CHAPTER 2
Basic techniques

Introduction

The initial impetus for gene manipulation *in vitro* came about in the early 1970s with the simultaneous development of techniques for:

- genetic transformation of *Escherichia coli*;
- cutting and joining DNA molecules;
- monitoring the cutting and joining reactions.

In order to explain the significance of these developments we must first consider the essential requirements of a successful gene-manipulation procedure.

The basic problems

Before the advent of modern gene-manipulation methods there had been many early attempts at transforming pro- and eukaryotic cells with foreign DNA. But, in general, little progress could be made. The reasons for this are as follows. Let us assume that the exogenous DNA is taken up by the recipient cells. There are then two basic difficulties. First, where detection of uptake is dependent on gene expression, failure could be due to lack of accurate transcription or translation. Secondly, and more importantly, the exogenous DNA may not be maintained in the transformed cells. If the exogenous DNA is integrated into the host genome, there is no problem. The exact mechanism whereby this integration occurs is not clear and it is usually a rare event. However this occurs, the result is that the foreign DNA sequence becomes incorporated into the host cell's genetic material and will subsequently be propagated as part of that genome. If, however, the exogenous DNA fails to be integrated, it will probably be lost during subsequent multiplication of the host cells. The reason for this is simple. In order to be replicated, DNA molecules must contain an *origin of replication*, and in bacteria and viruses there is usually only one per genome. Such molecules are called *replicons*. Fragments of DNA are not replicons and in the absence of replication will be diluted out of their host cells. It should be noted that, even if a DNA molecule contains an origin of replication, this may not function in a foreign host cell.

There is an additional, subsequent problem. If the early experiments were to proceed, a method was required for assessing the fate of the donor DNA. In particular, in circumstances where the foreign DNA was maintained because it had become integrated in the host DNA, a method was required for mapping the foreign DNA and the surrounding host sequences.

The solutions: basic techniques

If fragments of DNA are not replicated, the obvious solution is to attach them to a suitable replicon. Such replicons are known as *vectors* or *cloning vehicles*. Small plasmids and bacteriophages are the most suitable vectors for they are replicons in their own right, their maintenance does not necessarily require integration into the host genome and their DNA can be readily isolated in an intact form. The different plasmids and phages which are used as vectors are described in detail in Chapters 4 and 5. Suffice it to say at this point that initially plasmids and phages suitable as vectors were only found in *E. coli*. An important consequence follows from the use of a vector to carry the foreign DNA: simple methods become available for purifying the vector molecule, complete with its foreign DNA insert, from transformed host cells. Thus not only does the vector provide the replicon function, but it also permits the easy bulk preparation of the foreign DNA sequence, free from host-cell DNA.

Composite molecules in which foreign DNA has been inserted into a vector molecule are sometimes called DNA *chimeras* because of their analogy with the Chimaera of mythology – a creature with the head of a lion, body of a goat and tail of a serpent. The construction of such composite or *artificial*

recombinant molecules has also been termed *genetic engineering* or *gene manipulation* because of the potential for creating novel genetic combinations by biochemical means. The process has also been termed *molecular cloning* or *gene cloning* because a line of genetically identical organisms, all of which contain the composite molecule, can be propagated and grown in bulk, hence *amplifying* the composite molecule and *any gene product whose synthesis it directs*.

Although conceptually very simple, cloning of a fragment of foreign, or *passenger*, or *target* DNA in a vector demands that the following can be accomplished.

- The vector DNA must be purified and cut open.
- The passenger DNA must be inserted into the vector molecule to create the artificial recombinant. DNA joining reactions must therefore be performed. Methods for cutting and joining DNA molecules are now so sophisticated that they warrant a chapter of their own (Chapter 3).
- The cutting and joining reactions must be readily monitored. This is achieved by the use of gel electrophoresis.
- Finally, the artificial recombinant must be transformed into *E. coli* or another host cell. Further details on the use of gel electrophoresis and transformation of *E. coli* are given in the next section. As we have noted, the necessary techniques became available at about the same time and quickly led to many cloning experiments, the first of which were reported in 1972 (Jackson *et al.* 1972, Lobban & Kaiser 1973).

Agarose gel electrophoresis

The progress of the first experiments on cutting and joining of DNA molecules was monitored by velocity sedimentation in sucrose gradients. However, this has been entirely superseded by gel electrophoresis. Gel electrophoresis is not only used as an analytical method, it is routinely used preparatively for the purification of specific DNA fragments. The gel is composed of polyacrylamide or agarose. Agarose is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb (Fig. 2.1). Polyacrylamide is preferred for smaller DNA fragments.

The mechanism responsible for the separation of DNA molecules by molecular weight during gel

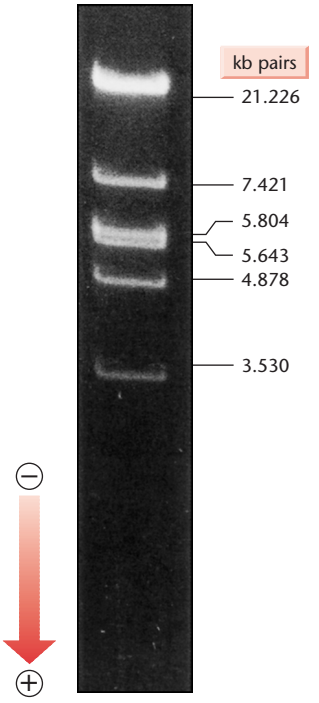


Fig. 2.1 Electrophoresis of DNA in agarose gels. The direction of migration is indicated by the arrow. DNA bands have been visualized by soaking the gel in a solution of ethidium bromide (see Fig. 2.3), which complexes with DNA by intercalating between stacked base-pairs, and photographing the orange fluorescence which results upon ultraviolet irradiation.

electrophoresis is not well understood (Holmes & Stellwagen 1990). The migration of the DNA molecules through the pores of the matrix must play an important role in molecular-weight separations since the electrophoretic mobility of DNA in free solution is independent of molecular weight. An agarose gel is a complex network of polymeric molecules whose average pore size depends on the buffer composition and the type and concentration of agarose used. DNA movement through the gel was originally thought to resemble the motion of a snake (reptation). However, real-time fluorescence microscopy of stained molecules undergoing electrophoresis has revealed more subtle dynamics (Schwartz & Koval 1989, Smith *et al.* 1989). DNA molecules display elastic behaviour by stretching in the direction of the applied field and then contracting into dense balls. The larger the pore size of the

gel, the greater the ball of DNA which can pass through and hence the larger the molecules which can be separated. Once the globular volume of the DNA molecule exceeds the pore size, the DNA molecule can only pass through by reptation. This occurs with molecules about 20 kb in size and it is difficult to separate molecules larger than this without recourse to pulsed electrical fields.

In pulsed-field gel electrophoresis (PFGE) (Schwartz & Cantor 1984) molecules as large as 10 Mb can be separated in agarose gels. This is achieved by causing the DNA to periodically alter its direction of migration by regular changes in the orientation of the electric field with respect to the gel. With each change in the electric-field orientation, the DNA must realign its axis prior to migrating in the new direction. Electric-field parameters, such as the direction, intensity and duration of the electric field, are set independently for each of the different fields and are chosen so that the net migration of the DNA is down the gel. The difference between the direction of migration induced by each of the electric fields is the *reorientation angle* and corresponds to the angle that the DNA must turn as it changes its direction of migration each time the fields are switched.

