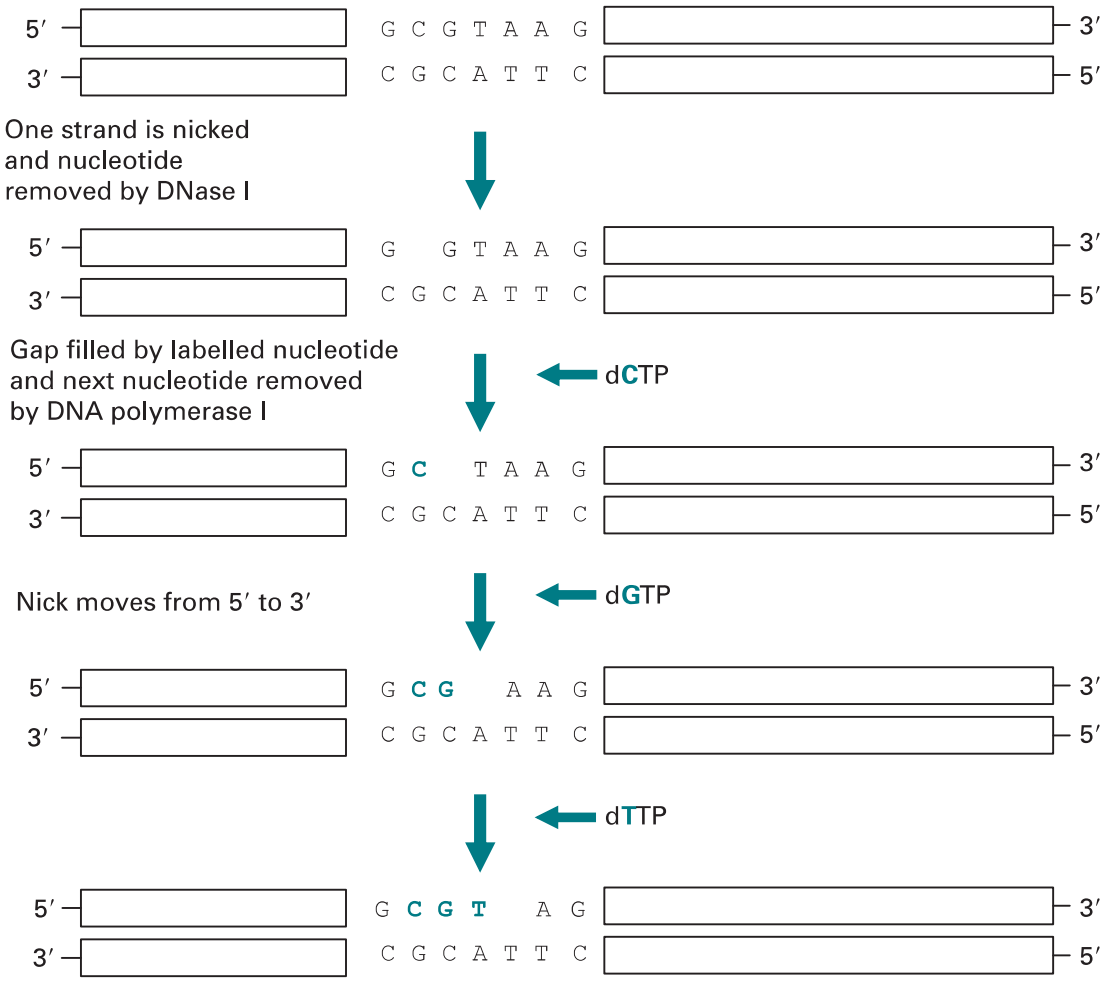

Fig. 5.31 Nick translation. The removal of nucleotides and their subsequent replacement with labelled nucleotides by DNA polymerase I increase the label in the gene probe as nick translation proceeds.

and a quencher molecule at the other. The oligonucleotide has a stem–loop structure where the stems place the fluorophore and quencher in close proximity. The loop structure is designed to be complementary to the target sequence. When the stem–loop structure is formed the fluorophore is quenched by Förster or fluorescence resonance energy transfer (FRET), i.e. the energy is transferred from the fluorophore to the quencher and given off as heat. The elegance of these types of probe lies in the fact that upon hybridisation to a target sequence the stem and loop move apart, the quenching is then lost and emission of light occurs from the fluorophore upon excitation. These types of probe have also been used to detect nucleic acid amplification system products such as the polymerase chain reaction (PCR) and have the advantage that it is unnecessary to remove the unhybridised probes.

