Chapter 9

The Polymerase Chain Reaction

Chapter contents

CHAPTER CONTENTS

As a result of the last seven chapters we have become familiar not only with the basic principles of gene cloning, but also with fundamental molecular biology techniques such as restriction analysis, gel electrophoresis, DNA labeling, and DNA–DNA hybridization. To complete our basic education in DNA analysis we must now return to the second major technique for studying genes, the polymerase chain reaction (PCR). PCR is a very uncomplicated technique: all that happens is that a short region of a DNA molecule, a single gene for instance, is copied many times by a DNA polymerase enzyme (see Figure 1.2). This might seem a rather trivial exercise, but it has a multitude of applications in genetics research and in broader areas of biology.

We begin this chapter with an outline of the polymerase chain reaction in order to understand exactly what it achieves. Then we will look at the key issues that determine whether or not an individual PCR experiment is successful, before examining some of the methods that have been devised for studying the amplified DNA fragments that are obtained.

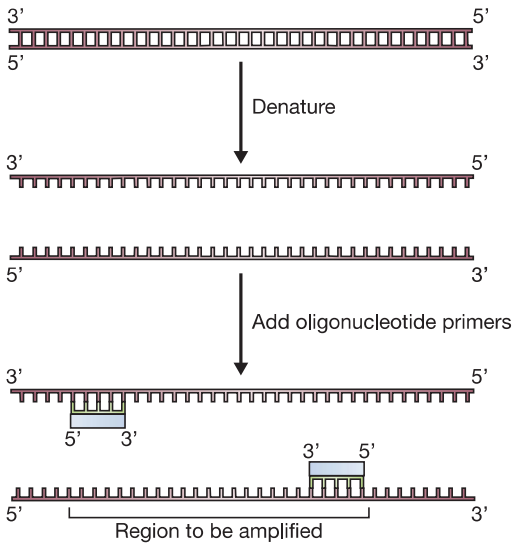
9.1 The polymerase chain reaction in outline

The polymerase chain reaction results in the selective amplification of a chosen region of a DNA molecule. Any region of any DNA molecule can be chosen, so long as the sequences at the borders of the region are known. The border sequences must be known because in order to carry out a PCR, two short oligonucleotides must hybridize to the

Gene Cloning and DNA Analysis: An Introduction. 6th edition. By T.A. Brown. Published 2010 by Blackwell Publishing.

Figure 9.1

Hybridization of the oligonucleotide primers to the template DNA at the beginning of a PCR.



DNA molecule, one to each strand of the double helix (Figure 9.1). These oligonucleotides, which act as primers for the DNA synthesis reactions, delimit the region that will be amplified.

Amplification is usually carried out by the DNA polymerase I enzyme from *Thermus aquaticus*. As mentioned on p. 49, this organism lives in hot springs, and many of its enzymes, including *Taq* polymerase, are thermostable, meaning that they are resistant to denaturation by heat treatment. As will be apparent in a moment, the thermostability of *Taq* polymerase is an essential requirement in PCR methodology.

To carry out a PCR experiment, the target DNA is mixed with *Taq* polymerase, the two oligonucleotide primers, and a supply of nucleotides. The amount of target DNA can be very small because PCR is extremely sensitive and will work with just a single starting molecule. The reaction is started by heating the mixture to 94°C. At this temperature the hydrogen bonds that hold together the two polynucleotides of the double helix are broken, so the target DNA becomes denatured into single-stranded molecules (Figure 9.2). The temperature is then reduced to 50–60°C, which results in some rejoining of the single strands of the target DNA, but also allows the primers to attach to their annealing positions. DNA synthesis can now begin, so the temperature is raised to 74°C, just below the optimum for *Taq* polymerase. In this first stage of the PCR, a set of "long products" is synthesized from each strand of the target DNA. These polynucleotides have identical 5′ ends but random 3′ ends, the latter representing positions where DNA synthesis terminates by chance.

The cycle of denaturation–annealing–synthesis is now repeated (Figure 9.3). The long products denature and the four resulting strands are copied during the DNA synthesis stage. This gives four double-stranded molecules, two of which are identical to the long products from the first cycle and two of which are made entirely of new DNA. During the third cycle, the latter give rise to "short products", the 5′ and 3′ ends of which are both set by the primer annealing positions. In subsequent cycles, the number of short products accumulates in an exponential fashion (doubling during each cycle) until one of the components of the reaction becomes depleted. This means that after 30 cycles, there will be over 130 million short products derived from each starting molecule. In real terms, this equates to several micrograms of PCR product from a few nanograms or less of target DNA.

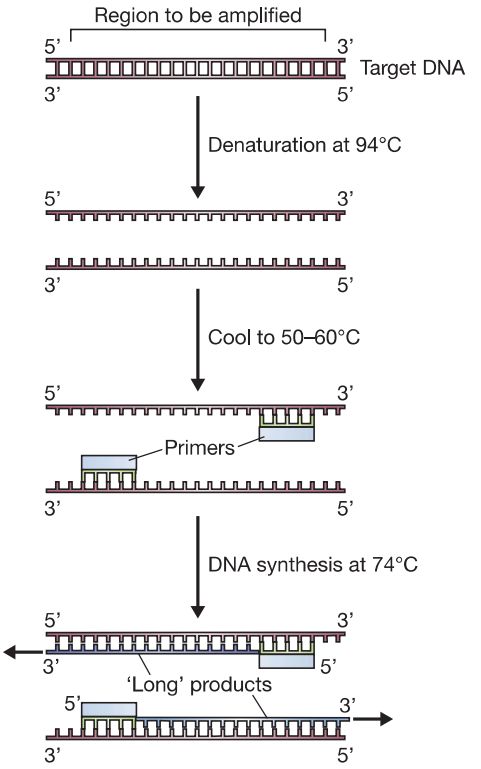


Figure 9.2
The first stage of a PCR, resulting in synthesis of the long products.