A major disadvantage of PFGE, as originally described, is that the samples do not run in straight lines. This makes subsequent analysis difficult. This problem has been overcome by the development of improved methods for alternating the electrical field. The most popular of these is contour-clamped homogeneous electrical-field electrophoresis (CHEF) (Chu *et al.* 1986). In early CHEF-type systems (Fig. 2.2) the reorientation angle was fixed at 120°. However, in newer systems, the reorientation angle can be varied and it has been found that for whole-yeast chromosomes the migration rate is much faster with an angle of 106° (Birren *et al.* 1988). Fragments of DNA as large as 200–300 kb are routinely handled in genomics work and these can be separated in a matter of hours using CHEF systems with a reorientation angle of 90° or less (Birren & Lai 1994).

Aaij and Borst (1972) showed that the migration rates of the DNA molecules were inversely proportional to the logarithms of the molecular weights. Subsequently, Southern (1979a,b) showed that plotting fragment length or molecular weight against the reciprocal of mobility gives a straight

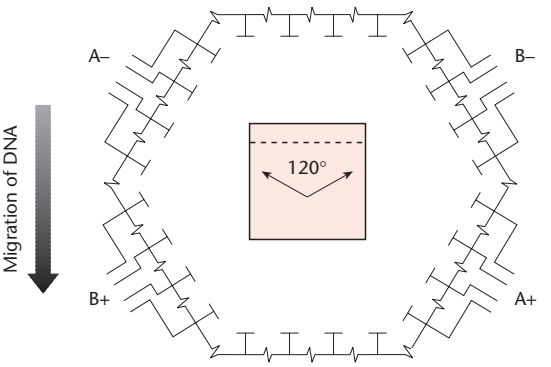


Fig. 2.2 Schematic representation of CHEF (contour-clamped homogeneous electrical field) pulsed-field gel electrophoresis.

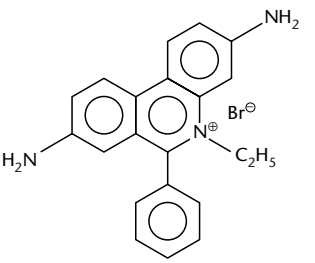


Fig. 2.3 Ethidium bromide.

line over a wider range than the semilogarithmic plot. In any event, gel electrophoresis is frequently performed with marker DNA fragments of known size, which allow accurate size determination of an unknown DNA molecule by interpolation. A particular advantage of gel electrophoresis is that the DNA bands can be readily detected at high sensitivity. The bands of DNA in the gel are stained with the intercalating dye ethidium bromide (Fig. 2.3), and as little as 0.05 μg of DNA in one band can be detected as visible fluorescence when the gel is illuminated with ultraviolet light.

In addition to resolving DNA fragments of different lengths, gel electrophoresis can be used to separate different molecular configurations of a DNA molecule. Examples of this are given in Chapter 4 (see p. 44). Gel electrophoresis can also be used for investigating protein–nucleic acid interactions in the so-called *gel retardation* or *band shift* assay. It is based on the observation that binding of a protein to DNA fragments usually leads to a reduction in

electrophoretic mobility. The assay typically involves the addition of protein to linear double-stranded DNA fragments, separation of complex and naked DNA by gel electrophoresis and visualization. A review of the physical basis of electrophoretic mobility shifts and their application is provided by Lane *et al*. (1992).

Nucleic acid blotting

Nucleic acid labelling and hybridization on membranes have formed the basis for a range of experimental techniques central to recent advances in our understanding of the organization and expression of the genetic material. These techniques may be applied in the isolation and quantification of specific nucleic acid sequences and in the study of their organization, intracellular localization, expression and regulation. A variety of specific applications includes the diagnosis of infectious and inherited disease. Each of these topics is covered in depth in subsequent chapters.

An overview of the steps involved in nucleic acid blotting and membrane hybridization procedures is shown in Fig. 2.4. *Blotting* describes the immobilization of sample nucleic acids on to a solid support, generally nylon or nitrocellulose membranes. The blotted nucleic acids are then used as 'targets' in subsequent hybridization experiments. The main blotting procedures are:

- blotting of nucleic acids from gels;
- dot and slot blotting;
- colony and plaque blotting.

Colony and plaque blotting are described in detail on pp. 104–105 and dot and slot blotting in Chapter 14.

Southern blotting

The original method of blotting was developed by Southern (1975, 1979b) for detecting fragments in an agarose gel that are complementary to a given RNA or DNA sequence. In this procedure, referred to as Southern blotting, the agarose gel is mounted on a filter-paper wick which dips into a reservoir containing transfer buffer (Fig. 2.5). The hybridization membrane is sandwiched between the gel and a stack of paper towels (or other absorbent material), which serves to draw the transfer buffer through the gel by capillary action. The DNA molecules are carried out of the gel by the buffer flow and immobilized on the membrane. Initially, the membrane material used was nitrocellulose. The main drawback with this membrane is its fragile nature. Supported nylon membranes have since been developed which have greater binding capacity for nucleic acids in addition to high tensile strength.

For efficient Southern blotting, gel pretreatment is important. Large DNA fragments (> 10 kb) require a longer transfer time than short fragments. To allow

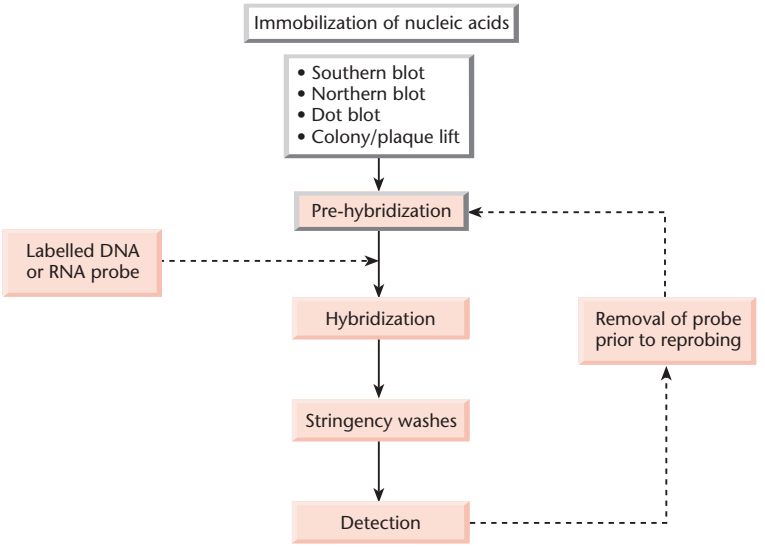


Fig. 2.4 Overview of nucleic acid blotting and hybridization (reproduced courtesy of Amersham Pharmacia Biotech).

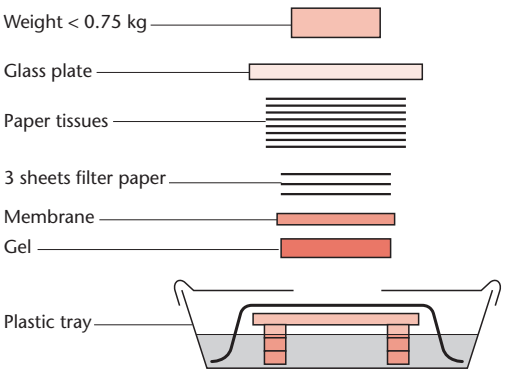


Fig. 2.5 A typical capillary blotting apparatus.

uniform transfer of a wide range of DNA fragment sizes, the electrophoresed DNA is exposed to a short depurination treatment (0.25 mol/l HCl) followed by alkali. This shortens the DNA fragments by alkaline hydrolysis at depurinated sites. It also denatures the fragments prior to transfer, ensuring that they are in the single-stranded state and accessible for probing. Finally, the gel is equilibrated in neutralizing solution prior to blotting. An alternative method uses positively charged nylon membranes, which remove the need for extended gel pretreatment. With them the DNA is transferred in native (non-denatured) form and then alkali-denatured *in situ* on the membrane.