5.10 THE POLYMERASE CHAIN REACTION (PCR)

5.10.1 Basic concept of the PCR

The polymerase chain reaction or PCR is one of the mainstays of molecular biology. One of the reasons for the wide adoption of the PCR is the elegant simplicity of the

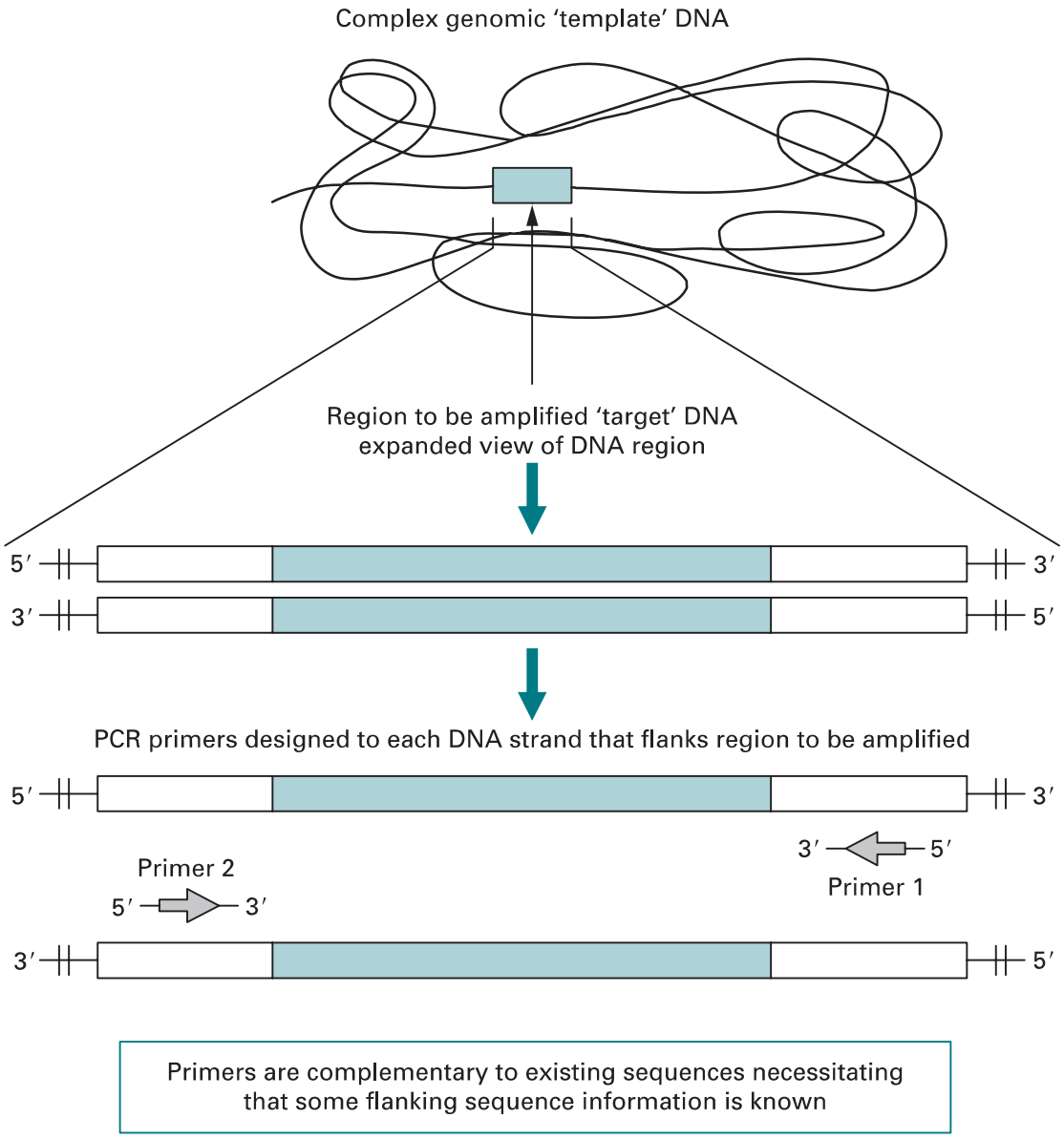

Fig. 5.32 The location of polymerase chain reaction (PCR) primers. PCR primers designed for sequences adjacent to the region to be amplified allow a region of DNA (e.g. a gene) to be amplified from a complex starting material of genomic template DNA.

reaction and relative ease of the practical manipulation steps. Indeed combined with the relevant bioinformatics resources for its design and for determination of the required experimental conditions it provides a rapid means for DNA identification and analysis. It has opened up the investigation of cellular and molecular processes to those outside the field of molecular biology.

