

HOW TO UNDERSTAND IT



Polymerase chain

Karen E. Morrison

Department of Clinical Neurosciences, The Medical School, University of Birmingham, Edgbaston, Birmingham; E-mail: k.morrison@bham.ac.uk
Practical Neurology, 2, 288–293

*Terms used throughout this paper are defined in the Glossary.

© 2002 Blackwell Science Ltd

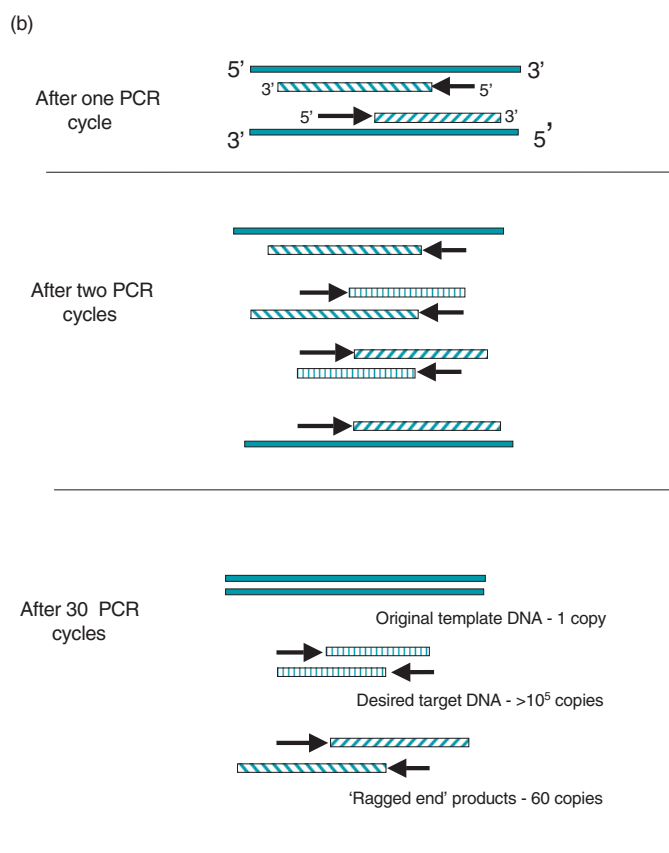
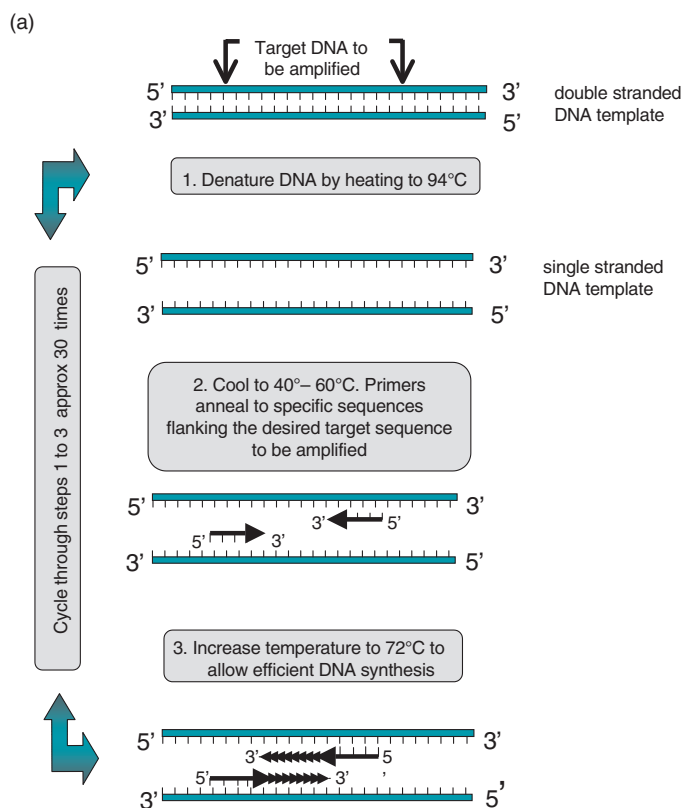
The polymerase chain reaction (PCR) is the single most important tool in biological sciences that has been developed in the last 20 years, with numerous applications in research and diagnosis. It is a test-tube method to selectively copy specific target DNA sequences – within a given source of DNA – and was ‘invented’ in 1986 following the identification of a specific DNA polymerase* that can withstand repeated heating to high temperatures. It came into widespread laboratory practice when efficient thermal cycling machines became available.