After transfer, the nucleic acid needs to be fixed to the membrane and a number of methods are available. Oven baking at 80°C is the recommended method for nitrocellulose membranes and this can also be used with nylon membranes. Due to the flammable nature of nitrocellulose, it is important that it is baked in a vacuum oven. An alternative fixation method utilizes ultraviolet cross-linking. It is based on the formation of cross-links between a small fraction of the thymine residues in the DNA and positively charged amino groups on the surface of nylon membranes. A calibration experiment must be performed to determine the optimal fixation period.

Following the fixation step, the membrane is placed in a solution of labelled (radioactive or non-radioactive) RNA, single-stranded DNA or oligodeoxynucleotide which is complementary in sequence to the blot-transferred DNA band or bands to be detected. Conditions are chosen so that the labelled nucleic acid hybridizes with the DNA on the membrane. Since this labelled nucleic acid is used to detect and locate the complementary sequence, it is called the *probe*. Conditions are chosen which maximize the rate of hybridization, compatible with a low background of non-specific binding on the membrane (see Box 2.1). After the hybridization reaction has been carried out, the membrane is washed to remove unbound radioactivity and regions of hybridization

Box 2.1 Hybridization of nucleic acids on membranes

The hybridization of nucleic acids on membranes is a widely used technique in gene manipulation and analysis. Unlike solution hybridizations, membrane hybridizations tend not to proceed to completion. One reason for this is that some of the bound nucleic acid is embedded in the membrane and is inaccessible to the probe. Prolonged incubations may not generate any significant increase in detection sensitivity.

The composition of the hybridization buffer can greatly affect the speed of the reaction and the sensitivity of detection. The key components of these buffers are shown below:

Rate enhancers	Dextran sulphate and other polymers act as volume excluders to increase both the rate and the extent of hybridization
Detergents and blocking agents	Dried milk, heparin and detergents such as sodium dodecyl sulphate (SDS) have been used to depress non-specific binding of the probe to the membrane. Denhardt's solution (Denhardt 1966) uses Ficoll, polyvinylpyrrolidone and bovine serum albumin
Denaturants	Urea or formamide can be used to depress the melting temperature of the hybrid so that reduced temperatures of hybridization can be used
Heterologous DNA	This can reduce non-specific binding of probes to non-homologous DNA on the blot

continued

Box 2.1 *continued*

Stringency control

Stringency can be regarded as the specificity with which a particular target sequence is detected by hybridization to a probe. Thus, at high stringency, only completely complementary sequences will be bound, whereas low-stringency conditions will allow hybridization to partially matched sequences. Stringency is most commonly controlled by the temperature and salt concentration in the post-hybridization washes, although these parameters can also be utilized in the hybridization step. In practice, the stringency washes are performed under successively more stringent conditions (lower salt or higher temperature) until the desired result is obtained.

The melting temperature (T_m) of a probe–target hybrid can be calculated to provide a starting-point for the determination of correct stringency. The T_m is the temperature at which the probe and target are 50% dissociated. For probes longer than 100 base pairs:

$$T_m = 81.5°C + 16.6 \log M + 0.41 (\%\ G + C)$$

where M = ionic strength of buffer in moles/litre. With long probes, the hybridization is usually carried out at $T_m - 25°C$. When the probe is used to detect partially matched sequences, the hybridization temperature is reduced by 1°C for every 1% sequence divergence between probe and target.

Oligonucleotides can give a more rapid hybridization rate than long probes as they can be used at a higher molarity. Also, in situations where target is in excess to the probe, for example dot blots, the hybridization rate is diffusion-limited and longer probes diffuse more slowly than oligonucleotides. It is standard practice to use oligonucleotides to analyse putative mutants following a site-directed mutagenesis experiment where the difference between parental and mutant progeny is often only a single base-pair change (see p. 132 *et seq.*).

The availability of the exact sequence of oligonucleotides allows conditions for hybridization and stringency washing to be tightly controlled so that the probe will only remain hybridized when it is 100% homologous to the target. Stringency is commonly controlled by adjusting the temperature of the wash buffer. The 'Wallace rule' (Lay Thein & Wallace 1986) is used to determine the appropriate stringency wash temperature:

$$T_m = 4 \times (\text{number of GC base pairs}) + 2 \times (\text{number of AT base pairs})$$

In filter hybridizations with oligonucleotide probes, the hybridization step is usually performed at 5°C below T_m for perfectly matched sequences. For every mismatched base pair, a further 5°C reduction is necessary to maintain hybrid stability.

The design of oligonucleotides for hybridization experiments is critical to maximize hybridization specificity. Consideration should be given to:

- probe length – the longer the oligonucleotide, the less chance there is of it binding to sequences other than the desired target sequence under conditions of high stringency;
- oligonucleotide composition – the GC content will influence the stability of the resultant hybrid and hence the determination of the appropriate stringency washing conditions. Also the presence of any non-complementary bases will have an effect on the hybridization conditions.

are detected autoradiographically by placing the membrane in contact with X-ray film (see Box 2.2). A common approach is to carry out the hybridization under conditions of relatively low stringency which permit a high rate of hybridization, followed by a series of post-hybridization washes of increasing stringency (i.e. higher temperature or, more commonly, lower ionic strength). Autoradiography following each washing stage will reveal any DNA bands that are related to, but not perfectly complementary with, the probe and will also permit an estimate of the degree of mismatching to be made.

Box 2.2 The principles of autoradiography

The localization and recording of a radiolabel within a solid specimen is known as autoradiography and involves the production of an image in a photographic emulsion. Such emulsions consist of silver halide crystals suspended in a clear phase composed mainly of gelatin. When a β-particle or γ-ray from a radionuclide passes through the emulsion, the silver ions are converted to silver atoms. This results in a latent image being produced, which is converted to a visible image when the image is developed. Development is a system of amplification in which the silver atoms cause the entire silver halide crystal to be reduced to metallic silver. Unexposed crystals are removed by dissolution in fixer, giving an autoradiographic image which represents the distribution of radiolabel in the original sample.

In direct autoradiography, the sample is placed in intimate contact with the film and the radioactive emissions produce black areas on the developed autoradiograph. It is best suited to detection of weak- to medium-strength β-emitting radionuclides (^{3}H, ^{14}C, ^{35}S). Direct autoradiography is not suited to the detection of highly energetic β-particles, such as those from ^{32}P, or for γ-rays emitted from isotopes like ^{125}I. These emissions pass through and beyond the film, with the majority of the energy being wasted. Both ^{32}P and ^{125}I are best detected by indirect autoradiography.

Indirect autoradiography describes the technique by which emitted energy is converted to light by means of a scintillator, using fluorography or intensifying screens. In fluorography the sample is impregnated with a liquid scintillator. The radioactive emissions transfer their energy to the scintillator molecules, which then emit photons which expose the photographic emulsion. Fluorography is mostly used to improve the detection of weak β-emitters (Fig. B2.1). Intensifying screens are