The PCR is used to amplify a precise fragment of DNA from a complex mixture of starting material usually termed the **template DNA** and in many cases requires little DNA purification. It does require the knowledge of some DNA sequence information which flanks the fragment of DNA to be amplified (**target DNA**). From this information two oligonucleotide primers may be chemically synthesised each complementary to a stretch of DNA to the 3′ side of the target DNA, one oligonucleotide for each of the two DNA strands (Fig. 5.32). It may be thought of as a technique

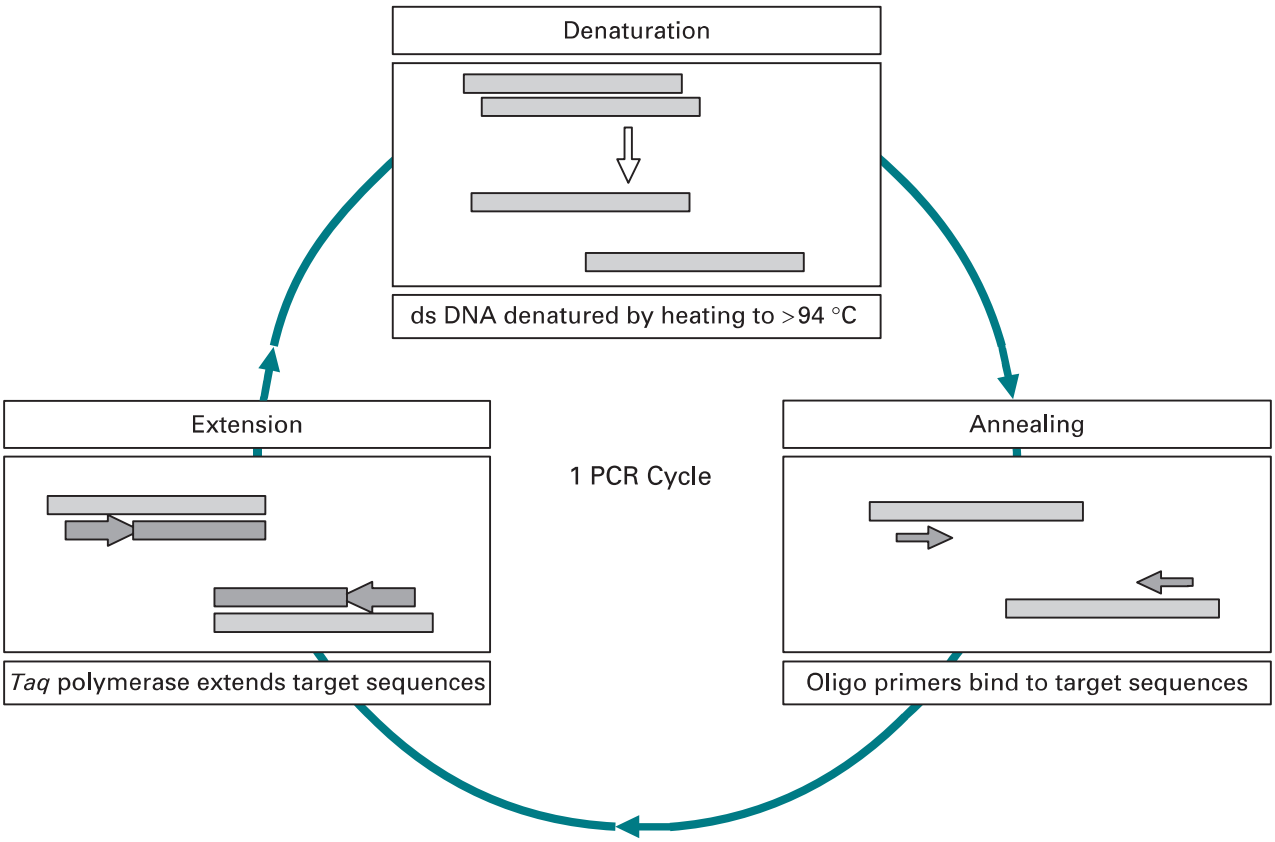

Fig. 5.33 A simplified scheme of one PCR cycle that involves denaturation, annealing and extension. ds, double-stranded.

analogous to the DNA replication process that takes place in cells since the outcome is the same: the generation of new complementary DNA stretches based upon the existing ones. It is also a technique that has replaced, in many cases, the traditional DNA cloning methods since it fulfils the same function, the production of large amounts of DNA from limited starting material; however, this is achieved in a fraction of the time needed to clone a DNA fragment (Chapter 6). Although not without its drawbacks the PCR is a remarkable development which is changing the approach of many scientists to the analysis of nucleic acids and continues to have a profound impact on core biosciences and biotechnology.

