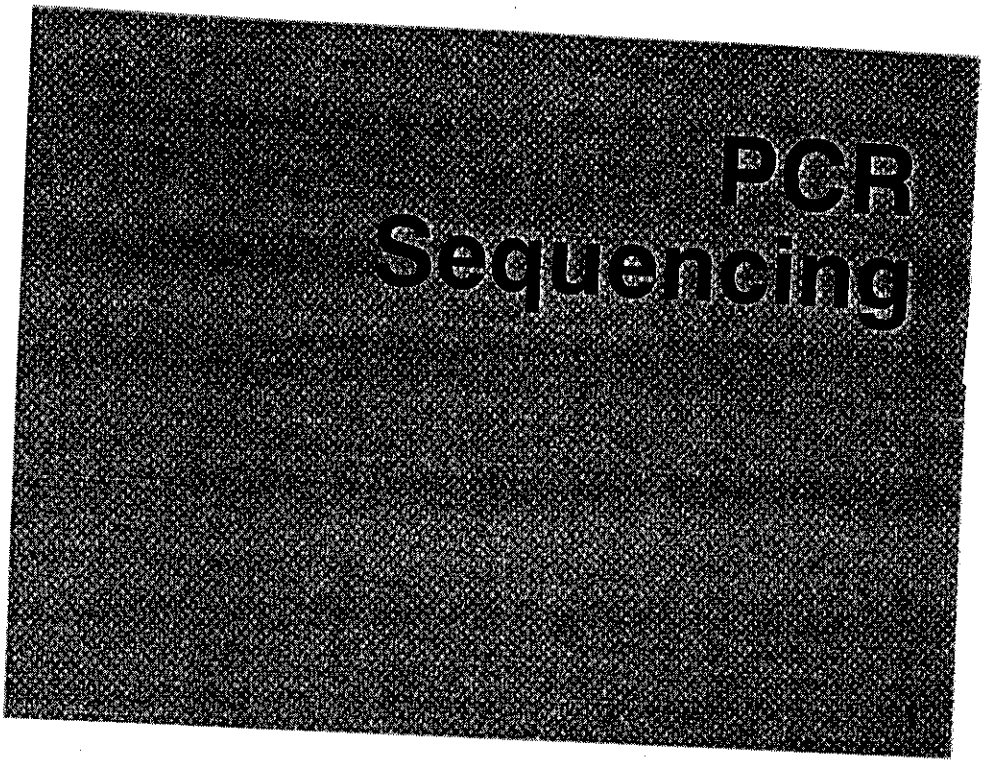

The standard DNA sequencing protocol based on chain termination using dideoxynucleotide triphosphates, first described by Sanger, has been modified by the availability of thermostable DNA polymerases. The two major advantages of using a thermostable enzyme in what is known as "cycle sequencing" over other DNA polymerases, such as Sequenase, used in standard DNA sequencing are the decreased amounts of template required to obtain a readable sequence ladder and the high temperature at which the reactions are performed. Carrying out sequencing reactions at higher temperatures facilitates the sequencing of double-strand templates, including PCR products, plasmids, lambda DNA, and cosmids, and is preferred for templates with a high G-C base content.

Two differences between cycle sequencing and a standard PCR should be noted. In cycle sequencing, only one primer is required and both dideoxynucleotide triphosphates and deoxynucleotide triphosphates are employed. Through the use of a single primer, the amplification of the template DNA is linear, as opposed to exponential. This linear amplification generates a signal corresponding to the template positions at which molecules terminate.

The DNA present in crude samples can be sequenced using a two-step process in which the DNA of interest is first PCR-amplified, followed by a cycle sequencing reaction. One of the two primers used to amplify this DNA, or a third primer that hybridizes to the PCR product, can be used for sequencing. This approach, described in detail in "Direct Sequencing of PCR-amplified DNA," bypasses the need to clone the amplified DNA prior to sequencing and is particularly useful when amplification reactions are performed with thermostable

polymerases that do not possess proofreading activity. Direc
ing of the amplified DNA guarantees that the obtained sequ
responds to that of the template, because any particular mu
"diluted" within the overall population of molecules. Alter
when a PCR product is cloned prior to sequencing, more t
clone should be sequenced, because mutations genera
nucleotide misincorporations during the PCR may be carried
clone selected. Also, when cloning a PCR product prior to sequ
the use of thermostable polymerases that have proofreading act
recommended to minimize nucleotide misincorporations.