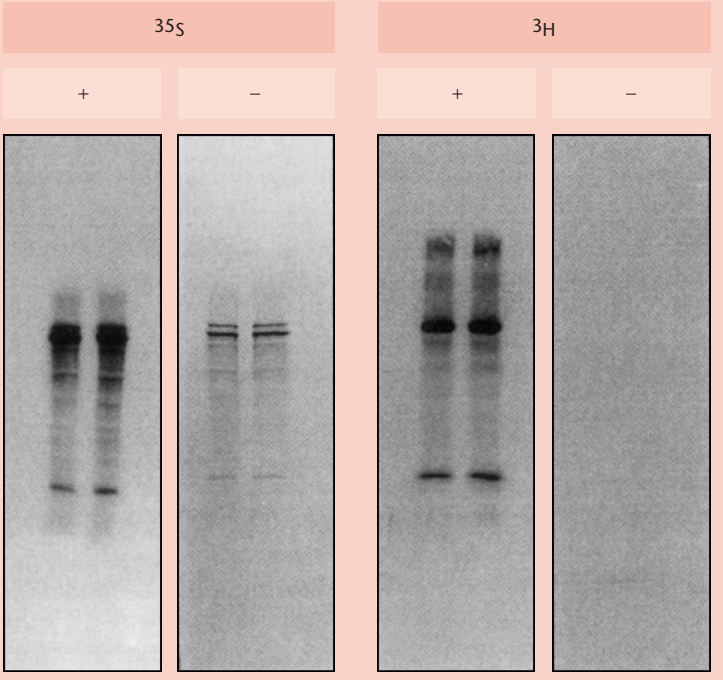


Fig. B2.1 Autoradiographs showing the detection of ^{35}S- and ^{3}H-labelled proteins in acrylamide gels with (+) and without (–) fluorography. (Photo courtesy of Amersham Pharmacia Biotech.)

continued

Box 2.2 *continued*

sheets of a solid inorganic scintillator which are placed behind the film. Any emissions passing through the photographic emulsion are absorbed by the screen and converted to light, effectively superimposing a photographic image upon the direct autoradiographic image.

The gain in sensitivity which is achieved by use of indirect autoradiography is offset by non-linearity of film response. A single hit by a β-particle or γ-ray can produce hundreds of silver atoms, but a single hit by a photon of light produces only a single silver atom. Although two or more silver atoms in a silver halide crystal are stable, a single silver atom is unstable and reverts to a silver ion very rapidly. This means that the probability of a second photon being captured before the first silver atom has reverted is greater for large amounts of radioactivity than for small amounts. Hence small amounts of radioactivity are under-represented with the use of fluorography and intensifying screens. This problem can be overcome by a combination of pre-exposing a film to an instantaneous flash of light (pre-flashing) and exposing the autoradiograph at −70°C. Pre-flashing provides many of the silver halide crystals of the film with a stable pair of silver atoms. Lowering the temperature to −70°C increases the stability of a single silver atom, increasing the time available to capture a second photon (Fig. B2.2).

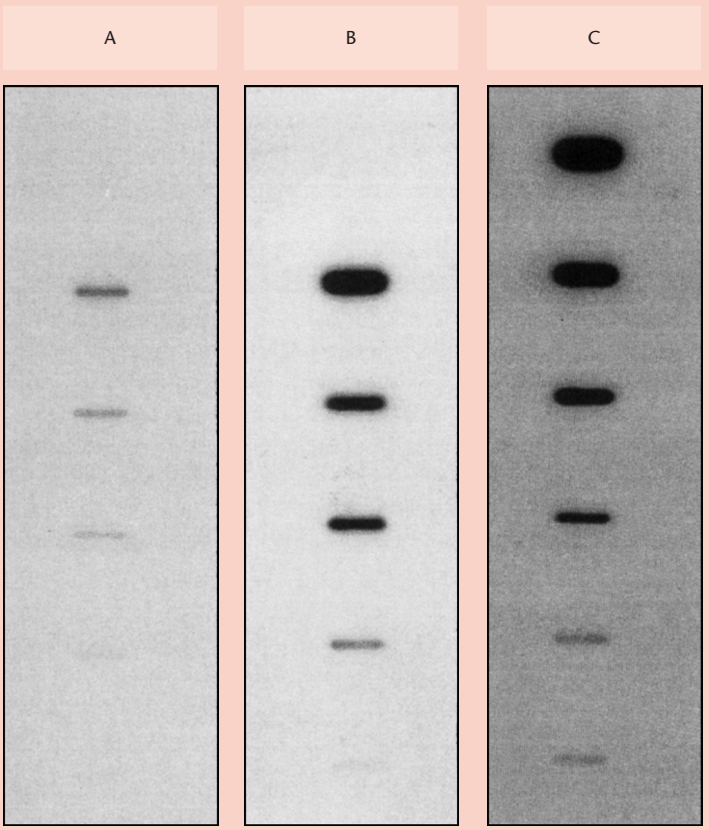


Fig. B2.2 The improvement in sensitivity of detection of ^{125}I-labelled IgG by autoradiography obtained by using an intensifying screen and pre-flashed film. A, no screen and no pre-flashing; B, screen present but film not pre-flashed; C, use of screen and pre-flashed film. (Photo courtesy of Amersham Pharmacia Biotech.)

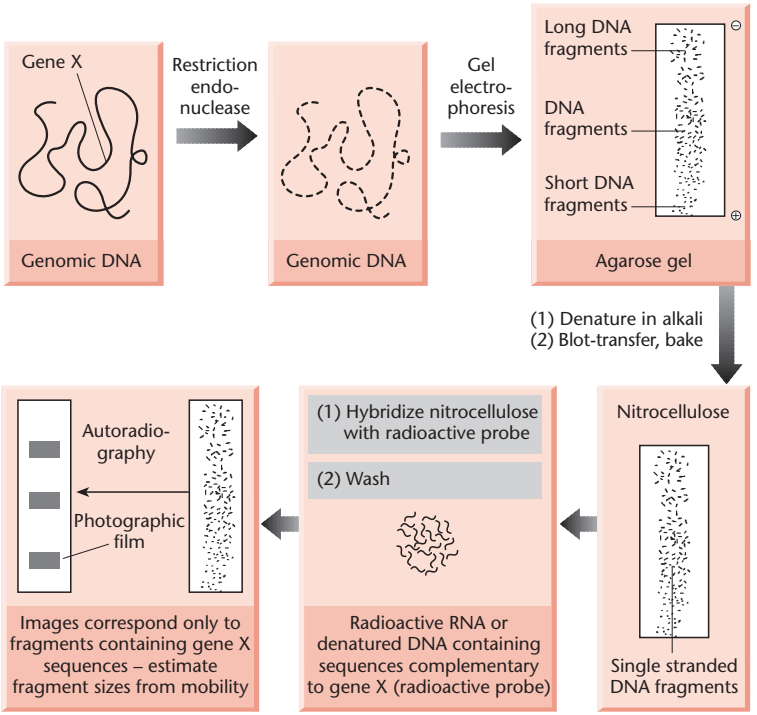


Fig. 2.6 Mapping restriction sites around a hypothetical gene sequence in total genomic DNA by the Southern blot method.

Genomic DNA is cleaved with a restriction endonuclease into hundreds of thousands of fragments of various sizes. The fragments are separated according to size by gel electrophoresis and blot-transferred on to nitrocellulose paper. Highly radioactive RNA or denatured DNA complementary in sequence to gene *X* is applied to the nitrocellulose paper bearing the blotted DNA. The radiolabelled RNA or DNA will hybridize with gene *X* sequences and can be detected subsequently by autoradiography, so enabling the sizes of restriction fragments containing gene *X* sequences to be estimated from their electrophoretic mobility. By using several restriction endonucleases singly and in combination, a map of restriction sites in and around gene *X* can be built up.

The Southern blotting methodology can be extremely sensitive. It can be applied to mapping restriction sites around a single-copy gene sequence in a complex genome such as that of humans (Fig. 2.6), and when a 'mini-satellite' probe is used it can be applied forensically to minute amounts of DNA (see Chapter 14).