At the end of a PCR a sample of the reaction mixture is usually analyzed by agarose gel electrophoresis, sufficient DNA having been produced for the amplified fragment to be visible as a discrete band after staining with ethidium bromide. This may by itself provide useful information about the DNA region that has been amplified, or alternatively the PCR product can be examined by techniques such as DNA sequencing.

9.2 PCR in more detail

Although PCR experiments are very easy to set up, they must be planned carefully if the results are to be of any value. The sequences of the primers are critical to the success of the experiment, as are the precise temperatures used in the heating and cooling stages of the reaction cycle.

9.2.1 Designing the oligonucleotide primers for a PCR

The primers are the key to the success or failure of a PCR experiment. If the primers are designed correctly the experiment results in amplification of a single DNA fragment, corresponding to the target region of the template molecule. If the primers are incorrectly designed the experiment will fail, possibly because no amplification occurs, or possibly because the wrong fragment, or more than one fragment, is amplified (Figure 9.4). Clearly a great deal of thought must be put into the design of the primers.

Working out appropriate sequences for the primers is not a problem: they must correspond with the sequences flanking the target region on the template molecule. Each primer must, of course, be complementary (not identical) to its template strand in

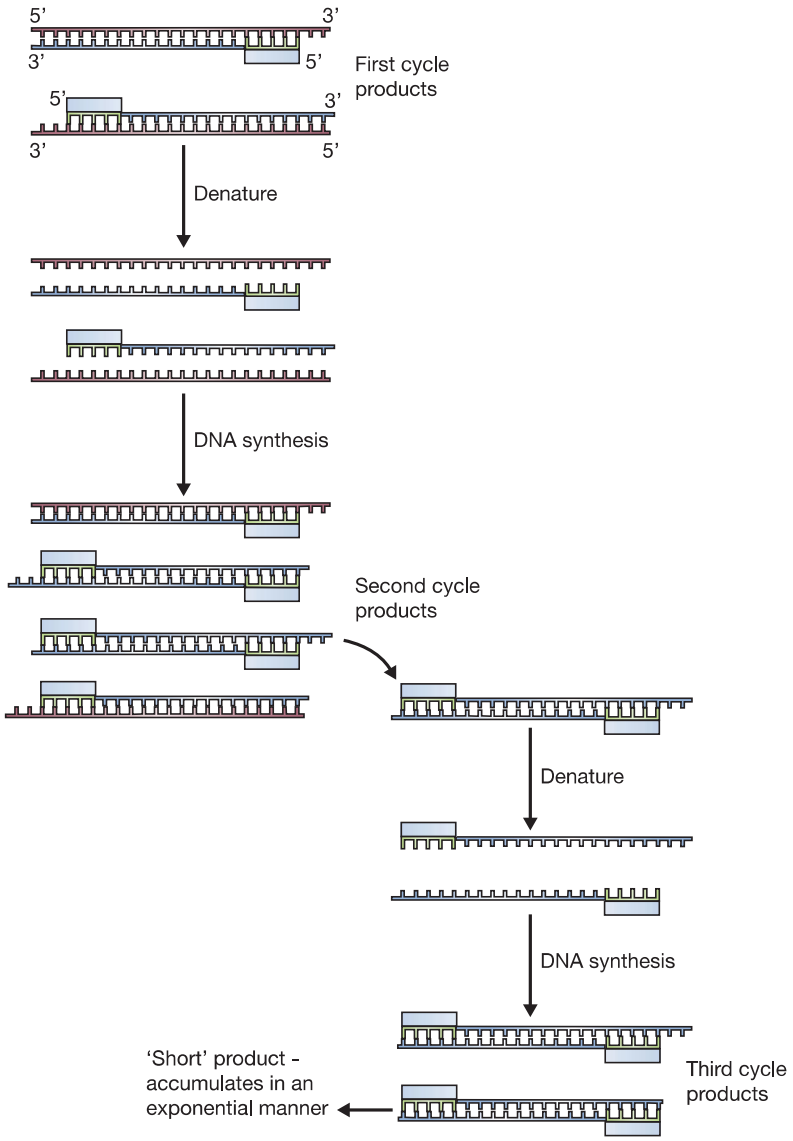
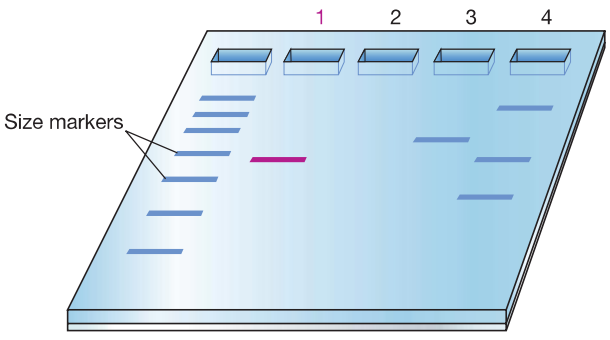


Figure 9.3

The second and third cycles of a PCR, during which the first short products are synthesized.

Figure 9.4

The results of PCRs with well designed and poorly designed primers. Lane 1 shows a single amplified fragment of the expected size, the result of a well designed experiment. In lane 2 there is no amplification product, suggesting that one or both of the primers were unable to hybridize to the template DNA. Lanes 3 and 4 show, respectively, an amplification product of the wrong size, and a mixture of products (the correct product plus two wrong ones); both results are due to hybridization of one or both of the primers to non-target sites on the template DNA molecule.