"Cycle Sequencing" describes the preparation of different
templates for cycle sequencing. It also discusses modifications o
cycle sequencing techniques aimed at maximizing the length o
readable sequence data.

TROD

SEC
STR

Direct Sequencing of PCR-amplified DNA

Venigalla B. Rao

Department of Biology, Institute for Biomolecular Studies, The Catholic University of America, Washington, DC 20064

NTRODUCTION

The amplification of target DNA by PCR (Mullis et al. 1986) followed by the direct sequencing of amplified DNA (Gyllensten 1989) has emerged as a powerful strategy for rapid molecular genetic analysis. Using this strategy, time-consuming cloning steps can be completely bypassed and the sequence of the target DNA can be determined directly from a crude biological sample. The crude sample can be cultured cells, bacteria, or a viral preparation. Furthermore, the copy number of target DNA in the sample can be as low as one to a few molecules of genomic DNA among a vast excess of contaminating nontarget DNA (Mullis 1991; Saunders et al. 1993).

This review describes two sequencing strategies that allow the generation of the DNA sequence from almost any PCR-amplified DNA template. In addition, factors that influence the sequencing reactions are discussed, which allow the manipulation of these strategies for specific sequencing needs. For a more comprehensive analysis of the direct sequencing strategy, the reader is referred to reviews by Gyllensten (1989) and Rao (1994).

QUENASE RATEGY

This strategy consists of three steps (Fig. 1) (Tabor and Richardson 1987). In the first step, the double-stranded, PCR-amplified DNA is denatured to single strands and the sequencing primer is annealed to the complementary sequence on one of the template strands. In the second step, the annealed primer is extended by 20–80 nucleotides by DNA polymerase, incorporating multiple radioactive labels into the newly synthesized DNA. This step is performed under nonoptimal reaction conditions so that the enzyme acts in a low-processive fashion, synthesizing only short stretches of DNA. In the third step, the labeled DNA chains are extended and terminated by the incorporation of the ddNMP (Fig. 1).

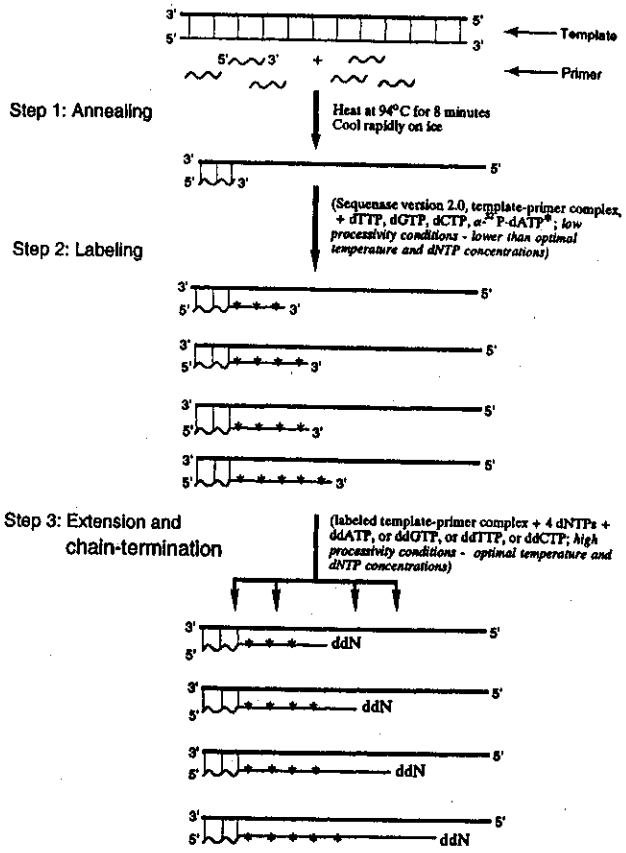

Figure 1 Schematic diagram of the Sequenase protocol. The wa... represents the primer. The template strand for sequencing of th... amplified DNA is represented as a thick line. The second st... represented as a thin line. The second strand is shown only in the temp... DNA at the top of the figure. It is not included in the rest of the figur... though it is present in the reaction mixture. The cross bars betwe... strands of the template DNA, and between the template and the pri... represent the classical Watson-Crick hydrogen bonds. The as... represent the radioactive labels incorporated as [^{32}P]dAMP into the ... synthesized DNA chains.