Northern blotting

Southern's technique has been of enormous value, but it was thought that it could not be applied directly to the blot-transfer of RNAs separated by gel electrophoresis, since RNA was found not to bind to nitrocellulose. Alwine *et al.* (1979) therefore devised a procedure in which RNA bands are blot-transferred from the gel on to chemically reactive paper, where they are bound covalently. The reactive paper is prepared by diazotization of aminobenzyloxymethyl paper (creating diazobenzyloxymethyl (DBM) paper), which itself can be prepared from Whatman 540 paper by a series of uncomplicated reactions. Once covalently bound, the RNA is available for hybridization with radiolabelled DNA probes. As before, hybridizing bands are located by autoradiography. Alwine *et al.*'s method thus extends that of Southern and for this reason it has acquired the jargon term *northern* blotting.

Subsequently it was found that RNA bands can indeed be blotted on to nitrocellulose membranes under appropriate conditions (Thomas 1980) and suitable nylon membranes have been developed. Because of the convenience of these more recent methods, which do not require freshly activated paper, the use of DBM paper has been superseded.

Western blotting

The term 'western' blotting (Burnette 1981) refers to a procedure which does not directly involve nucleic acids, but which is of importance in gene manipulation. It involves the transfer of electrophoresed protein bands from a polyacrylamide gel on to a membrane of nitrocellulose or nylon, to which they bind strongly (Gershoni & Palade 1982, Renart & Sandoval 1984). The bound proteins are then avail-

able for analysis by a variety of specific protein–ligand interactions. Most commonly, antibodies are used to detect specific antigens. Lectins have been used to identify glycoproteins. In these cases the probe may itself be labelled with radioactivity, or some other 'tag' may be employed. Often, however, the probe is unlabelled and is itself detected in a 'sandwich' reaction, using a second molecule which is labelled, for instance a species-specific second antibody, or protein A of *Staphylococcus aureus* (which binds to certain subclasses of IgG antibodies), or streptavidin (which binds to antibody probes that have been biotinylated). These second molecules may be labelled in a variety of ways with radioactive, enzyme or fluorescent tags. An advantage of the sandwich approach is that a single preparation of labelled second molecule can be employed as a general detector for different probes. For example, an antiserum may be raised in rabbits which reacts with a range of mouse immunoglobins. Such a rabbit anti-mouse (RAM) antiserum may be radiolabelled and used in a number of different applications to identify polypeptide bands probed with different, specific, monoclonal antibodies, each monoclonal antibody being of mouse origin. The sandwich method may also give a substantial increase in sensitivity, owing to the multivalent binding of antibody molecules.

Alternative blotting techniques

The original blotting technique employed capillary blotting but nowadays the blotting is usually accomplished by electrophoretic transfer of polypeptides from an SDS-polyacrylamide gel on to the membrane (Towbin *et al.* 1979). Electrophoretic transfer is also the method of choice for transferring DNA or RNA from low-pore-size polyacrylamide gels. It can also be used with agarose gels. However, in this case, the rapid electrophoretic transfer process requires high currents, which can lead to extensive heating effects, resulting in distortion of agarose gels. The use of an external cooling system is necessary to prevent this.

Another alternative to capillary blotting is vacuum-driven blotting (Olszewska & Jones 1988), for which several devices are commercially available. Vacuum blotting has several advantages over capillary or electrophoretic transfer methods: transfer is very rapid and gel treatment can be performed *in situ* on the vacuum apparatus. This ensures minimal gel handling and, together with the rapid transfer, prevents significant DNA diffusion.

Transformation of *E. coli*

Early attempts to achieve transformation of *E. coli* were unsuccessful and it was generally believed that *E. coli* was refractory to transformation. However, Mandel and Higa (1970) found that treatment with $CaC1_2$ allowed *E. coli* cells to take up DNA from bacteriophage λ. A few years later Cohen *et al.* (1972) showed that $CaC1_2$-treated *E. coli* cells are also effective recipients for plasmid DNA. Almost any strain of *E. coli* can be transformed with plasmid DNA, albeit with varying efficiency, whereas it was thought that only *rec*BC$^-$ mutants could be transformed with linear bacterial DNA (Cosloy & Oishi 1973). Later, Hoekstra *et al.* (1980) showed that *rec*BC$^+$ cells can be transformed with linear DNA, but the efficiency is only 10% of that in otherwise isogenic *rec*BC$^-$ cells. Transformation of *rec*BC$^-$ cells with linear DNA is only possible if the cells are rendered recombination-proficient by the addition of a *sbc*A or *sbc*B mutation. The fact that the *rec*BC gene product is an exonuclease explains the difference in transformation efficiency of circular and linear DNA in *rec*BC$^+$ cells.

As will be seen from the next chapter, many bacteria contain restriction systems which can influence the efficiency of transformation. Although the complete function of these restriction systems is not yet known, one role they do play is the recognition and degradation of foreign DNA. For this reason it is usual to use a restriction-deficient strain of *E. coli* as a transformable host.

Since transformation of *E. coli* is an essential step in many cloning experiments, it is desirable that it be as efficient as possible. Several groups of workers have examined the factors affecting the efficiency of transformation. It has been found that *E. coli* cells and plasmid DNA interact productively in an environment of calcium ions and low temperature (0–5°C), and that a subsequent heat shock (37–45°C) is important, but not strictly required. Several other factors, especially the inclusion of metal ions in

addition to calcium, have been shown to stimulate the process.

A very simple, moderately efficient transformation procedure for use with *E. coli* involves resuspending log-phase cells in ice-cold 50 mmol/l calcium chloride at about 10^{10} cells/ml and keeping them on ice for about 30 min. Plasmid DNA (0. 1 μg) is then added to a small aliquot (0.2 ml) of these now *competent* (i.e. competent for transformation) cells, and the incubation on ice continued for a further 30 min, followed by a heat shock of 2 min at 42°C. The cells are then usually transferred to nutrient medium and incubated for some time (30 min to 1 h) to allow phenotypic properties conferred by the plasmid to be expressed, e.g. antibiotic resistance commonly used as a selectable marker for plasmid-containing cells. (This so-called *phenotypic lag* may not need to be taken into consideration with high-level ampicillin resistance. With this marker, significant resistance builds up very rapidly, and ampicillin exerts its effect on cell-wall biosynthesis only in cells which have progressed into active growth.) Finally the cells are plated out on selective medium. Just why such a transformation procedure is effective is not fully understood (Huang & Reusch 1995). The calcium chloride affects the cell wall and may also be responsible for binding DNA to the cell surface. The actual uptake of DNA is stimulated by the brief heat shock.

Hanahan (1983) has re-examined factors that affect the efficiency of transformation, and has devised a set of conditions for optimal efficiency (expressed as transformants per μg plasmid DNA) applicable to most *E. coli* K12 strains. Typically, efficiencies of 10^7 to 10^9 transformants/μg can be achieved depending on the strain of *E. coli* and the method used (Liu & Rashidbaigi 1990). Ideally, one wishes to make a large batch of competent cells and store them frozen for future use. Unfortunately, competent cells made by the Hanahan procedure rapidly lose their competence on storage. Inoue *et al.* (1990) have optimized the conditions for the preparation of competent cells. Not only could they store cells for up to 40 days at −70°C while retaining efficiencies of $1–5 \times 10^9$ cfu/μg, but competence was affected only minimally by salts in the DNA preparation.

There are many enzymic activities in *E. coli* which can destroy incoming DNA from non-homologous sources (see Chapter 3) and reduce the transformation efficiency. Large DNAs transform less efficiently, on a molar basis, than small DNAs. Even with such improved transformation procedures, certain potential gene-cloning experiments requiring large numbers of clones are not reliable. One approach which can be used to circumvent the problem of low transformation efficiencies is to package recombinant DNA into virus particles *in vitro*. A particular form of this approach, the use of cosmids, is described in detail in Chapter 5. Another approach is electroporation, which is described below.