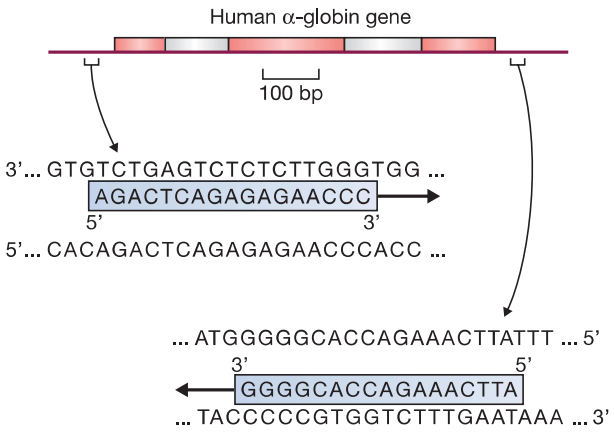


Figure 9.5
A pair of primers designed to amplify the human α_1-globin gene. The exons of the gene are shown as closed boxes, the introns as open boxes.

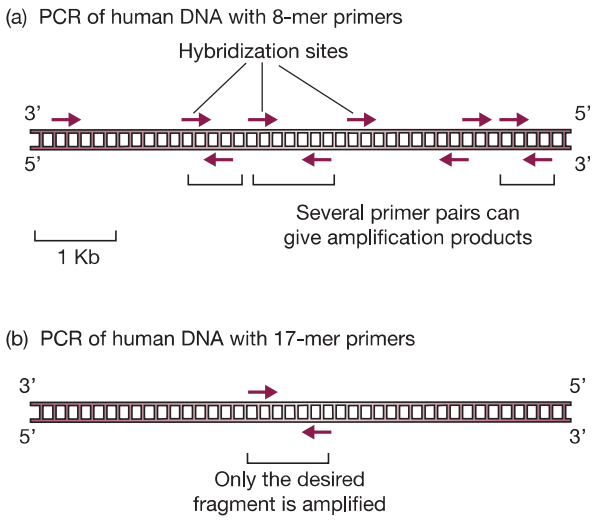


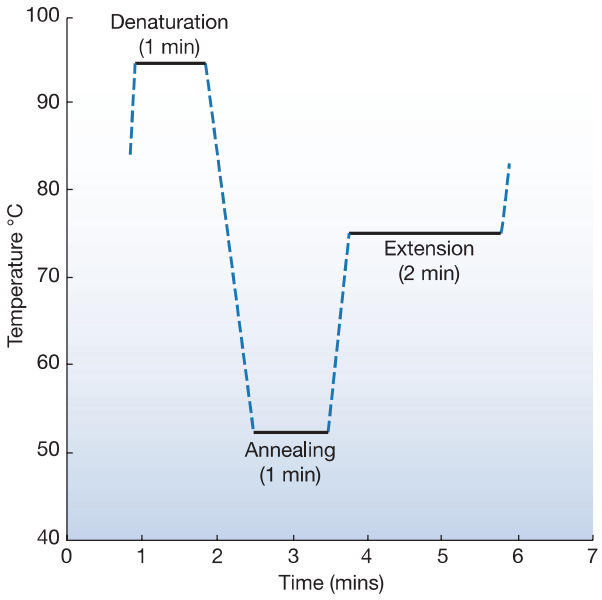
Figure 9.6
The lengths of the primers are critical for the specificity of the PCR.

order for hybridization to occur, and the 3′ ends of the hybridized primers should point toward one another (Figure 9.5). The DNA fragment to be amplified should not be greater than about 3 kb in length and ideally less than 1 kb. Fragments up to 10 kb can be amplified by standard PCR techniques, but the longer the fragment the less efficient the amplification and the more difficult it is to obtain consistent results. Amplification of very long fragments—up to 40 kb—is possible, but requires special methods.

The first important issue to address is the length of the primers. If the primers are too short they might hybridize to non-target sites and give undesired amplification products. To illustrate this point, imagine that total human DNA is used in a PCR experiment with a pair of primers eight nucleotides in length (in PCR jargon, these are called "8-mers"). The likely result is that a number of different fragments will be amplified. This is because attachment sites for these primers are expected to occur, on average, once every $4^8 = 65{,}536$ bp, giving approximately 49,000 possible sites in the 3,200,000 kb of nucleotide sequence that makes up the human genome. This means that it would be very unlikely that a pair of 8-mer primers would give a single, specific amplification product with human DNA (Figure 9.6a).

What if the 17-mer primers shown in Figure 9.5 are used? The expected frequency of a 17-mer sequence is once every $4^{17} = 17{,}179{,}869{,}184$ bp. This figure is over five

Figure 9.7
A typical temperature profile for a PCR.



times greater than the length of the human genome, so a 17-mer primer would be expected to have just one hybridization site in total human DNA. A pair of 17-mer primers should therefore give a single, specific amplification product (Figure 9.6b).

Why not simply make the primers as long as possible? The length of the primer influences the rate at which it hybridizes to the template DNA, long primers hybridizing at a slower rate. The efficiency of the PCR, measured by the number of amplified molecules produced during the experiment, is therefore reduced if the primers are too long, as complete hybridization to the template molecules cannot occur in the time allowed during the reaction cycle. In practice, primers longer than 30-mer are rarely used.