REAGENTS

DNA template: The PCR-amplified DNA should be first separated fr... the unused dNTPs and primers by QIAquick ion-exchange co... chromatography (QIAGEN). The concentration of recovered ... template can be estimated by agarose gel electrophoresis ... ethidium bromide staining of a small aliquot of the purified DN... comparing the intensity of the stained band with that of a stan... DNA of known concentration. Alternately, the DNA concentratio... be estimated by using a DNA dipstick (Invitrogen). A DNA concen... tion in the range of about 1 pmole is desirable to generate hi... intensity sequence ladders. However, concentrations as low ... pmole can be used.

Sequencing primer: A 20-nucleotide DNA primer synthesized by an oligonucleotide synthesizer can be used directly as a sequencing primer without any purification steps involving high-performance liquid chromatography or polyacrylamide gel electrophoresis. The concentration of the primer is estimated by spectroscopy at 260 nm (one absorbance unit is equivalent to a primer concentration of 33 μg/ml).

DNA polymerase: Sequenase, version 2.0 (USB/Amersham), is the most suitable enzyme for performing this protocol. Sequenase, version 2.0, is a genetically modified phage T7 DNA polymerase that has no $3' \rightarrow 5'$ exonuclease activity. Sequenase is active at low temperatures and incorporates multiple radioactive labels efficiently to generate high-intensity sequence ladders.

Sequencing buffer: The composition of the sequencing buffer varies with the polymerase used. For Sequenase, the buffer is 40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl.

Radiolabeled dNTP: Approximately 5 μCi of either [α-^{32}P]dATP, [α-^{33}P]dATP, or [α-^{35}S]dATP is used for each set of sequencing reactions. However, [α-^{32}P]dATP is the preferred radionucleotide for this protocol.

PROTOCOL

The following is a Sequenase protocol (USB/Amersham) modified for sequencing PCR-amplified DNA.

1. Set up the extension-termination reaction mixtures. Transfer 2.5 μl of each of the four dNTP/ddNTP mixtures to four independent tubes. Each tube receives 80 μM of each of the four dNTPs and 8 μM of either ddATP, ddTTP, ddGTP, or ddCTP. Preincubate these tubes for 5 minutes at 37°C before initiating the extension and chain-termination reactions (see step 4 below).

   For sequencing GC-rich templates, dGTP should be replaced with 7-deaza-dGTP (USB/Amersham) to overcome compression artifacts that are known to occur during sequencing gel electrophoresis (Barr et al. 1986). dGTP can also be replaced with dITP, but dITP may not be a good substrate for all DNA polymerases.

2. Mix about 1 pmole of PCR-amplified DNA, 10 pmoles of sequencing primer, 2 μl of 5x sequencing buffer, and H_2O in a total reaction volume of 10 μl. The addition of nonionic detergents such as Tween 20 and NP-40 to a final concentration of 0.5% has been reported to improve the specificity of annealing of the sequencing primer to the template (Bachmann et al. 1990). Incubate the samples in a heat block for 8 minutes at 94–96°C. Chill the tubes on ice for 1 minute (Kusukawa et al. 1990). Centrifuge in a micro-

fuge at 10,000 rpm for 10 seconds. Transfer the tubes to ice, a
immediately proceed to the next step.

Note that the above annealing conditions are applicable only
double-stranded, PCR-amplified DNA templates. For single
stranded templates, such as the products of asymmetric PCR (Gy
lensten and Erlich 1988), the gradual cooling technique
preferred. In this procedure, the above mixture is incubated for
minutes in a 65°C heat block and is gradually cooled to 30°C b
turning off the heat block.