Electroporation

A rapid and simple technique for introducing cloned genes into a wide variety of microbial, plant and animal cells, including *E. coli*, is electroporation. This technique depends on the original observation by Zimmerman & Vienken (1983) that high-voltage electric pulses can induce cell plasma membranes to fuse. Subsequently it was found that, when subjected to electric shock, the cells take up exogenous DNA from the suspending solution. A proportion of these cells become stably transformed and can be selected if a suitable marker gene is carried on the transforming DNA. Many different factors affect the efficiency of electroporation, including temperature, various electric-field parameters (voltage, resistance and capacitance), topological form of the DNA, and various host-cell factors (genetic background, growth conditions and post-pulse treatment). Some of these factors have been reviewed by Hanahan *et al.* (1991).

With *E. coli*, electroporation has been found to give plasmid transformation efficiencies (10^9 cfu/μg DNA) comparable with the best $CaC1_2$ methods (Dower *et al.* 1988). More recently, Zhu and Dean (1999) have reported 10-fold higher transformation efficiencies with plasmids (9×10^9 transformants/μg) by co-precipitating the DNA with transfer RNA (tRNA) prior to electroporation. With conventional $CaCl_2$-mediated transformation, the efficiency falls off rapidly as the size of the DNA molecule increases and is almost negligible when the size exceeds 50 kb. While size also affects the efficiency of electroporation (Sheng *et al.* 1995), it is possible to get transformation efficiencies of 10^6 cfu/μg DNA with molecules

as big as 240 kb. Molecules three to four times this size also can be electroporated successfully. This is important because much of the work on mapping and sequencing of genomes demands the ability to handle large fragments of DNA (see p. 64 and p. 126).

Transformation of other organisms

Although *E. coli* often remains the host organism of choice for cloning experiments, many other hosts are now used, and with them transformation may still be a critical step. In the case of Gram-positive bacteria, the two most important groups of organisms are *Bacillus* spp. and actinomycetes. That *B. subtilis* is naturally competent for transformation has been known for a long time and hence the genetics of this organism are fairly advanced. For this reason *B. subtilis* is a particularly attractive alternative prokaryotic cloning host. The significant features of transformation with this organism are detailed in Chapter 8. Of particular relevance here is that it is possible to transform protoplasts of *B. subtilis*, a technique which leads to improved transformation frequencies. A similar technique is used to transform actinomycetes, and recently it has been shown that the frequency can be increased considerably by first entrapping the DNA in liposomes, which then fuse with the host-cell membrane.

In later chapters we discuss ways, including electroporation, in which cloned DNA can be introduced into eukaryotic cells. With animal cells there is no great problem as only the membrane has to be crossed. In the case of yeast, protoplasts are required (Hinnen *et al.* 1978). With higher plants one strategy that has been adopted is either to package the DNA in a plant virus or to use a bacterial plant pathogen as the donor. It has also been shown that protoplasts prepared from plant cells are competent for transformation. A further remarkable approach that has been demonstrated with plants and animals (Klein & Fitzpatrick-McElligott 1993) is the use of microprojectiles shot from a gun (p. 238).

Animal cells, and protoplasts of yeast, plant and bacterial cells are susceptible to transformation by liposomes (Deshayes *et al.* 1985). A simple transformation system has been developed which makes use of liposomes prepared from a cationic lipid (Felgner *et al.* 1987). Small unilamellar (single-bilayer) vesicles are produced. DNA in solution spontaneously and efficiently complexes with these liposomes (in contrast to previously employed liposome encapsidation procedures involving non-ionic lipids). The positively charged liposomes not only complex with DNA, but also bind to cultured animal cells and are efficient in transforming them, probably by fusion with the plasma membrane. The use of liposomes as a transformation or transfection system is called *lipofection*.

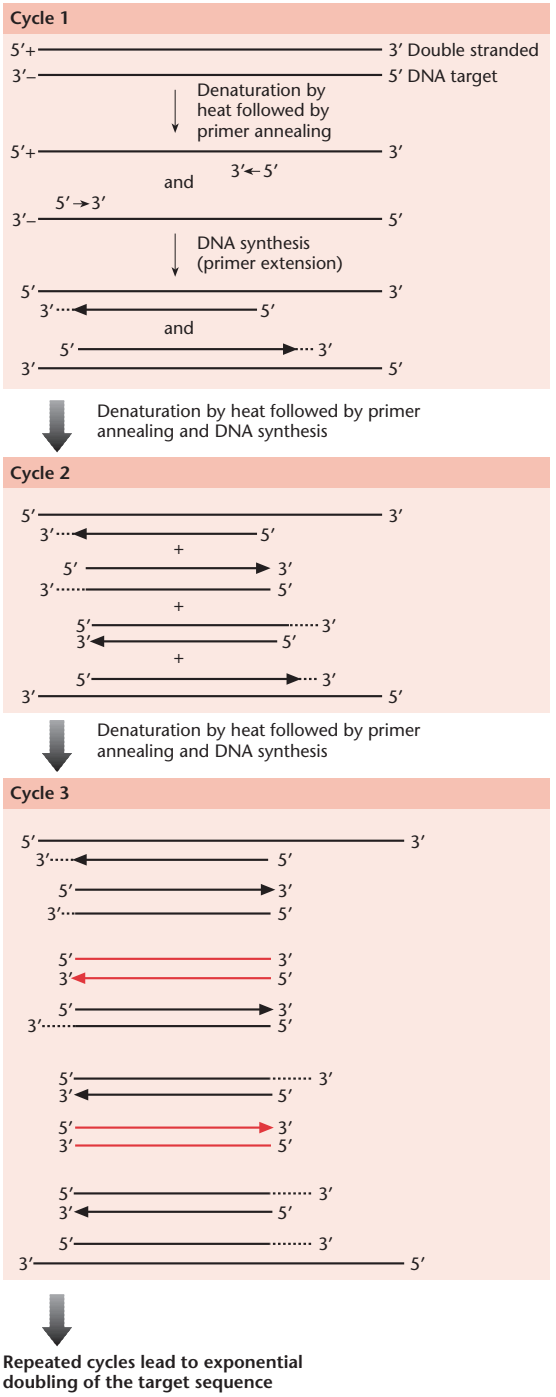
The polymerase chain reaction (PCR)

The impact of the PCR upon molecular biology has been profound. The reaction is easily performed, and leads to the amplification of specific DNA sequences by an enormous factor. From a simple basic principle, many variations have been developed with applications throughout gene technology (Erlich 1989, Innis *et al.* 1990). Very importantly, the PCR has revolutionized prenatal diagnosis by allowing tests to be performed using small samples of fetal tissue. In forensic science, the enormous sensitivity of PCR-based procedures is exploited in DNA profiling; following the publicity surrounding *Jurassic Park*, virtually everyone is aware of potential applications in palaeontology and archaeology. Many other processes have been described which should produce equivalent results to a PCR (for review, see Landegran 1996) but as yet none has found widespread use.

In many applications of the PCR to gene manipulation, the enormous amplification is secondary to the aim of altering the amplified sequence. This often involves incorporating extra sequences at the ends of the amplified DNA. In this section we shall consider only the amplification process. The applications of the PCR will be described in appropriate places.

Basic reaction

First we need to consider the basic PCR. The principle is illustrated in Fig. 2.7. The PCR involves two oligonucleotide primers, 17–30 nucleotides in length, which flank the DNA sequence that is to be amplified. The primers hybridize to opposite strands of the DNA after it has been denatured, and are



orientated so that DNA synthesis by the polymerase proceeds through the region between the two primers. The extension reactions create two double-stranded target regions, each of which can again be denatured ready for a second cycle of hybridization and extension. The third cycle produces two double-stranded molecules that comprise precisely the target region in double-stranded form. By repeated cycles of heat denaturation, primer hybridization and extension, there follows a rapid exponential accumulation of the specific target fragment of DNA. After 22 cycles, an amplification of about 10^6-fold is expected (Fig. 2.8), and amplifications of this order are actually attained in practice.