9.2.2 Working out the correct temperatures to use

During each cycle of a PCR, the reaction mixture is transferred between three temperatures (Figure 9.7):

- The denaturation temperature, usually 94°C, which breaks the base pairs and releases single-stranded DNA to act as templates in the next round of DNA synthesis;
- The hybridization or annealing temperature, at which the primers attach to the templates;
- The extension temperature, at which DNA synthesis occurs. This is usually set at 74°C, just below the optimum for *Taq* polymerase.

The annealing temperature is the important one because, again, this can affect the specificity of the reaction. DNA–DNA hybridization is a temperature-dependent phenomenon. If the temperature is too high no hybridization takes place; instead the primers and templates remain dissociated (Figure 9.8a). However, if the temperature is too low, mismatched hybrids—ones in which not all the correct base pairs have formed—are stable (Figure 9.8b). If this occurs the earlier calculations regarding the appropriate lengths for the primers become irrelevant, as these calculations assumed that only perfect primer–template hybrids are able to form. If mismatches are tolerated,

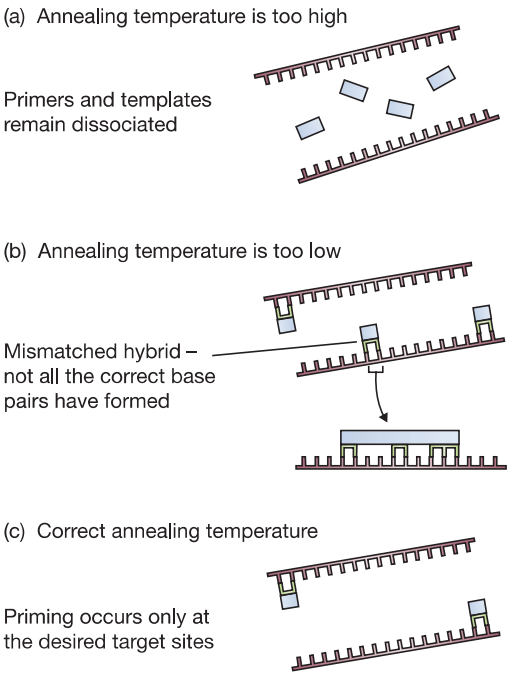


Figure 9.8
Temperature has an important effect on the hybridization of the primers to the template DNA.

the number of potential hybridization sites for each primer is greatly increased, and amplification is more likely to occur at non-target sites in the template molecule.

The ideal annealing temperature must be low enough to enable hybridization between primer and template, but high enough to prevent mismatched hybrids from forming (Figure 9.8c). This temperature can be estimated by determining the **melting temperature** or T_m of the primer–template hybrid. The T_m is the temperature at which the correctly base-paired hybrid dissociates ("melts"). A temperature 1–2°C below this should be low enough to allow the correct primer–template hybrid to form, but too high for a hybrid with a single mismatch to be stable. The T_m can be determined experimentally but is more usually calculated from the simple formula (Figure 9.9):

$$T_m = (4 \times [G + C]) + (2 \times [A + T])°C$$

in which $[G + C]$ is the number of G and C nucleotides in the primer sequence, and $[A + T]$ is the number of A and T nucleotides.

The annealing temperature for a PCR experiment is therefore determined by calculating the T_m for each primer and using a temperature of 1–2°C below this figure. Note that this means the two primers should be designed so that they have identical T_ms. If this is not the case, the appropriate annealing temperature for one primer may be too high or too low for the other member of the pair.

9.3 After the PCR: studying PCR products

PCR is often the starting point for a longer series of experiments in which the amplification product is studied in various ways in order to gain information about the DNA molecule that acted as the original template. We will encounter many studies of this type in Parts II and III, when we examine the applications of gene cloning and PCR

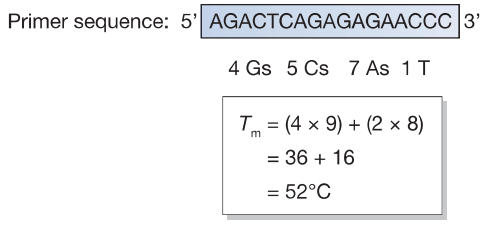


Figure 9.9
Calculating the T_m of a primer.

in research and biotechnology. Although a wide range of procedures have been devised for studying PCR products, three techniques are particularly important:

- Gel electrophoresis of PCR products
- Cloning of PCR products
- Sequencing of PCR products.

The first two of these techniques are dealt with in this chapter. The third technique is deferred until Chapter 10, when all aspects of DNA sequencing will be covered.

9.3.1 Gel electrophoresis of PCR products

The results of most PCR experiments are checked by running a portion of the amplified reaction mixture in an agarose gel. A band representing the amplified DNA may be visible after staining, or if the DNA yield is low the product can be detected by Southern hybridization (p. 142). If the expected band is absent, or if additional bands are present, something has gone wrong and the experiment must be repeated.

In some cases, agarose gel electrophoresis is used not only to determine if a PCR experiment has worked, but also to obtain additional information. For example, the presence of restriction sites in the amplified region of the template DNA can be determined by treating the PCR product with a restriction endonuclease before running the sample in the agarose gel (Figure 9.10). This is a type of **restriction fragment length polymorphism (RFLP) analysis** and is important both in the construction of genome maps (p. 180) and in studying genetic diseases (p. 257).

Alternatively, the exact size of the PCR product can be used to establish if the template DNA contains an insertion or deletion mutation in the amplified region (Figure 9.10). Length mutations of this type form the basis of **DNA profiling**, a central technique in forensic science (Chapter 16).

In some experiments, the mere presence or absence of the PCR product is the diagnostic feature. An example is when PCR is used as the screening procedure to identify a desired gene from a genomic or cDNA library. Carrying out PCRs with every clone in a genomic library might seem to be a tedious task, but one of the advantages of PCR is that individual experiments are quick to set up and many PCRs can be performed in parallel. The workload can also be reduced by **combinatorial screening**, an example of which is shown in Figure 9.11.