3. To the above mixture on ice, add 2 µl of cold dNTP mix containin
0.75 µM each of dTTP, dGTP, and dCTP, and 5 µCi of [α-^{32}P]dAT
(sp. act. 3000 Ci/mmole, Amersham), 1 µl of 0.1 M dithiothreitol
H_2O, and 2 units of freshly diluted Sequenase, version 2.0, to a final
reaction volume of 15.5 µl. Incubate the reaction mixture on ice for
2 minutes to label the DNA.

4. Transfer 3.5 µl of the above mixture to each of the four
dNTP/ddNTP tubes that have been preincubated for 5 minutes at
37°C (from step 1). Allow the extension and chain-termination
reactions to proceed for 5 minutes at 37°C.

5. Stop the extension and chain-termination reactions by adding 4 µl
of Stop buffer containing 95% formamide, 20 mM EDTA, 0.05%
bromophenol blue, and 0.05% xylene cyanol FF. If the samples are
not used immediately, they can be stored frozen at –70°C for about
a week. However, it is preferable to use the samples within 2 days.
Heat the samples for 3 minutes in an 80°C heat block, and load a 2-
to 3-µl aliquot of the sample in each lane of a sequencing gel.

Important Characteristics of This Protocol

1. The labeling step is the most critical step in this protocol and should be performed under well-controlled conditions. In this step, the annealed primer is extended by only 20–80 nucleotides to incorporate multiple radioactive labels into the newly synthesized DNA. Because the Sequenase enzyme is highly processive and synthesizes several thousand nucleotides (~4000 nucleotides with version 2.0) at a stretch before dissociating from the complex, this step should be performed under low-processivity conditions such as low temperature and low concentrations of dNTPs (Tabor and Richardson 1987).

2. A common pitfall of this procedure is the appearance of either very faint sequence ladders or no sequence ladders on the final autoradiogram. However, an intense band is seen in the high-

molecular-weight position. This characteristic pattern is due to a lack of control at the labeling step. If the labeling reactions are not controlled well, the Sequenase enzyme, instead of synthesizing a short stretch of DNA, predominantly synthesizes the full-length product. Consequently, the sequence ladders representing the shorter chain-terminated products constitute a minor fraction and therefore appear as faint bands. This can be controlled by modifying the labeling conditions in such a way that the enzyme acts in a low-processive manner. These modifications include decreasing the enzyme concentration, decreasing the unlabeled dNTP concentration, reducing the reaction time, and increasing the template concentration.

3. This protocol requires the incorporation of at least one radioactive label into the newly synthesized DNA to visualize a band on the autoradiogram. Therefore, the sequence of the first nucleotide that can be determined by this protocol depends on the distance between the 3′ end of the primer and the first labeled [^{32}P]dAMP incorporated. If the template is known to be, or suspected to be, a GC-rich template, the use of [α-^{32}P]dCTP rather than [α-^{32}P]dATP is recommended.

CYCLE SEQUENCING STRATEGY

Based on the report by Murray (1989), a number of PCR-directed sequencing strategies (Lee 1991; Ruano and Kidd 1991; Rao and Saunders 1992), referred to as cycle sequencing strategies, have been developed for the direct sequencing of PCR-amplified DNA. These strategies take advantage of the powerful automated cycling capability of thermal cyclers to amplify chain-terminated sequencing products and generate high-intensity sequence ladders. Each sequencing cycle consists of three steps (Fig. 2). First, the PCR-amplified DNA is denatured to single strands. This is followed by the annealing of a ^{32}P-labeled sequencing primer (or a biotinylated primer) to the complementary sequence on one of the strands. In the final step, the annealed primer is extended and chain-terminated by a thermostable DNA polymerase. The resulting partially double-stranded chain-terminated product is then denatured in the next sequencing cycle, releasing the template strand for another round of priming reactions, while accumulating chain-terminated products in each cycle. These steps are repeated for 20–40 cycles to amplify the chain-terminated products in a linear fashion (Fig. 2).