COMPONENTS OF THE PCR

The five components of the basic PCR are:

- 1 A source of template DNA containing the target sequence to be amplified (Fig. 1). This template can be either total genomic DNA containing exons* and introns*, or complex cDNA* mixtures.
- 2 Two short single-stranded sequences of DNA termed ‘primers’ or ‘amplimers’, designed on the basis of previous knowledge of the DNA target sequence. They flank the region of interest to be amplified. The ‘upstream’ primer has a sequence complementary to the se-

Figure 1 Schematic outline of PCR. (a) Outline of temperature cycling in PCR. Step 1: double stranded DNA template containing the target sequence of nucleotides for amplification is denatured by heating to 94 °C. Step 2: on cooling to between 40 °C and 60 °C, primers flanking the target sequence anneal to complementary single strands of the DNA template. Step 3: on raising the temperature to 72 °C nucleotides are added to the 3' end of the primers, according to the sequence of the single stranded DNA template. Temperature cycling steps 1–3 are repeated approximately 30 times. Oligonucleotide primers are denoted by single filled arrows, the newly synthesized DNA strands by arrowheads. (b) Outline of products obtained using PCR. After one cycle, two new single stranded DNA strands are formed that extend beyond the target DNA sequence for amplification. These 'ragged ended' strands are denoted by diagonal stripes. After two cycles, six new strands are formed, four of which are 'ragged ended' and extend beyond the ends of the target sequence and two of which are the desired target sequence, denoted by vertical stripes. After 30 cycles, the desired target sequence has been exponentially amplified $> 10^5$ times, because amplified strands act as templates for further reaction cycles. There are additionally 60 'ragged ended' strands in the final reaction product due to two amplified strands (from forward and reverse primers) extending beyond the target region in each amplification cycle.



reaction

quences upstream (5 prime: 5') of the target region, and the reverse primer has a sequence complementary to the opposite DNA strand downstream (3 prime: 3') of the target DNA sequence. Usually primers are between 15 and 25 oligonucleotides long.

- 3 A suitable heat-stable DNA polymerase*.
- 4 DNA precursors, in the form of the four deoxynucleotide triphosphates* dATP, dCTP, dGTP, dTTP (A:adenosine; C:cytidine; G:guanosine; T:thymidine).
- 5 Suitable buffers containing optimum concentrations of salts such as magnesium and potassium.

PCR is a reliable,
quick and simple
way of amplifying
large quantities
of target DNA
sequences

TEMPERATURE CYCLING FOR PCR

The reaction components are mixed and then subjected to repeated and sequential cycles of temperature variation. The first step is to heat the reaction mixture to around 94 °C. This is termed 'denaturation', during which the previously double-stranded DNA molecules melt apart into single-stranded molecules. In the second step, the temperature is reduced, typically to between 40 °C and 60 °C, when the single stranded oligonucleotide primers* anneal to their complementary sequences in the denatured, and now single-stranded template. In the third stage of the thermal cycling, the temperature is raised to 72 °C, the temperature at which the DNA polymerase is most efficient at adding deoxynucleotide triphosphate* residues to the 3' end of the growing DNA strands. The DNA polymerase adds residues that are complementary to those in the single-stranded template, thus ensuring that the sequence of the newly formed DNA strand is exactly complementary to that of the template.

The three stages of denaturation, primer annealing and DNA strand extension are then repeated approximately 25–40 times. With such repeated temperature cycling, the desired target sequence is amplified exponentially, and over 10⁵ copies are usually obtained. It is this ability to rapidly and simply produce huge quantities of highly specific target DNA for further use (such as DNA sequencing, mutation detection or therapeutic purposes) that gives PCR its supreme advantage over the other main method of target DNA amplification, namely cloning into bacteria.

Because the repeated cycles of denaturation, annealing and extension are the key to PCR, it is clear just how crucial it was finding the heat-stable DNA polymerases that could repeatedly withstand high temperatures. The widely used DNA polymerase enzyme, Taq polymerase, is obtained from the microorganism *Thermus aquaticus*, found in hot springs. The DNA polymerases that occur naturally in man are rapidly denatured by heating to 94 °C and would have to be repeatedly added to the reaction mix after each denaturation step.

ADVANTAGES OF PCR

PCR is a reliable, quick and simple way of amplifying large quantities of target DNA sequences. A typical PCR reaction takes around 15 min to set up, and the temperature cycling for 30

repeated cycles in an automated thermal cycler takes about 3 h. Time is obviously required for designing the specific primer sequences, although computer programs are available to help with this, and the primers themselves are usually commercially synthesized within a couple of days of ordering. The whole PCR process lends itself to automation and is widely used in research and diagnostic laboratories.