9.3.2 Cloning PCR products

Some applications require that after a PCR the resulting products are ligated into a vector and examined by any of the standard methods used for studying cloned DNA. This may sound easy, but there are complications.

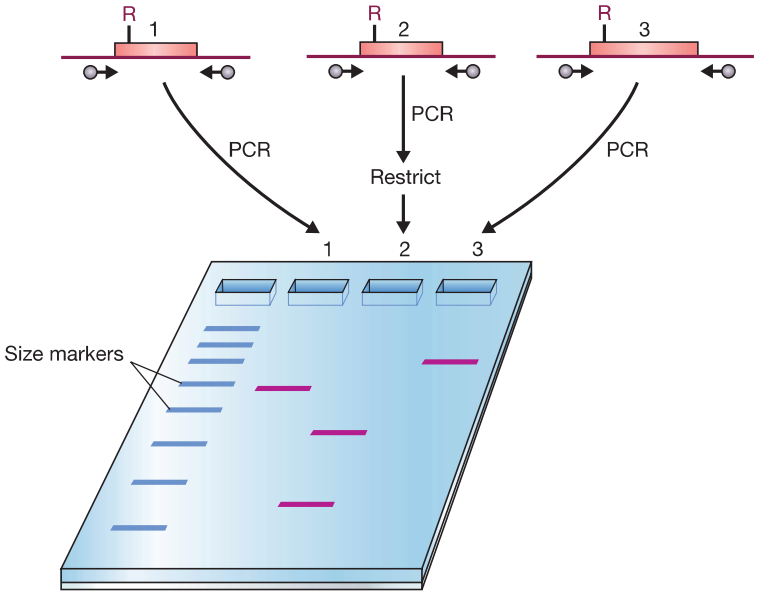


Figure 9.10

Gel electrophoresis of the PCR product can provide information on the template DNA molecule. Lanes 1 and 2 show, respectively, an unrestricted PCR product and a product restricted with the enzyme that cuts at site R. Lane 3 shows the result obtained when the template DNA contains an insertion in the amplified region.

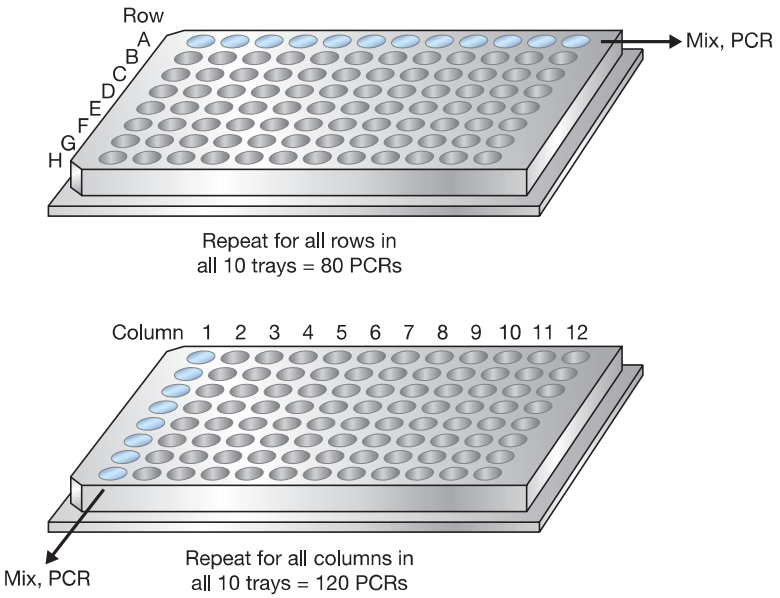


Figure 9.11

Combinatorial screening of clones in microtiter trays. A library of 960 clones is screened by a series of PCRs, each with a combination of clones. The clone combinations that give positive results enable the well(s) containing positive clone(s) to be identified. For example, if positive PCRs are given with row A of tray 2, row D of tray 6, column 7 of tray 2, and column 9 of tray 6, then it can be deduced that there are positive clones in well A7 of tray 2 and well D9 of tray 6. Although there are 960 clones, unambiguous identification of the positive clones is therefore achieved after just 200 PCRs.

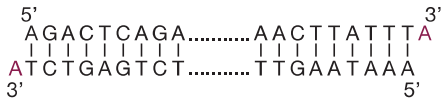


Figure 9.12
Polynucleotides synthesized by *Taq* polymerase often have an extra adenosine at their 3′ ends.

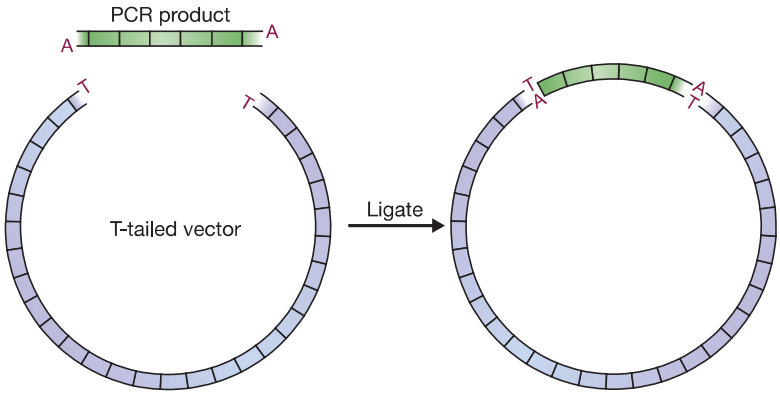


Figure 9.13
Using a special T-tailed vector to clone a PCR product.