In the original description of the PCR method (Mullis & Faloona 1987, Saiki *et al.* 1988, Mullis 1990), Klenow DNA polymerase was used and, because of the heat-denaturation step, fresh enzyme had to be added during each cycle. A breakthrough came with the introduction of *Taq* DNA polymerase (Lawyer *et al.* 1989) from the thermophilic bacterium *Thermus aquaticus*. The *Taq* DNA polymerase is resistant to high temperatures and so does not need to be replenished during the PCR (Erlich *et al.* 1988, Sakai *et al.* 1988). Furthermore, by enabling the extension reaction to be performed at higher temperatures, the specificity of the primer annealing is not compromised. As a consequence of employing the heat-resistant enzyme, the PCR could be automated very simply by placing the assembled reaction in a heating block with a suitable thermal cycling programme (see Box 2.3).

Fig. 2.7 (*left*) The polymerase chain reaction. In cycle 1 two primers anneal to denatured DNA at opposite sides of the target region, and are extended by DNA polymerase to give new strands of variable length. In cycle 2, the original strands and the new strands from cycle 1 are separated, yielding a total of four primer sites with which primers anneal. The primers that are hybridized to the new strands from cycle 1 are extended by polymerase as far as the end of the template, leading to a precise copy of the target region. In cycle 3, double-stranded DNA molecules are produced (highlighted in colour) that are precisely identical to the target region. Further cycles lead to exponential doubling of the target region. The original DNA strands and the variably extended strands become negligible after the exponential increase of target fragments.

Cycle number	Number of double-stranded target molecules
1	0
2	0
3	2
4	4
5	8
6	16
7	32
8	64
9	128
10	256
11	512
12	1024
13	2048
14	4096
15	8192
16	16,384
17	32,768
18	65,536
19	131,072
20	262,144
21	524,288
22	1,048,576
23	2,097,152
24	4,194,304
25	8,388,608
26	16,777,216
27	33,554,432
28	67,108,864
29	134,217,728
30	268,435,456

Fig. 2.8 Theoretical PCR amplification of a target fragment with increasing number of cycles.

Recent developments have sought to minimize amplification times. Such systems have used small reaction volumes in glass capillaries to give large surface area-to-volume ratios. This results in almost instantaneous temperature equilibration and minimal annealing and denaturation times. This, accompanied by temperature ramp rates of 10–20°C/s, made possible by the use of turbulent forced hot-air systems to heat the sample, results in an amplification reaction completed in tens of minutes.

While the PCR is simple in concept, practically there are a large number of variables which can influence the outcome of the reaction. This is especially important when the method is being used with rare samples of starting material or if the end result has diagnostic or forensic implications. For a detailed analysis of the factors affecting the PCR, the reader should consult McDowell (1999). There are many substances present in natural samples (e.g. blood, faeces, environmental materials) which can interfere with the PCR, and ways of eliminating them have been reviewed by Bickley and Hopkins (1999).

RT-PCR

The thermostable polymerase used in the basic PCR requires a DNA template and hence is limited to the amplification of DNA samples. There are numerous instances in which the amplification of RNA would be preferred. For example, in analyses involving the diffierential expression of genes in tissues during development or the cloning of DNA derived from an mRNA (complementary DNA or *cDNA*), particularly a rare mRNA. In order to apply PCR methodology to the study of RNA, the RNA sample must first be reverse-transcribed to cDNA to provide the necessary DNA template for the thermostable polymerase. This process is called reverse transcription (RT), hence the name RT-PCR.

Avian myeloblastosis virus (AMV) or Moloney murine leukaemia virus (MuLV) reverse transcriptases are generally used to produce a DNA copy of the RNA template. Various strategies can be adopted for first-strand cDNA synthesis (Fig. 2.9).

Long accurate PCR (LA-PCR)

Amplification of long DNA fragments is desirable for numerous applications of gene manipulation. The basic PCR works well when small fragments are amplified. The efficiency of amplification and therefore the yield of amplified fragments decrease significantly as the size of the amplicon increases over 5 kb. This decrease in yield of longer amplified fragments is attributable to partial synthesis across the desired sequence, which is not a suitable substrate for the subsequent cycles. This is demonstrated by the presence of smeared, as opposed to discrete, bands on a gel.

Barnes (1994) and Cheng *et al.* (1994) examined the factors affecting the thermostable polymerization across larger regions of DNA and identified key variables affecting the yield of longer PCR fragments. Most significant of these was the absence of a 3′–5′ exonuclease (proofreading) activity in *Taq* polymerase. Presumably, when the *Taq* polymerase misincorporates a dNTP, subsequent extension of the strand either proceeds very slowly or stops completely. To overcome this problem, a second

Box 2.3 The polymerase chain reaction achieves enormous amplifications, of specific target sequence, very simply

The reaction is assembled in a single tube, and then placed in a thermal cycler (a programmable heating/cooling block), as described below.

A typical PCR for amplifying a human genomic DNA sequence has the following composition. The reaction volume is 100 µl.

Input genomic DNA, 0.1–1 µg
Primer 1, 20 pmol
Primer 2, 20 pmol
20 mmol/l Tris-HCl, pH 8.3 (at 20°C)
1.5 mmol/l magnesium chloride
25 mmol/l potassium chloride
50 mmol/l each deoxynucleoside triphosphate (dATP, dCTP, dGTP, dTTP)
2 units *Taq* DNA polymerase

A layer of mineral oil is placed over the reaction mix to prevent evaporation.

The reaction is cycled 25–35 times, with the following temperature programme:

Denaturation	94°C, 0.5 min
Primer annealing	55°C,1.5 min
Extension	72°C, 1 min

Typically, the reaction takes some 2–3 h overall.

Notes:

- The optimal temperature for the annealing step will depend upon the primers used.
- The pH of the Tris-HCl buffer decreases markedly with increasing temperature. The actual pH varies between about 6.8 and 7.8 during the thermal cycle.
- The time taken for each cycle is considerably longer than 3 min (0.5 + 1.5 + 1 min), depending upon the rates of heating and cooling between steps, but can be reduced considerably by using turbo systems (p. 21).
- The standard PCR does not efficiently amplify sequences much longer than about 3 kb.

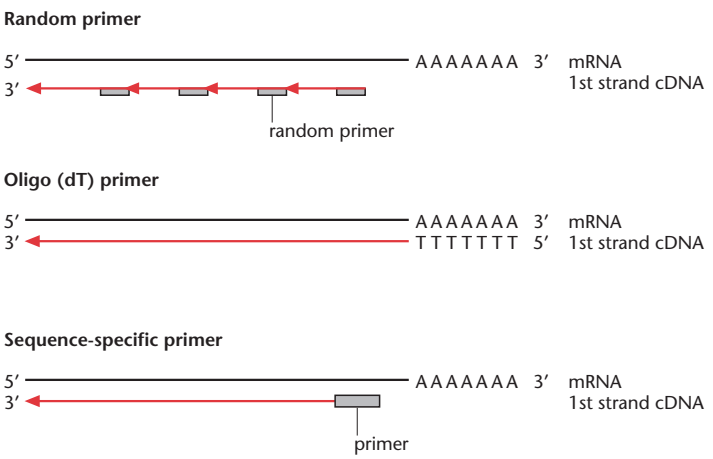


Fig. 2.9 Three strategies for synthesis of first-strand cDNA. (a) Random primer; (b) oligo (dT) primer; (c) sequence-specific primer.

thermostable polymerase with proofreading capability is added. Thermostable DNA polymerases with proofreading capabilities are listed in Table 2.1.

