Demystified . . . In situ amplification. *J Mol Pathol: Clin Pathol* 1998;**51**:119–130.
6 Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning a laboratory handbook*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1982.
7 Jackson DP, Hayden JD, Quirke P. Extraction of nucleic acids from fresh and archival material. In: McPherson MJ, Quirke P, Taylor GR, eds. *PCR a practical approach*. Oxford: IRL Press, 1992:29–50.
8 Lench N, Stanier P, Williamson R. Simple non invasive method to obtain DNA for gene analysis. *Lancet* 1988;**I**: 1356–8.

9 Kawasaki ES. Sample preparation from blood, cells, and other fluids. In: Innis MA, Gelfand DH, Sninsky JJ, *et al.*, eds. *PCR protocols a guide to methods and applications*. New York: Academic Press, 1990:146–52.

10 Baginski I, Ferrie A, Watson R, *et al.* Detection of hepatitis B virus. In: Innis MA, Gelfand DH, Sninsky JJ, *et al.*, eds. *PCR protocols a guide to methods and applications*. New York, Academic Press, 1990:348–55.

11 Jackson DP, Lewis FA, Taylor GR, *et al.* Tissue extraction of DNA and RNA and analysis by the polymerase chain reaction. *J Clin Pathol* 1990;**43**:499–504.

12 Warford A, Pringle JH, Hay J, *et al.* Southern blot analysis of DNA extracted from formal saline fixed and paraffin wax embedded tissue. *J Pathol* 1988;**154**:313–20.

13 Innis MA, Gelfand DH. Optimisation of PCRs. In: Innis MA, Gelfand DH, Sninsky JJ, *et al.*, eds. *PCR protocols a guide to methods and applications*. New York, Academic Press, 1990:3–12.

14 Kocher TD, Wilson AC. DNA amplification by the polymerase chain reaction. In: Brown TA, ed. *Essential molecular biology a practical approach*, Vol. II. Oxford: IRL Press, 1991:185–207.

15 Old RW, Primrose SB. *Principles of gene manipulation. An introduction to genetic engineering*, 5th ed. Oxford: Blackwell Science Ltd, 1994:178–90.

16 Taylor GR. Polymerase chain reaction: basic principles and automation. In: McPherson MJ, Quirke P, Taylor GR, eds. *PCR a practical approach*. Oxford: IRL Press, 1992:1–14.

17 Rybicki EP. Standard PCR protocol. In: Coyne VE, James MD, Reid SJ, *et al.*, eds. *Molecular biology techniques manual*, 3rd ed. Capetown: University of Capetown, 1994. (Web based techniques manual.) Http://www.uct.ac.za/depts/microbiology/pcrcond.htm

18 Gelfand DH. Taq DNA polymerase In: Erlich HA, ed. *Principles and applications for DNA amplification* 1992:17–22.

19 Eckert KA, Kunkel TA. The fidelity of DNA polymerase used in the polymerase chain reaction. In: McPherson MJ, Quirke P, Taylor GR, eds. *PCR a practical approach*. Oxford: IRL Press, 1992:225–44.

20 Innis MA, Myambo KB, Gelfand DH, *et al.* DNA sequencing with Thermus-aquaticus DNA-polymerase and direct sequencing of polymerase chain reaction amplified DNA. *Proc Natl Acad Sci USA* 1988;**85**:9436–40.

21 Nelson PN, Lever AML, Bruckner FE, *et al.* Polymerase chain reaction fails to incriminate exogenous retroviruses HTLV-I and HIV-1 in rheumatological diseases although a minority of sera cross-react with retroviral antigens. *Ann Rheum Dis* 1994;**53**:749–54.

22 Kwok S. Procedures to minimise PCR product carry-over. In: Innis MA, Gelfand DH, Sninsky JJ, *et al.*, eds. *PCR protocols a guide to methods and applications*. New York: Academic Press 1990:142–5.

23 Kwok S, Higuchi R. Avoiding false positives with PCR. *Nature* 1990;**339**:237–8.

24 Sarker G, Sommer SS. Shedding light on PCR contamination. *Nature* 1990;**345**:27.

25 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;**98**:503–17.

26 Southern EM. Gel electrophoresis of restriction fragments. *Methods Enzymol* 1979;**68**:152–76.

27 Howe C. *Gene cloning and manipulation*. Cambridge: Cambridge University Press, 1995.

28 Wu R. Recombinant DNA. Part 1. *Methods Enzymol* 1993;**218**:3–306.

29 Kawasaki ES. Amplification of RNA. In: Innis MA, Gelfand DH, Sninsky JJ, *et al.*, eds. *PCR protocols a guide to methods and applications*. New York: Academic Press, 1990:21–7.

30 Taylor-Wiedman J, Sissons JGP, Borysiewicz LK, *et al.* Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. *J Gen Virol* 1991;**72**:2059–64.

31 Nelson PN, Rawal BK, Mathers K, *et al.* Application of reverse transcriptase polymerase chain reaction to detect a late-splice transcript of cytomegalovirus in blood specimens from bone marrow transplant recipients. *J Virol Methods* 1996;**56**:139–48.

32 Dennes WJB, Slater DM, Bennett PR. Nitric oxide synthase mRNA expression in human fetal membranes: a possible role in parturition. *Biochem Biophys Res Commun* 1997;**233**:276–8.

33 Marriott JB, Cookson S, Carlin E, *et al.* A double-blind placebo-controlled phase II trail of thalidomide in asymptomatic HIV-positive patients: clinical tolerance and effect on activation markers and cytokines. *AIDS Res Hum Retroviruses* 1997;**13**:1625–31.

34 Montague CT, Prins JB, Sanders L, *et al.* Depot- and sex-specific differences in human leptin mRNA expression. *Diabetes* 1997;**46**:342–7.

35 Rees HE, Nelson PN, Dalziel M, *et al.* Investigations to detect Borrelia burgdorferi DNA in the synovial fluid of patients with an undiagnosed large joint arthritis. *Ann Rheum Dis* 1993;**52**:404.

36 Simmonds P, Balfe P, Peuthere JF, *et al.* Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear-cells and at low copy numbers. *J Virol* 1990;**64**:864–72.

37 Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. *Genetics* 1988;**120**:621–3.

38 Silver J, Keerikatte J. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J Virol* 1989;**63**:1924–8.

39 Triglia T, Peterson MG, Kemp DJ. A procedure for in-vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* 1988;**16**:8186.

40 MacDonald F, Ford CHJ. *Molecular biology of cancer*. Bios Scientific Publishers, 1997:183–208.

41 Grompe M. The Rapid detection of unknown mutations in nucleic acids. *Nat Genet* 1993;**5**:111–17.

42 Pan LX, Diss TC, Isaacson PG. The polymerase chain reaction in histopathology. *Histopathology* 1995;**26**:201–17.