The first problem concerns the ends of the PCR products. From an examination of Figure 9.3 it might be imagined that the short products resulting from PCR amplification are blunt-ended. If this was the case they could be inserted into a cloning vector by blunt-end ligation, or alternatively the PCR products could be provided with sticky ends by the attachment of linkers or adaptors (p. 64). Unfortunately, the situation is not so straightforward. *Taq* polymerase tends to add an additional nucleotide, usually an adenosine, to the end of each strand that it synthesizes. This means that a double-stranded PCR product is not blunt-ended, and instead most 3′ termini have a single nucleotide overhang (Figure 9.12). The overhangs could be removed by treatment with an exonuclease enzyme, resulting in PCR products with true blunt ends, but this is not a popular approach as it is difficult to prevent the exonuclease from becoming over-active and causing further damage to the ends of the molecules.

One solution is to use a special cloning vector which carries thymidine (T) overhangs and which can therefore be ligated to a PCR product (Figure 9.13). These vectors are usually prepared by restricting a standard vector at a blunt-end site, and then treating with *Taq* polymerase in the presence of just 2′-deoxythymidine 5′-triphosphate (dTTP). No primer is present so all the polymerase can do is add a T nucleotide to the 3′ ends of the blunt-ended vector molecule, resulting in the T-tailed vector into which the PCR products can be inserted. Special vectors of this type have also been designed for use with the topoisomerase ligation method described on p. 69, and this is currently the most popular way of cloning PCR products.

A second solution is to design primers that contain restriction sites. After PCR the products are treated with the restriction endonuclease, which cuts each molecule within the primer sequence, leaving sticky-ended fragments that can be ligated efficiently into a standard cloning vector (Figure 9.14). The approach is not limited to those instances where the primers span restriction sites that are present in the template DNA. Instead, the restriction site can be included within a short extension at the 5′ end of each primer (Figure 9.15). These extensions cannot hybridize to the template molecule, but they are copied during the PCR, resulting in PCR products that carry terminal restriction sites.

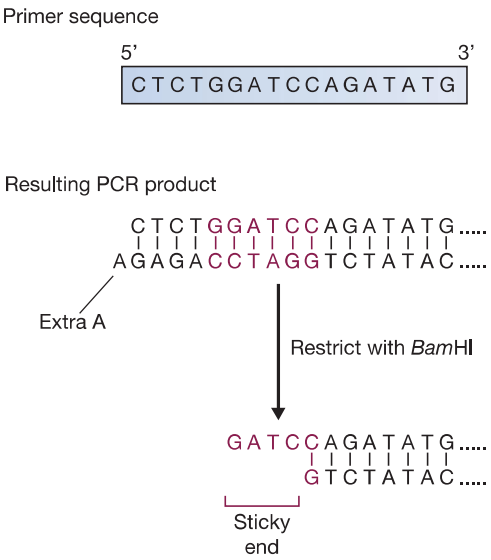


Figure 9.14
Obtaining a PCR product with a sticky end through use of a primer whose sequence includes a restriction site.

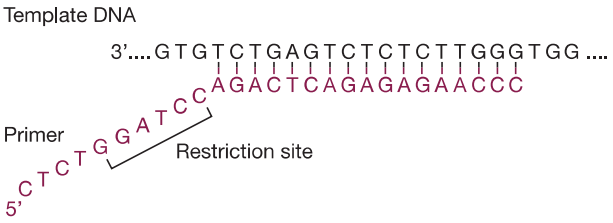


Figure 9.15
A PCR primer with a restriction site present within an extension at the 5′ end.

9.3.3 Problems with the error rate of Taq polymerase

All DNA polymerases make mistakes during DNA synthesis, occasionally inserting an incorrect nucleotide into the growing DNA strand. Most polymerases, however, are able to rectify these errors by reversing over the mistake and resynthesizing the correct sequence. This property is referred to as the "proofreading" function and depends on the polymerase possessing a 3′ to 5′ exonuclease activity (p. 168).

Taq polymerase lacks a proofreading activity and as a result is unable to correct its errors. This means that the DNA synthesized by *Taq* polymerase is not always an accurate copy of the template molecule. The error rate has been estimated at one mistake for every 9000 nucleotides of DNA that is synthesized, which might appear to be almost insignificant but which translates to one error in every 300 bp for the PCR products obtained after 30 cycles. This is because PCR involves copies being made of copies of copies, so the polymerase-induced errors gradually accumulate, the fragments produced at the end of a PCR containing copies of earlier errors together with any new errors introduced during the final round of synthesis.

For many applications this high error rate does not present a problem. In particular, sequencing of a PCR product provides the correct sequence of the template, even though the PCR products contain the errors introduced by *Taq* polymerase. This is because the errors are distributed randomly, so for every molecule that has an error at a particular nucleotide position, there will be many molecules with the correct sequence. In this context the error rate is indeed insignificant.