PCR ALLOWS VERY SPECIFIC AMPLIFICATION TO OCCUR

With appropriately-designed primer sequences and choice of temperature cycling steps, the target sequence amplification is highly-specific – only the required sequences are amplified. Primer sequences are generally chosen to be between 15 and 25 oligonucleotides long. The chance of a spurious match between a primer and some unintended sequence elsewhere in the genome (total 3 × 10⁹ base pairs) is very low. As is evident from Fig. 1, even if one of the primers does bind spuriously to a target other than the one intended, the newly-formed strand resulting from this priming will not contain the complement of the other primer in subsequent amplification cycles. Thus there is no exponential amplification of the misprimed, erroneous target. Primer sequences are usually chosen to avoid repetitive DNA sequences, and designed so that forward and reverse primers have a roughly similar adenosine and thymidine, and guanosine and cytidine oligonucleotide content, so their optimum annealing temperature (which depends on their sequence) with their DNA targets is similar. Further factors to be taken into account in designing primers ensure that their 3' ends do not themselves overlap, because this can result in much spurious amplification of small molecular weight fragments (so-called primer dimers*) and less amplification of the correct target sequence.

PCR IS AN EXQUISITELY SENSITIVE TECHNIQUE

The great advantage of PCR is the ability to amplify even minute quantities of starting target DNA. It is this property that has been exploited for amplifying even the DNA from a single cell, and in forensic science and molecular palaeontology studies. While high sensitivity is clearly an advantage in these instances, this does mean that the technique is very prone to error if the target DNA source is contaminated with an extraneous source of the same DNA. One has to be

particularly mindful of this when relying on the technique diagnostically, for example in the case of PCR-based tests for the presence of specific viral sequences in CSF samples. To interpret the results, one must know the sensitivity and specificity levels for a given diagnostic PCR assay in the individual service laboratory.

PCR IS A VERY ROBUST METHOD

PCR relies on the presence of target DNA sequences containing the sequence of interest to be amplified. However, the DNA does not have to be extracted intact. Thus DNA sequences can be amplified by PCR from tissue that has been degraded or indeed from formalin-fixed tissue samples such as stored muscle biopsies, or blood spots on Guthrie cards, sources from which other DNA cloning methods are not successful.

DISADVANTAGES OF PCR

The main limitation to the use of PCR is that one has got to know the sequences flanking the target DNA region of interest so that specific primer sequences can be made. In practice, this absolute requirement for precise DNA sequence information can sometimes be circumvented. For example, if we know the protein amino acid sequence flanking the area of interest in the whole protein, then we can work out all the corresponding potential DNA sequences. From these, degenerate oligonucleotide primers can be made, encoding all possible codons* for the flanking amino acids. Then, with appropriate PCR cycling conditions (for example see 'touch-down PCR' below), the correct gene sequence can be amplified. Alternatively, if we are interested in cloning families of similar genes or proteins, we can design degenerate oligonucleotide primers based on the sequence of known genes or proteins from various species and again, with appropriate cycling parameters, amplify desired DNA targets in the absence of prior specific sequence information. PCR with highly degenerate primer sequences has been recently exploited in 'whole genome amplification'. In this method, highly degenerate random primer sequences are used to amplify essentially all the target sequences in the starting DNA. This is of value if one is limited in the amount of starting DNA available, for example in just a single blood sample from a patient to be used in extensive genotyping* studies.

Another disadvantage of PCR is that the commonly used DNA polymerase enzyme Taq polymerase, is not error free, and will occasion-

ally misincorporate an incorrectly paired DNA base. Such incorrect DNA strands will serve as templates in future rounds of target amplification, and themselves be exponentially amplified. The level of such misincorporation depends on which specific enzyme is used, but is sufficiently frequent that the final product of a PCR reaction usually contains a mixture of extremely similar, but not identical, DNA sequences. Attempts are being made to overcome this difficulty by identifying DNA polymerase enzymes that have associated 3'-5' exonuclease* activity, able to check and correct mispaired sequences.

A further disadvantage of PCR for some applications is the requirement that the flanking primers should not be too far apart. As yet the available thermostable DNA polymerases are inefficient at amplifying DNA fragments of more than around 3000–5000 base pairs. The average length of an exon* in the human genome is around 200 base pairs, so in practice this size restriction is not greatly limiting for many current applications.