5.10.2 Stages in the PCR

The PCR consists of three defined sets of times and temperatures termed steps: (i) denaturation, (ii) annealing and (iii) extension. Each of these steps is repeated 30–40 times, termed cycles (Fig. 5.33). In the first cycle the double-stranded template DNA is (i) denatured by heating the reaction to above 90 °C. Within the complex DNA the region to be specifically amplified (target) is made accessible. The temperature is then cooled to 40–60 °C. The precise temperature is critical and each PCR system has to be defined and optimised. One useful technique for optimisation is touchdown PCR where a programmable cycler is used to incrementally decrease the annealing temperature until the optimum is derived. Reactions that are not optimised may give rise to other DNA products in addition to the specific target or may not produce any

amplified products at all. The annealing step allows the hybridisation of the two oligonucleotide primers, which are present in excess, to bind to their complementary sites that flank the target DNA. The annealed oligonucleotides act as primers for DNA synthesis, since they provide a free 3′ hydroxyl group for DNA polymerase. The DNA synthesis step is termed extension and is carried out by a thermostable DNA polymerase, most commonly *Taq* DNA polymerase.

DNA synthesis proceeds from both of the primers until the new strands have been extended along and beyond the target DNA to be amplified. It is important to note that, since the new strands extend beyond the target DNA, they will contain a region near their 3′ ends that is complementary to the other primer. Thus, if another round of DNA synthesis is allowed to take place, not only the original strands will be used as templates but also the new strands. Most interestingly, the products obtained from the new strands will have a precise length, delimited exactly by the two regions complementary to the primers. As the system is taken through successive cycles of denaturation, annealing and extension all the new strands will act as templates and so there will be an exponential increase in the amount of DNA produced. The net effect is to selectively amplify the target DNA and the primer regions flanking it (Fig. 5.34).

One problem with early PCR reactions was that the temperature needed to denature the DNA also denatured the DNA polymerase. However the availability of a thermostable DNA polymerase enzyme isolated from the thermophilic bacterium *Thermus aquaticus* found in hot springs provided the means to automate the reaction. *Taq* DNA polymerase has a temperature optimum of 72 °C and survives prolonged exposure to temperatures as high as 96 °C and so is still active after each of the denaturation steps. The widespread utility of the technique is also due to the ability to automate the reaction and as such many thermal cyclers have been produced in which it is possible to program in the temperatures and times for a particular PCR reaction.

5.10.3 PCR primer design and bioinformatics

The specificity of the PCR lies in the design of the two oligonucleotide primers. These have to not only be complementary to sequences flanking the target DNA but also must not be self-complementary or bind each other to form dimers since both prevent DNA amplification. They also have to be matched in their GC content and have similar annealing temperatures. The increasing use of bioinformatics resources such as Oligo, Generunner and Genefisher in the design of primers makes the design and the selection of reaction conditions much more straightforward. These resources allow the sequences to be amplified, primer length, product size, GC content, etc. to be input and, following analysis, provide a choice of matched primer sequences. Indeed the initial selection and design of primers without the aid of bioinformatics would now be unnecessarily time-consuming.

It is also possible to design primers with additional sequences at their 5′ end such as restriction endonuclease target sites or promoter sequences. However modifications such as these require that the annealing conditions be altered to compensate for the areas of non-homology in the primers. A number of PCR methods have been developed where either one of the primers or both are random. This gives rise to