This is not the case if the PCR products are cloned. Each resulting clone contains multiple copies of a single amplified fragment, so the cloned DNA does not necessarily

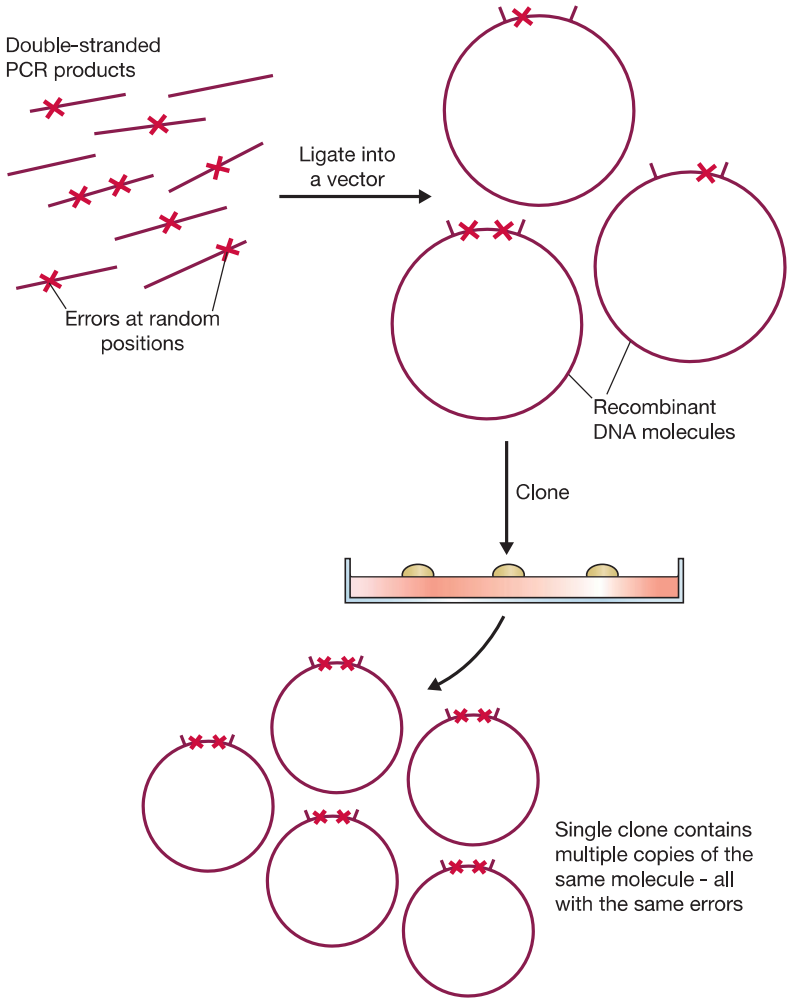


Figure 9.16
The high error rate of *Taq* polymerase becomes a factor when PCR products are cloned.

have the same sequence as the original template molecule used in the PCR (Figure 9.16). This possibility lends an uncertainty to all experiments carried out with cloned PCR products and dictates that, whenever possible, the amplified DNA should be studied directly rather than being cloned.

9.4 Real-time PCR enables the amount of starting material to be quantified

The amount of product that is synthesized during a set number of cycles of a PCR depends on the number of DNA molecules that are present in the starting mixture (Table 9.1). If there are only a few DNA molecules at the beginning of the PCR then relatively little product will be made, but if there are many starting molecules then the product yield will be higher. This relationship enables PCR to be used to quantify the number of DNA molecules present in an extract.

Table 9.1

Number of short products synthesized after 25 cycles of PCR with different numbers of starting molecule.

NUMBER OF STARTING MOLECULES	NUMBER OF SHORT PRODUCTS
1	4,194,304
2	8,388,608
5	20,971,520
10	41,943,040
25	104,857,600
50	209,715,200
100	419,430,400

Note: The numbers assume that amplification is 100% efficient, none of the reactants becoming limiting during the course of the PCR.

9.4.1 Carrying out a quantitative PCR experiment

In **quantitative PCR (qPCR)** the amount of product synthesized during a test PCR is compared with the amounts synthesized during PCRs with known quantities of starting DNA. In the early procedures, agarose gel electrophoresis was used to make these comparisons. After staining the gel, the band intensities were examined to identify the control PCR whose product was most similar to that of the test (Figure 9.17). Although easy to perform, this type of qPCR is imprecise, because large differences in the amount of starting DNA give relatively small differences in the band intensities of the resulting PCR products.

Today, quantification is carried out by **real-time PCR**, a modification of the standard PCR technique in which synthesis of the product is measured over time, as the PCR proceeds through its series of cycles. There are two ways of following product synthesis in real time:

- A dye that gives a fluorescent signal when it binds to double-stranded DNA can be included in the PCR mixture. This method measures the total amount of double-stranded DNA in the PCR at any particular time, which may over-estimate the actual amount of the product because sometimes the primers anneal to one another in various non-specific ways, increasing the amount of double-stranded DNA that is present.
- A short oligonucleotide called a **reporter probe**, which gives a fluorescent signal when it hybridizes to the PCR product, can be used. Because the probe only hybridizes to the PCR product, this method is less prone to inaccuracies caused by primer-primer annealing. Each probe molecule has pair of labels. A fluorescent dye

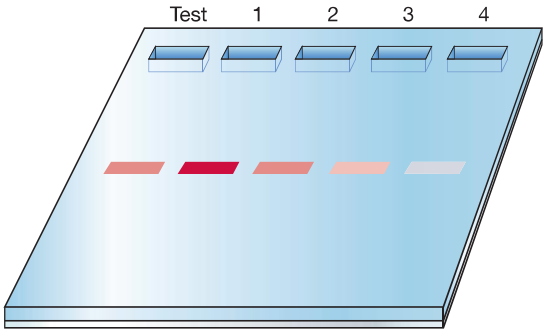


Figure 9.17

Using agarose gel electrophoresis to quantify the amount of DNA in a test PCR. Lanes 1 to 4 are control PCRs carried out with decreasing amounts of template DNA. The intensity of staining for the test band suggests that this PCR contained approximately the same amount of DNA as the control run in lane 2.

Figure 9.18
Hybridization of a reporter probe to its target DNA.