Key factors affecting the PCR

The specificity of the PCR depends crucially upon the primers. The following factors are important in choosing effective primers.

- Primers should be 17 to 30 nucleotides in length.
- A GC content of about 50% is ideal. For primers with a low GC content, it is desirable to choose a long primer so as to avoid a low melting temperature.
- Sequences with long runs (i.e. more than three or four) of a single nucleotide should be avoided.
- Primers with significant secondary structure are undesirable.

Table 2.1 Sources of thermostable DNA polymerases with proofreading (3′–5′ exonuclease) activity.

DNA polymerase	Source
Tma	*Thermotoga maritima*
Deep Vent™	*Pyrococcus* sp.
Tli	*Thermococcus litoralis*
Pfu	*Pyrococcus furiosus*
Pwo	*Pyrococcus woesi*

• There should be no complementarity between the two primers. The great majority of primers which conform with these guidelines can be made to work, although not all comparable primer sets are equally effective even under optimized conditions.

In carrying out a PCR it is usual to employ a *hot-start* protocol. This entails adding the DNA polymerase after the heat-denaturation step of the first cycle, the addition taking place at a temperature at or above the annealing temperature and just prior to the annealing step of the first cycle. The hot start overcomes the problem that would arise if the DNA polymerase were added to complete the assembly of the PCR reaction mixture at a relatively low temperature. At low temperature, below the desired hybridization temperature for the primer (typically in the region 45–60°C), mismatched primers will form and may be extended somewhat by the polymerase. Once extended, the mismatched primer is stabilized at the unintended position. Having been incorporated into the extended DNA during the first cycle, the primer will hybridize efficiently in subsequent cycles and hence may cause the amplification of a spurious product.

Alternatives to the hot-start protocol include the use of *Taq* polymerase antibodies, which are inactivated as the temperature rises (Taylor & Logan 1995), and AmpliTaq Gold™, a modified *Taq* polymerase that is inactive until heated to 95°C (Birch 1996). Yet another means of inactivating *Taq* DNA polymerase at ambient temperatures is the SELEX method (systematic evolution of ligands by exponential enrichment). Here the polymerase is reversibly inactivated by the binding of nanomolar amounts of a 70-mer, which is itself a poor polymerase substrate and should not interfere with the amplification primers (Dang & Jayasena 1996).

In order to minimize further the amplification of spurious products, the strategy of *nested primers* may be deployed. Here the products of an initial PCR amplification are used to seed a second PCR amplification, in which one or both primers are located internally with respect to the primers of the first PCR. Since there is little chance of the spurious products containing sequences capable of hybridizing with the second primer set, the PCR with these nested primers selectively amplifies the sought-after DNA.

As noted above, the *Taq* DNA polymerase lacks a 3′–5′ proofreading exonuclease. This lack appears to contribute to errors during PCR amplification due to misincorporation of nucleotides (Eckert & Kunkel 1990). Partly to overcome this problem, other thermostable DNA polymerases with improved fidelity have been sought, although the *Taq* DNA polymerase remains the most widely used enzyme for PCR. In certain applications, especially where amplified DNA is cloned, it is important to check the nucleotide sequence of the cloned product to reveal any mutations that may have occurred during the PCR. The fidelity of the amplification reaction can be assessed by cloning, sequencing and comparing several independently amplified molecules.

Real-time quantitative PCR

There are many applications of the PCR where it would be advantageous to be able to quantify the amount of starting material. Theoretically, there is a quantitative relationship between the amount of starting material (target sequence) and the amount of PCR product at any given cycle. In practice, replicate reactions yield different amounts of product, making quantitation unreliable. Higuchi *et al.* (1992, 1993) pioneered the use of ethidium bromide to quantify PCR products as they accumulate. Amplification produces increasing amounts of double-stranded DNA, which binds ethidium bromide, resulting in an increase in fluorescence. By plotting the increase in fluorescence versus cycle number it is possible to analyse the PCR kinetics in real time. This is much more satisfactory than analysing product accumulation after a fixed number of cycles.

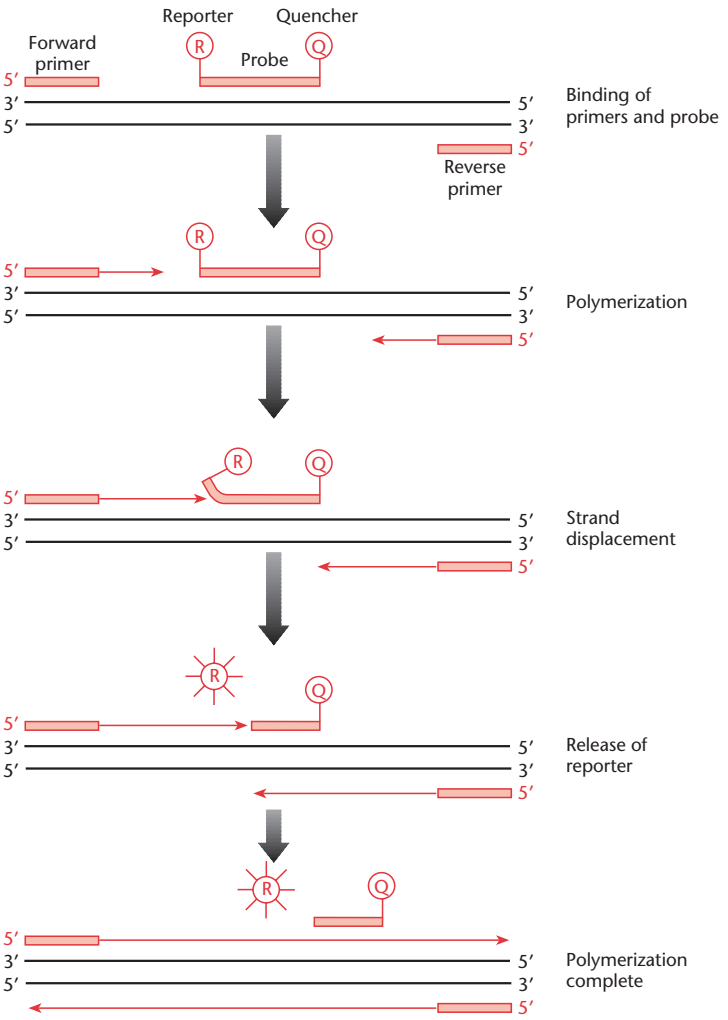


Fig. 2.10 Real-time quantitative PCR. See text for details.

The principal drawback to the use of ethidium bromide is that both specific and non-specific products generate a signal. This can be overcome by the use of probe-based methods for assaying product accumulation (Livak *et al.* 1995). The probes used are oligonucleotides with a reporter fluorescent dye attached at the 5′ end and a quencher dye at the 3′ end. While the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter dye. If the target sequence is present, the probe anneals downstream from one of the primer sites. As the primer is extended, the probe is cleaved by the 5′ nuclease activity of the *Taq* polymerase (Fig. 2.10). This cleavage of the probe separates the reporter dye from the quencher dye, thereby increasing the reporter-dye signal. Cleavage removes the probe from the target strand, allowing primer extension to continue to the end of the template strand. Additional reporter-dye molecules are cleaved from their respective probes with each cycle, effecting an increase in fluorescence intensity proportional to the amount of amplicon produced.

Instrumentation has been developed which combines thermal cycling with measurement of fluorescence, thereby enabling the progress of the PCR to be monitored in real time. This revolutionizes

the way one approaches PCR-based quantitation of DNA. Reactions are characterized by the point in time during cycling when amplification of a product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the target, the sooner a significant increase in fluorescence is noted. Quantitation of the amount of target in unknown samples is achieved by preparing a standard curve, using different starting copy numbers of the target sequence.