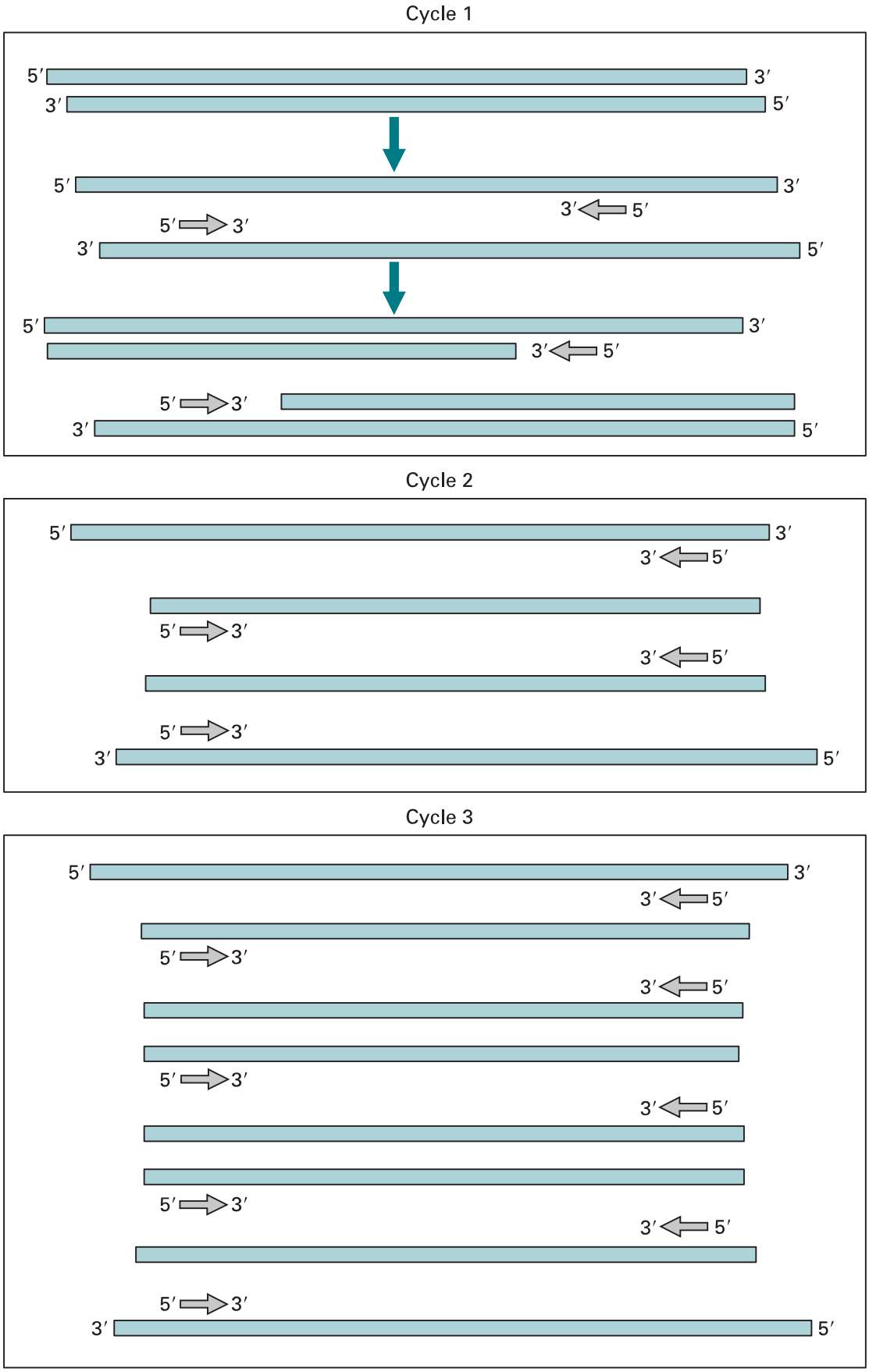

Fig. 5.34 Three cycles in the PCR. As the number of cycles in the PCR increases, the DNA strands that are synthesised and become available as templates are delimited by the ends of the primers. Thus specific amplification of the desired target sequence flanked by the primers is achieved. Primers are denoted as 5′ to 3′.

arbitrary priming in genomic templates but interestingly may give rise to discrete banding patterns when analysed by gel electrophoresis. In many cases this technique may be used reproducibly to identify a particular organism or species. This is sometimes referred to as random amplified polymorphic DNA (RAPD) and has been used successfully in the detection and differentiation of a number of pathogenic strains of bacteria. In addition primers can now be synthesised with a variety of labels such as fluorophores bound to them allowing easier detection and quantitation using techniques such as qPCR (Section 5.10.7).

5.10.4 PCR amplification templates

DNA from a variety of sources may be used as the initial source of amplification templates. It is also a highly sensitive technique and requires only one or two molecules for successful amplification. Unlike many manipulation methods used in current molecular biology the PCR technique is sensitive enough to require very little template preparation. The extraction from many prokaryotic and eukaryotic cells may involve a simple boiling step. Indeed the components of many extraction techniques such as SDS and proteinase K may adversely affect the PCR. The PCR may also be used to amplify RNA, a process termed RT–PCR (reverse transcriptase–PCR). Initially a reverse transcription reaction which converts the RNA to cDNA is carried out (Section 6.2.5). This reaction normally involves the use of the enzyme reverse transcriptase although some thermostable DNA polymerases used in the PCR such as *Tth* have a reverse transcriptase activity under certain buffer conditions. This allows mRNA transcription products to be effectively analysed. It may also be used to differentiate latent viruses (detected by standard PCR) or active viruses which replicate and thus produce transcription products and are thus detectable by RT–PCR (Fig. 5.35). In addition the PCR may be extended to determine relative amounts of a transcription product.