GENTS

DNA template: As in the Sequenase protocol, the PCR-amplified DNA should be first purified by QIAquick column chromatography (QIAGEN Inc.) to remove unused dNTPs and primers. It is possible to use 1–2 µl of the unpurified PCR product directly for cycle sequenc-

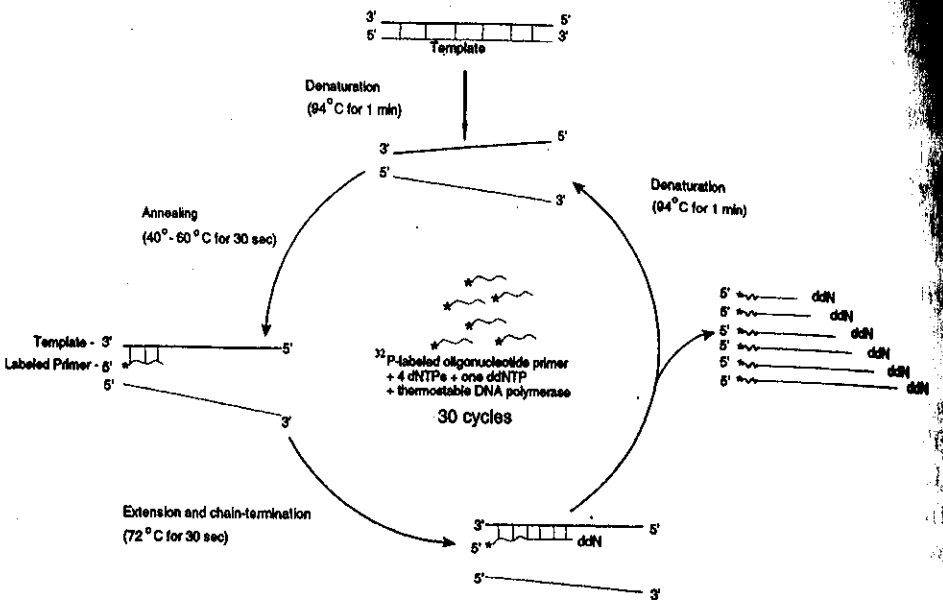

Figure 2 Schematic diagram of the cycle sequencing protocol. The wavy line represents the primer. The template strand for sequencing of the PCR-amplified DNA is represented as a thick line. The second strand is represented as a thin line. The cross bars between the strands of the template DNA, and between the template and the primer, represent the classical Watson-Crick hydrogen bonds. The asterisks represent the ^{32}P-label incorporated into the 5′ end of the primer as a result of phosphorylation by T4 polynucleotide kinase.

ing. However, the overall quality of the sequence ladders generated is not as good as with the purified DNA. Therefore, the purification of amplified DNA is recommended. When sequencing a large number of templates routinely, the purification step can be bypassed by using the recently described modified cycle sequencing strategies (Ruano and Kidd 1991; Rao and Saunders 1992). In these strategies, lower concentrations of dNTPs (10–20 μM of each dNTP) and primers (10 pmoles of each primer) are used for the amplification of the target DNA. As a result, the carryover of unused dNTPs and primers into the sequencing reactions is minimized. It is desirable to have a template DNA concentration of at least 0.1 pmole for each set of sequencing reactions to generate high-intensity sequencing ladders. Lower template concentrations can be used, but the sequence ladders generated will be of low intensity.

Sequencing primer: The sequencing primer is labeled with ^{32}P at the 5′ end using [γ-^{32}P]ATP and T4 polynucleotide kinase. The ratio of primer, kinase, and [γ-^{32}P]ATP should be maintained at an optimal level to achieve high-specific-activity labeling of the sequencing primer (see below). High specific activity of the primer, but not the to-

tal activity, is critical for generating high-intensity sequence ladders. Alternatively, it is possible to use [α-^{32}P]dATP or [α-^{33}P]dATP to label newly synthesized DNA, rather than using a 5′-labeled primer (Promega). However, this generates higher background (see below). In addition, there is a greater degree of nonuniformity in the intensity of the bands generated. For example, chain-terminated products of high molecular weight incorporate radioactive labels at severalfold higher frequency when compared to those of low molecular weight.