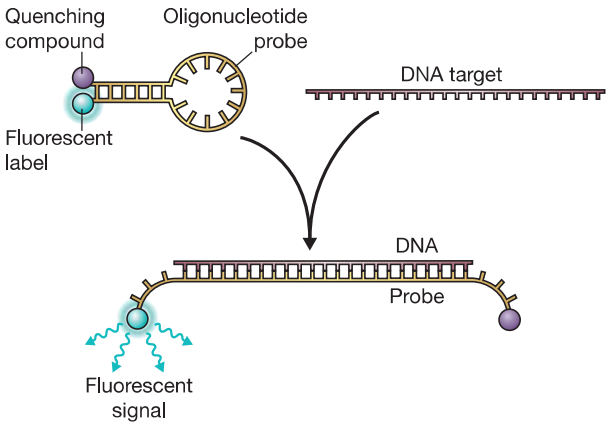
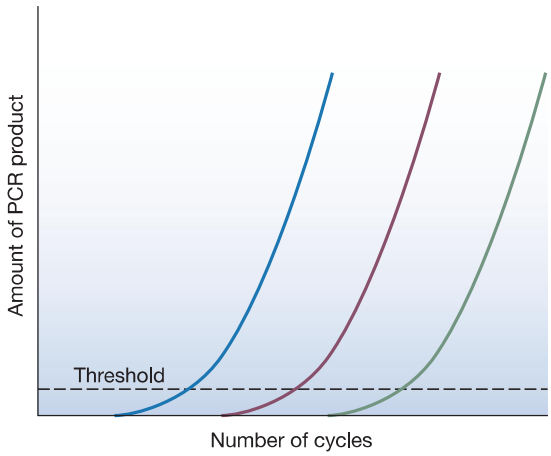


Figure 9.19
Quantification by real-time PCR. The graph shows product synthesis during three PCRs, each with a different amount of starting DNA. During a PCR, product accumulates exponentially, the amount present at any particular cycle proportional to the amount of starting DNA. The blue curve is therefore the PCR with the greatest amount of starting DNA, and the green curve is the one with the least starting DNA. If the amounts of starting DNA in these three PCRs are known, then the amount in a test PCR can be quantified by comparison with these controls. In practice, the comparison is made by identifying the cycle at which product synthesis moves above a threshold amount, indicated by the horizontal line on the graph.



is attached to one end of the oligonucleotide, and a quenching compound, which inhibits the fluorescent signal, is attached to the other end (Figure 9.18). Normally there is no fluorescence because the oligonucleotide is designed in such a way that its two ends base pair to one another, placing the quencher next to the dye. Hybridization between the oligonucleotide and the PCR product disrupts this base pairing, moving the quencher away from the dye and enabling the fluorescent signal to be generated.

Both systems enable synthesis of the PCR product to be followed by measuring the fluorescent signal. Quantification again requires comparison between test and control PCRs, usually by identifying the stage in the PCR at which the amount of fluorescent signal reaches a pre-set threshold (Figure 9.19). The more rapidly the threshold is reached, the greater the amount of DNA in the starting mixture.

9.4.2 Real-time PCR can also quantify RNA

Real-time PCR is often used to quantify the amount of DNA in an extract, for example to follow the progression of a viral infection by measuring the amount of pathogen DNA that is present in a tissue. More frequently, however, the method is used as a means of measuring RNA amounts, in particular to determine the extent of expression of a particular gene by quantifying its mRNA. The gene under study might be one that is switched on in cancerous cells, in which case quantifying its mRNA will enable the

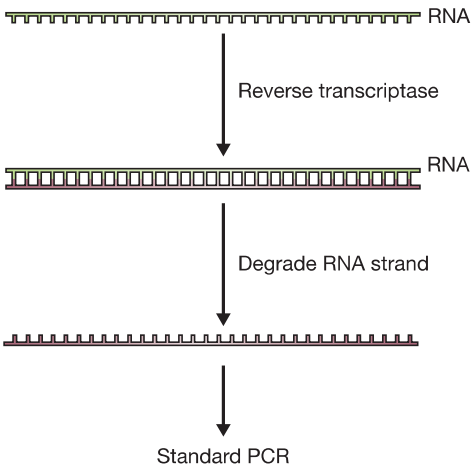


Figure 9.20
Reverse transcriptase PCR.

development of the cancer to be monitored and the effects of subsequent treatment to be assessed.

How do we carry out PCR if RNA is the starting material? The answer is to use **reverse transcriptase PCR**. The first step in this procedure is to convert the RNA molecules into single-stranded complementary DNA (cDNA) (Figure 9.20). Once this preliminary step has been carried out, the PCR primers and *Taq* polymerase are added and the experiment proceeds exactly as in the standard technique. Some thermostable polymerases are able to make DNA copies of both RNA and DNA molecules (i.e., they have both reverse transcriptase and DNA-dependent DNA polymerase activities) and so can carry out all the steps of this type of PCR in a single reaction.

Further reading

Higuchi, R., Dollinger, G., Walsh, P.S. & Griffith, R. (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, 10, 413–417. [The first description of real-time PCR].

Marchuk, D., Drumm, M., Saulino, A. & Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research*, 19, 1154.

Rychlik, W., Spencer, W.J. & Rhoads, R.E. (1990) Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Research*, 18, 6409–6412.

Saiki, R.K., Gelfand, D.H., Stoffel, S. et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, 487–491. [The first description of PCR with *Taq* polymerase.]

Tindall, K.R. & Kunkel, T.A. (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry*, 27, 6008–6013. [Describes the error rate of *Taq* polymerase.]

VanGuilder, H.D., Vrana, K.E. & Freeman, W.M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, 44, 619–624.