5.10.5 Sensitivity of the PCR

The enormous sensitivity of the PCR system is also one of its main drawbacks since the very large degree of amplification makes the system vulnerable to contamination. Even a trace of foreign DNA, such as that even contained in dust particles, may be amplified to significant levels and may give misleading results. Hence cleanliness is paramount when carrying out PCR, and dedicated equipment and in some cases dedicated laboratories are used. It is possible that amplified products may also contaminate the PCR although this may be overcome by UV irradiation to damage already amplified products so that they cannot be used as templates. A further interesting solution is to incorporate uracil into the PCR and then treat the products with the enzyme uracil *N*-glycosylase (UNG) which degrades any PCR amplicons with incorporated uracil rendering them useless as templates. In addition most PCRs are now undertaken using hotstart. Here the reaction mixture is physically separated from the template or the enzyme: when the reaction begins mixing occurs and thus avoids any mispriming that may have arisen.

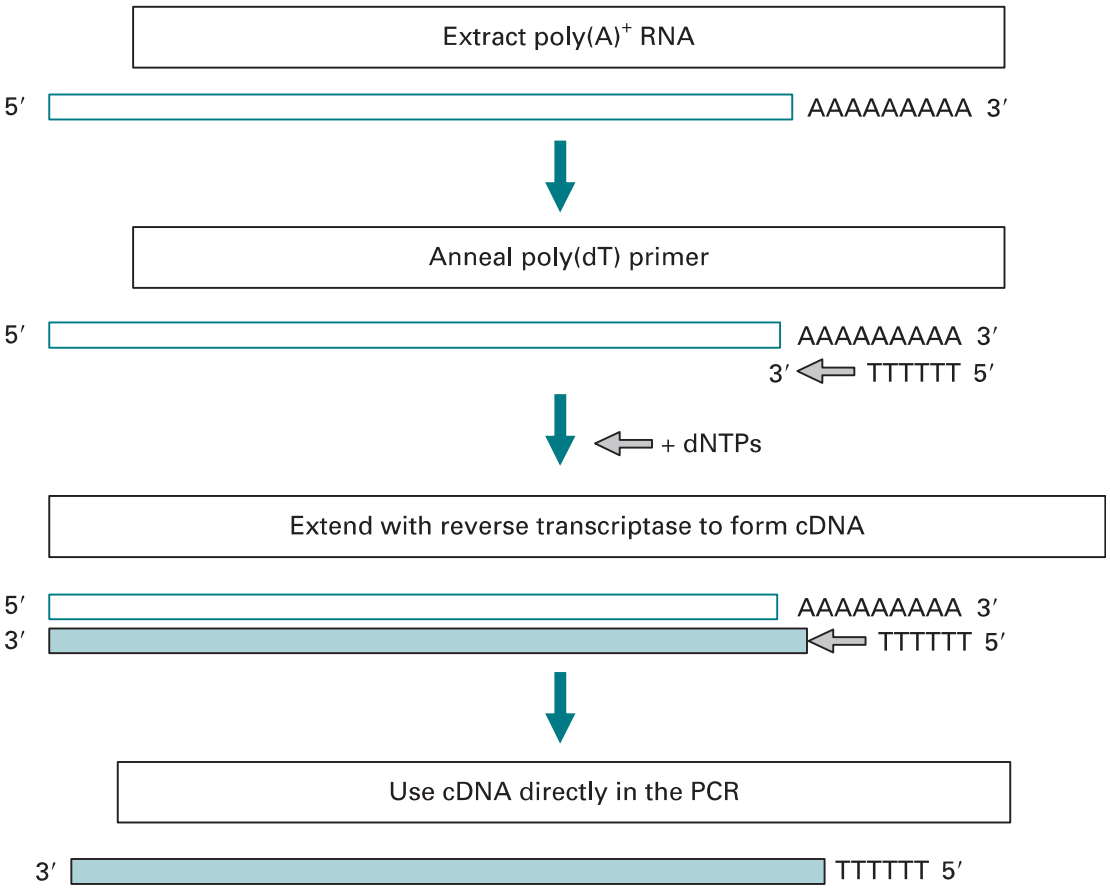

Fig. 5.35 Reverse transcriptase–PCR (RT–PCR): mRNA is converted to complementary DNA (cDNA) using the enzyme reverse transcriptase. The cDNA is then used directly in the PCR.