DNA polymerase: Any thermostable DNA polymerase that lacks 3′→ 5′ exonuclease activity can be used. *Taq* DNA polymerase (Perkin-Elmer), which has no 3′→ 5′ exonuclease activity, is the most widely used enzyme (Innis et al. 1988). Other enzymes that are used for cycle sequencing include *Pfu* polymerase (exonuclease-minus) (Stratagene), Vent polymerase (exonuclease-minus) (New England Biolabs), and *Tub* polymerase (Amersham).

PROTOCOL

The following is a basic protocol for *Taq* DNA polymerase. The same can be used as well for other thermostable polymerases, but the buffer conditions and dNTP/ddNTP ratios should be modified accordingly for each polymerase used. Many factors affect the optimal dNTP/ddNTP ratios required for generating high-intensity sequence ladders with low background (see below). Therefore, it is highly recommended that the dNTP/ddNTP ratios be optimized for a given sequencing application.

1. For labeling the primer, mix 10–15 pmoles of sequencing primer, 50 μCi of [γ-^{32}P]ATP (sp. act. 6000 Ci/mmole or >6000 Ci/mmole [NEG-035C, NEN]), 5 μl of 10x kinase buffer (70 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 5 mM dithiothreitol, or buffer supplied by the vendor), and H_2O, to a final reaction volume of 50 μl. Preincubate the reaction mixture for 5 minutes at 37°C. Add 1 μl of freshly diluted T4 polynucleotide kinase (10 units) and incubate for 30 minutes at 37°C. Add an additional 1-μl aliquot of freshly diluted kinase (10 units) and continue incubation for 30 minutes at 37°C. The removal of unincorporated [γ-^{32}P]ATP, although not critical, is recommended. This can be accomplished by gel filtration through a Biospin-10 column (Bio-Rad Laboratories). The labeled primer can be stored at –70°C for at least 2 weeks.

2. Transfer 2 μl of each of the four dNTP/ddNTP extension-termination mixtures to four tubes. Each tube receives 30 μM of each of four dNTPs and either 1.2 mM of ddATP, 1.2 mM of ddTTP, 90 μM of ddGTP, or 600 μM of ddCTP. As in the Sequenase protocol, dGTP should be replaced with 7-deaza-dGTP (USB/Amersham) for

GC-rich templates to overcome the compression artifacts that are known to occur during sequencing gel electrophoresis (Barr et al. 1986).

3. Mix 0.1–0.2 pmole of QIAGEN-purified PCR-amplified DNA, 1–2 pmoles of 5′ ^{32}P-labeled sequencing primer, 3 µl of 10x sequencing buffer (500 mM Tris-HCl, pH 8.8, and 20 mM $MgCl_2$, or buffer supplied by the vendor), and 5 units of *Taq* DNA polymerase in a total reaction volume of 20 µl. Reagents such as dimethyl sulfoxide, Triton X-100, Tween 20, or NP-40 may be added to this mixture to enhance the quality of sequence ladders (Bachmann et al. 1990).

4. Transfer 4 µl of the above mixture to each of the four tubes containing dNTP/ddNTP mixtures (from step 2). Mix the contents and layer with 20 µl of mineral oil.

5. Place the tubes in a thermal cycler that has been preheated to 94°C to initiate thermal cycling. Each cycle consists of denaturation for 1 minute at 94°C, annealing for 30 seconds at 40–60°C, and extension and chain-termination for 30 seconds at 72°C. These steps are repeated for 20–40 cycles.

6. After the completion of thermal cycling, add 4 µl of Stop buffer containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.5% xylene cyanol FF to the reaction mixture. Mix and centrifuge for a few seconds in a microfuge to separate the layers. If the samples are not used immediately, they can be stored frozen at −70°C for about a week. However, it is preferable to use the samples within 2 days. Heat the samples for 3 minutes in an 80°C heat block, and load a 2–3 µl aliquot of the aqueous phase in a single well of the sequencing gel.