MODIFICATIONS OF THE BASIC PCR

Many modifications of the basic PCR have been developed, most of them aiming to further increase the specificity of the target DNA amplification. Nested primer PCR involves taking some of the product of an initial PCR and then using a new pair of primers in a second reaction. The primers for this second reaction are designed to bind close to the binding sites of the original PCR primers, but internal to them. In this way even small amounts of the original target sequence will be selectively amplified in the second amplification reaction.

Hot-start PCR refers to a technique in which the DNA polymerase enzyme is only added to the rest of the reaction mix once the template DNA strands have been completely denatured. Denaturing of the double stranded template DNA is not an instantaneous phenomenon. The reaction mixes are usually set up on ice and then placed in a thermal cycler, where they are heated to 94 °C over approximately 15 s. During this ramping time there is the opportunity for spurious, non-specific binding of the primers to sites in the partially denatured template. The chances of such spurious amplification can be reduced by having the enzyme (and its salt buffer solution) separated from the rest of the reaction mix by a wax plug. Only when the reaction mix is heated sufficiently does the wax melt and the

PCR relies on the presence of target DNA sequences containing the sequence of interest to be amplified

enzyme and buffer come into contact with the rest of the reaction components, increasing the chance of correct specific target amplification.

Touch-down PCR is a further modification designed to increase specificity of the target amplification. It relies on the fact that if a high annealing temperature is chosen, while the absolute amount of primer template paired strands is low as a proportion of all paired strands, the exactly matched species will predominate over those with mismatches. By starting with such very stringent annealing temperatures for the first few cycles, fewer spurious products are produced. In subsequent cycles the annealing temperature can be reduced, as increased amounts of the correct target sequence are then available for further exponential amplification.

APPLICATIONS OF PCR (TABLE 1)

Mutation detection is an area in which PCR-based methodologies are very common. DNA mutations, such as insertions or deletions of several base pairs, can be readily identified because the size of the resulting amplified product will differ between the mutated and the normal allele*. PCR reactions can also be 'multiplexed', with several pairs of primers flanking products of different sizes being added to the same tem-

plate in a single reaction tube. Such multiplex PCR is commonly used in the identification of deletions in dystrophin in DNA from Duchenne muscular dystrophy patients, when deletions often extend across several exons. A further method of analysing DNA products following PCR amplification, other than by sizing products by electrophoresis on a gel, is by DNA sequencing. PCR-based techniques are key in such sequencing, both manually and in automated systems. Methods have also been developed to allow more rapid and cost effective identification of DNA sequence alterations in amplified DNA compared to full DNA sequencing. These methods, which include single-stranded conformation polymorphism analysis (SSCP) and denaturing high performance liquid chromatography (DHPLC), make extensive use of PCR.

Genetic linkage analysis relies on the statistical analysis of the inheritance pattern of specific genetic markers throughout families in which a particular disease is segregating. Key to such analysis is the availability of several thousand genetic markers* that allow one to determine the parent of origin of a particular stretch of DNA sequence in an offspring. PCR is extensively used in the genotyping of genetic markers, including microsatellite repeat sequences* (Fig. 2), single nucleotide polymorphisms* and restriction site fragment length polymorphisms*. Association studies are another form of genetic analysis that compares the occurrence of specific genetic markers in individuals with a particular disease to the frequency of their occurrence in individuals without the disease. The success of such studies relies on identifying and genotyping appropriate genetic markers, again making extensive use of PCR.

USE IN RESEARCH METHODOLOGIES

PCR is widely used in innumerable research methodologies. It is a key step in many of the protocols now established for DNA sequencing. It is also used to radiolabel DNA species, a widespread application in molecular biological research. Although I have concentrated on the use of PCR in amplifying DNA using DNA as a template, a further use of the reaction is in amplifying DNA from templates of cDNA derived from RNA from various species and tissues. This method, termed reverse transcriptase PCR (RT-PCR) allows amplification and subsequent analysis of the coding regions of genes, because

Table 1 Some applications of PCR

DNA mutation detection
detection of base pair substitutions
detection of small insertions and deletions
detection of large insertions and deletions
Typing genetic markers
used in genetic linkage and association studies; used in genotyping microsatellite markers, single nucleotide polymorphisms, restriction fragment length polymorphisms*
DNA sequencing
amplification of stretches of DNA for 'conventional' chemical DNA sequencing methods and for PCR-based 'cycle-sequencing' methods
Radio-labelling of DNA
extensive use in research molecular biological methodologies
Cloning of genomic DNA
used when sequence flanking target is known, or unknown using degenerate oligonucleotide primers; whole genome amplification protocols being developed; high sensitivity of the technique allows use in forensic science and palaeontology
Analysis of RNA
reverse transcription-PCR used to synthesize cDNA using RNA as template; widespread use in gene expression studies
Diagnostic uses
for example, in detecting genomic DNA sequences from specific pathogens in body fluids/tissues/pathological samples