5.10.6 Applications of the PCR

Many traditional methods in molecular biology have now been superseded by the PCR and the applications for the technique appear to be unlimited. Some of the main techniques derived from the PCR are introduced in Chapter 6 while some of the main areas to which the PCR has been put to use are summarised in Table 5.5. The success of the PCR process has given impetus to the development of other amplification techniques that are based on either thermal cycling or non-thermal cycling (isothermal) methods. The most popular alternative to the PCR is termed the ligase chain reaction or LCR. This operates in a similar fashion to the PCR but a thermostable DNA ligase joins sets of primers together which are complementary to the target DNA. Following this a similar exponential amplification reaction takes place producing amounts of DNA that are similar to the PCR. A number of alternative amplification techniques are listed in Table 5.6.

5.10.7 Quantitative PCR (qPCR)

One of the most useful PCR applications is quantitative PCR or qPCR. This allows the PCR to be used as a means of identifying the initial concentrations of DNA or cDNA template used. Early qPCR methods involved the comparison of a standard or

Table 5.5 **Selected applications of the PCR. A number of the techniques are described in the text of Chapters 5 and 6**

Field or area of study	Application	Specific examples or uses
General molecular biology	DNA amplification	Screening gene libraries
Gene probe production	Production/labelling	Use with blots/hybridisations
RNA analysis	RT–PCR	Active latent viral infections
Forensic science	Scenes of crime	Analysis of DNA from blood
Infection/disease monitoring	Microbial detection	Strain typing/analysis RAPDs
Sequence analysis	DNA sequencing	Rapid sequencing possible
Genome mapping studies	Referencing points in genome	Sequence-tagged sites (STS)
Gene discovery	mRNA analysis	Expressed sequence tags (EST)
Genetic mutation analysis	Detection of known mutations	Screening for cystic fibrosis
Quantification analysis	Quantitative PCR	5′ Nuclease (TaqMan assay)
Genetic mutation analysis	Detection of unknown mutations	Gel-based PCR methods (DGGE)
Protein engineering	Production of novel proteins	PCR mutagenesis
Molecular archaeology	Retrospective studies	Dinosaur DNA analysis
Single-cell analysis	Sexing or cell mutation sites	Sex determination of unborn
In situ analysis	Studies on frozen sections	Localisation of DNA/RNA

Notes: RT, reverse transcriptase; RAPDs, rapid amplification polymorphic DNA; DDGE, denaturing gradient gel electrophoresis.

control DNA template amplified with separate primers at the same time as the specific target DNA. However these types of quantitation rely on the fact that all the reactions are identical and so any factors affecting this may also affect the result. The introduction of thermal cyclers that incorporate the ability to detect the accumulation of DNA through fluorescent dyes binding to the DNA has rapidly transformed this area.

In its simplist form a PCR is set up that includes a DNA-binding cyanine dye such as SYBR green. This dye binds to the major groove of double-stranded DNA but not single-stranded DNA and so as amplicons accumulate during the PCR process SYBR green binds the double-stranded DNA proportionally and fluorescence emission of the dye can be detected following excitation. Thus the accumulation of DNA amplicons can be followed in real time during the reaction run. In order to quantitate unknown DNA templates a standard dilution is prepared using DNA of known concentration. As the DNA accumulates during the early exponential phase of the reaction an arbitrary point is taken where each of the dilluted DNA samples cross. This is termed the crossing threshold on Ct value. From the various Ct values a log

Table 5.6 **Selected alternative amplification techniques to the PCR. Two broad methodologies exist that either amplify the target molecules such as DNA and RNA or detect the target and amplify a signal molecule bound to it**

Technique	Type of assay	Specific examples or uses
Target amplification methods		
Ligase chain reaction (LCR)	Non-isothermal, employs thermostable DNA ligase	Mutation detection
Nucleic acid sequence based amplification (NASBA)	Isothermal, involving use of RNA, RNase H/reverse transcriptase, and T7 DNA polymerase	Viral detection, e.g. HIV
Signal amplification methods		
Branched DNA amplification (b-DNA)	Isothermal microwell format using hybridisation or target/capture probe and signal amplification	Mutation detection

Note: HIV, human immunodeficiency virus.

graph is prepared from which an unknown concentration can be deduced. Since SYBR green and similar DNA-binding dyes are non-specific, in order to determine if a correctly sized PCR product is present most qPCR cyclers have a built-in melting curve function. This gradually increases the temperature of each tube until the double-stranded PCR product denatures or melts and allows a precise although not definitive determination of the product. Confirmation of the product is usually obtained by DNA sequencing.