Manipulation of the Size Range of Sequence Ladders

Most of the compositions of dNTP/ddNTP mixtures reported in the literature or available commercially are optimized to generate sequence ladders of high intensity and uniformity in the range of 50–200 nucleotides from the 3′ end of the primer. Generally, the sequence ladders closer to the primer (1–50 nucleotides from the 3′ end of the primer) are of low intensity under these conditions. If high-intensity sequence ladders are desired very close to the primer (in the 1- to 100-nucleotide range), Mn^{++} can be added to the sequencing reactions (Tabor and Richardson 1989). DNA polymerases incorporate ddNTPs about five- to tenfold more frequently in the presence of Mn^{++}. Because Mn^{++} effects are seen in the presence of Mg^{++}, no changes to

the basic protocols are required other than the addition of Mn^{++}. Different polymerases require different Mn^{++} concentrations for optima results (Tabor and Richardson 1989; Rao and Saunders 1992). A Mn^{++} concentration of 3.5 mM is optimal for Sequenase for increasing the frequency of terminations by about fivefold. (Mn^{++}, prepared as a stock solution of 100 mM $MnCl_2$ in 150 mM sodium isocitrate, is available from USB/Amersham.) If it is desired to generate sequence ladders far from the primer (in the 200- to 400-nucleotide range), the dNTP/ddNTP ratios should be increased by simply adding an appropriate aliquot (this varies depending on the polymerase used) of a dNTP stock solution to the extension-termination mixture. This decreases the frequency of terminations, and therefore, increases the average length of the chain-terminated products.

Carryover Nucleotides and Primers

In a standard PCR, about 200 μM each of four dNTPs and 50 pmoles each of two primers are used to amplify several micrograms of target DNA. Of these, >97% of dNTPs and >90% of primers remain unused at the end of PCR (Rao and Saunders 1992; Rao 1994). Unless these are removed, direct use of PCR-amplified DNA as a sequencing template results in the carryover of unused dNTPs and primers to the sequencing reactions. These interfere with the sequencing reactions in the following ways: (1) The carryover dNTPs alter the dNTP/ddNTP ratios that are required for optimal chain-termination reactions; (2) since the carryover primers can also prime DNA synthesis, they compete with the sequencing primer and titrate out polymerase as well as dNTPs and ddNTPs; and (3) as a result of priming by carryover primers, a mixture of sequence ladders is generated by the Sequenase protocol because the DNA synthesized from the PCR primers is also radioactively labeled.

As a consequence of the above interferences, the sequence ladders generated are of low intensity with high background. It is difficult to decipher the DNA sequence from such ladders; therefore, it is essential to remove the unused dNTPs and primers from PCR mixtures. A number of strategies have been reported for the separation of low-molecular-weight dNTPs and primers from the high-molecular-weight, PCR-amplified DNA (Rao 1994). These include differential precipitation, ion-exchange chromatography, gel filtration, and streptavidin chromatography. Of these, ion-exchange chromatography may be the best way to remove quantitatively the low-molecular-weight dNTPs and primers from the high-molecular-weight PCR-amplified DNA. In our experience, QIAquick spin column separation gives clean DNA templates and generates high-intensity sequence ladders. These columns are designed to resolve either the single-

stranded or the double-stranded PCR-amplified DNA from dNTPs and PCR primers that are less than 50 nucleotides long.

Heterogeneity of DNA Template

PCR-amplified DNA is a product of virtually billions of in vitro priming and extension reactions. Therefore, inherent in the PCR process is the amplification of heterogeneous DNA in addition to the unique target DNA. These heterogeneous DNA molecules differ in sequence and size, and arise as a result of secondary reactions (Rao 1994). These include (1) partial products generated by the premature termination of DNA synthesis, (2) mosaic products generated by random intermolecular recombination between target DNA strands, and (3) multiple products generated as a result of priming at sequences that have either accidental homology or functional relatedness to the target DNA of interest.