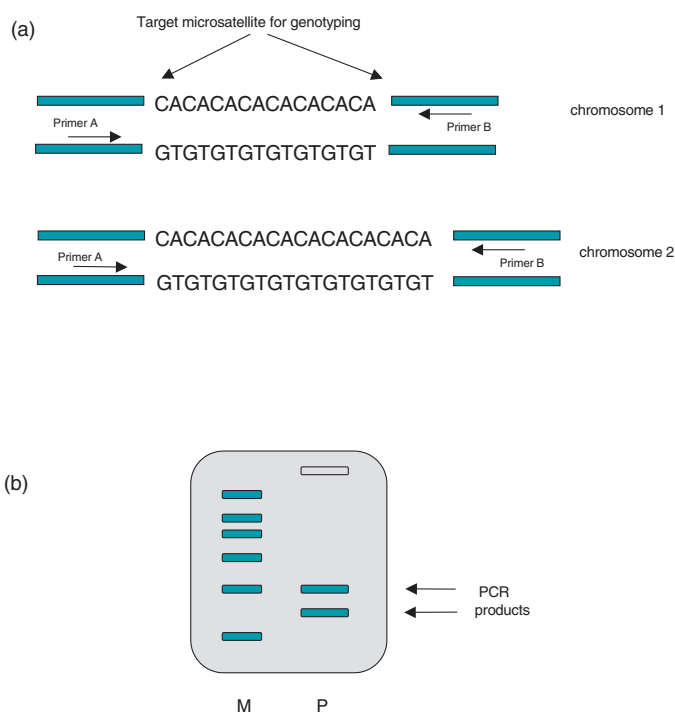


Figure 2 Use of PCR in genotyping a microsatellite marker. (a) A microsatellite is shown containing a variable number of CA (Cytidine-Adenosine) repeat units on each of two chromosomes. Primers A and B are designed to bind to specific DNA sequences flanking this microsatellite. (b) Following PCR, using a mix of DNA from chromosomes 1 and 2 as template, the products are separated according to their size by electrophoresis on an agarose gel. M denotes a DNA size marker standard. P denotes the product of the PCR, consisting of two bands that differ in length by four base pairs, reflecting the difference in number of CA repeat units flanked by the specific primers A and B.

RNA does not contain intervening intronic sequences. This is of great value if the structure of the gene (such as the intron/exon boundaries) under study is not known.

SUMMARY

PCR is a highly versatile, robust and rapid method which selectively amplifies target DNA sequences within a given source of DNA. It has broad application throughout biology, both in research and diagnostic laboratories. The widespread, and growing, uses of PCR are testament to the crucial importance of this simple reaction.

FURTHER READING

Strachan T, Read AP (1999) *Human Molecular Genetics* (1999) 2nd Edn. BIOS Scientific Publishers Ltd, Oxford.

McPherson MJ, Taylor GR, Quirke P (1991) *PCR: A Practical Approach*. IRL Press, Oxford.

GLOSSARY

allele – one of alternative forms of a gene sequence

codon – group of three adjacent nucleotides which encodes a specific amino acid

cdNA – synthetically derived DNA copied from an RNA template

deoxynucleotide triphosphate – the building block of DNA

DNA polymerase – enzyme that synthesizes DNA from the nucleotide building blocks, along a single stranded DNA template

exon – stretch of DNA sequence that determines an amino acid sequence

3'-5' exonuclease – enzyme that can remove a deoxynucleotide triphosphate residue from the 3' end of DNA

genetic marker – specific variable DNA sequence

genome – the complete DNA sequence of an organism

genotype/genotyping – the specific sequence of a DNA region known to vary between individuals/the determination of such a variable sequence

intron – stretch of DNA sequence that does not encode an amino acid sequence

microsatellite repeat sequence – stretch of repetitive DNA sequence consisting of 2–5 base pairs repeated a variable number of times

oligonucleotide primer ('oligo') – short stretch of single stranded DNA sequence composed of between 15 and 30 nucleotide residues

primer dimer – spurious, small (< 50 base pairs) amplification product due to anomalous primer self-annealing sometimes seen during PCR

restriction fragment length polymorphism – region of DNA of variable sequence identified by the ability of a bacterial enzyme (a restriction endonuclease) to specifically cleave the DNA

single nucleotide polymorphism – a particular nucleotide at a specific site in a DNA sequence, the sequence of which varies between different chromosomes