5.10.8 The TaqMan system

In order to make qPCR specific a number of strategies may be employed that rely on specific hybridisation probes. One ingenious method is called the TaqMan assay or 5′ nuclease assay. Here the probe consists of an oligonucleotide labelled with a fluorescent reporter at one end of the molecule and quencher at the other end.

The PCR proceeds as normal and the oligonucleotide probe binds to the target sequence in the annealing step. As the *Taq* polymerase extends from the primer its 5′ exonuclease activity degrades the hybridisation probe and releases the reporter from the quencher. A signal is thus generated which increases in direct proportion to the number of starting molecules and fluorescence can be detected in real time as the PCR proceeds (Fig. 5.36). Although relatively expensive in comparison to other methods for determining expression levels it is simple, rapid and reliable and now in use in many research and clinical areas. Further developments in probe-based PCR systems have also been used and include scorpion probe systems, amplifluor and real-time LUX probes.

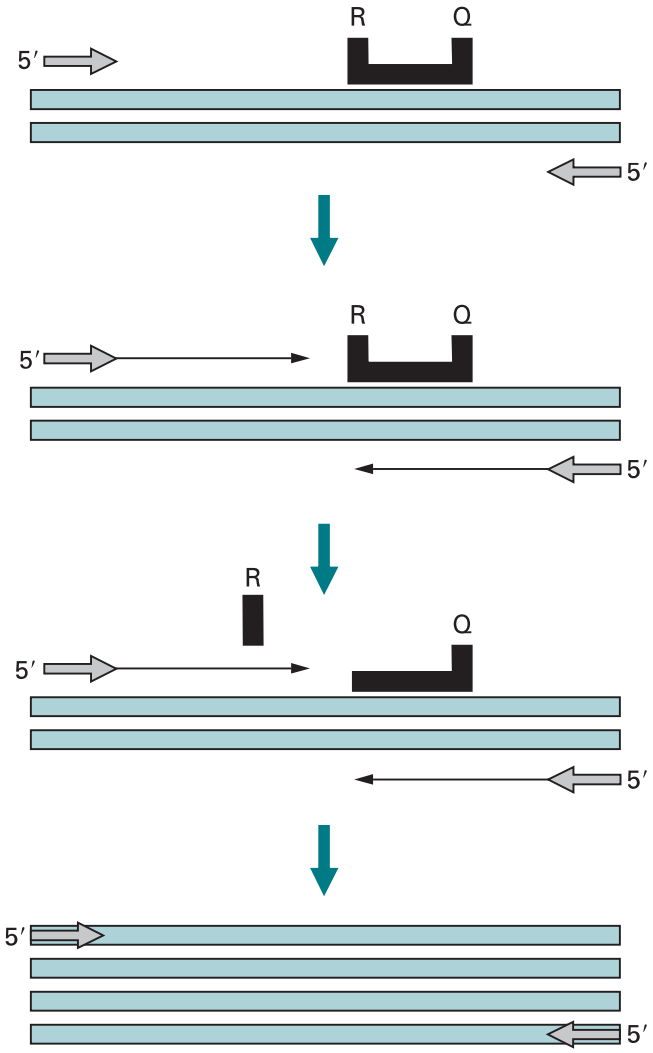

Fig. 5.36 5′ Nuclease assay (TaqMan assay). PCR is undertaken with RQ probe (reporter/quencher dye). As R–Q are in close proximity, fluorescence is quenched. During extension by *Taq* polymerase the probe is cleaved as a result of *Taq* having 5′ nuclease activity. This cleaves R–Q probe and the reporter is released. This results in detectable increase in fluorescence and allows real-time PCR detection.

5.11 NUCLEOTIDE SEQUENCING OF DNA

5.11.1 Concepts of nucleic acid sequencing

The determination of the order or sequence of bases along a length of DNA is one of the central techniques in molecular biology. Although it is now possible to derive amino acid sequence information with a degree of reliability it is frequently more convenient and rapid to analyse the DNA coding information. The precise usage of codons, information regarding mutations and polymorphisms and the identification of gene regulatory control sequences are also only possible by analysing DNA sequences. Two techniques have been developed for this, one based on an enzymatic method frequently termed Sanger sequencing after its developer, and a chemical method called Maxam and Gilbert, named for the same reason. At present Sanger