Most of the standard PCRs amplify predominantly the unique target DNA, whereas the heterogeneous DNA constitutes a minor fraction. This fraction usually appears as a smear upon agarose gel electrophoresis and ethidium bromide staining. However, it is not uncommon to see that a major portion of the amplified DNA is constituted by heterogeneous DNA. This happens particularly when the copy number of starting sample is low or when PCRs are performed under low-stringency conditions (Arnheim and Erlich 1992; Saunders et al. 1993). The heterogeneous DNA, depending on the amount present in the PCR-amplified DNA, accordingly contributes to the background in the final sequence ladders.

Two major factors that lead to the formation of heterogeneous DNA are strand annealing and random priming during PCR amplification (see below). Therefore, careful consideration should be given in designing the parameters for PCR amplification. In particular, primer design, annealing temperature, and conditions or treatments prior to the initiation of the first PCR cycle should be stringently controlled to amplify only the target DNA. In general, a high-quality PCR-amplified product is generated by choosing highly specific PCR primers with a GC-content of >50%, using a high annealing temperature that is very close to the estimated T_m value of the primers (in the range of 50–60°C), initiating PCR by hot start, and if necessary, performing a second nested PCR with an internal set of primers (Arnheim and Erlich 1992). In addition, PCR should be performed for as few cycles as possible using a high copy number of target DNA. Despite these modifications, if considerable background still exists, or when multiple products are amplified, it is essential to purify the desired amplified product by agarose gel electrophoresis. The DNA fragment is extracted from the agarose gel by the procedure given below (Rao 1994),

or by any standard procedure, and is used as a template for sequencing reactions.

PROTOCOL

Purification of PCR-amplified DNA by Agarose Gel Electrophoresis

1. Extract the aqueous phase (100 µl) of the PCR mix twice with an equal volume of chloroform, and concentrate it to ~25 µl by extracting twice with an equal volume of *n*-butanol. More than one PCR mixture can be pooled and concentrated in this way. Extract once with an equal volume of water-saturated ether to remove traces of *n*-butanol. Evaporate the ether for 3 minutes at 67°C.

2. Load the entire sample into a single well of an 0.8% low-melting-temperature agarose gel containing 1 µg/ml of ethidium bromide, and electrophorese at 150 volts for 1 hour. Excise the DNA band(s) of interest and suspend each agarose slice in about 300 µl of a buffer containing 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 100 mM NaCl. Melt the gel for 20–30 minutes at 67°C. It is critical to melt the gel completely. When the tube is tapped, the solution should appear clear, and no fine gel particles should remain. Otherwise, poor recovery of DNA will result.

3. Add an equal volume of phenol to the melted gel while it is still at 67°C. Mix immediately by vortexing for a few seconds. Centrifuge in a microfuge at the maximum speed for 15 minutes. The melted gel material forms a thick precipitate at the interface. Remove the top aqueous phase and extract it once again with phenol, twice with an equal volume of chloroform, and once with an equal volume of water-saturated ether. Each extraction is done for a few seconds either by tapping the tube vigorously or by vortexing. Between extractions, the aqueous phase is recovered by centrifugation in a microfuge for a few seconds at the maximum speed. Residual ether after the last extraction is removed by heating the sample for 3 minutes at 67°C.

4. Precipitate the DNA with three volumes of ethanol for 30 minutes to overnight at –70°C. Sediment the DNA by centrifugation at the maximum speed for 15 minutes. Wash the DNA pellet twice with 1 ml each of 80% ethanol. Remove residual ethanol by drying the pellet for 2 minutes at 37°C. Dissolve the DNA in 20 µl of sterile H_2O. Some insoluble material may remain in the DNA pellet, but this does not interfere with the sequencing reactions. A 1- to 3-µl aliquot of the DNA is used for the sequencing reactions.

   Occasionally, the final DNA preparation may contain some inhibitors (most likely residual phenol), which inactivate the Se-

quenase enzyme during DNA synthesis and generate considerable ddNTP-independent terminations. These appear as background bands in all four lanes. This background can be eliminated by repeating the extractions starting from the chloroform extraction in step 3.

Random Priming

Random priming, as opposed to specific priming by the sequencing primer, refers to the priming of DNA synthesis at random points on any DNA strand in the reaction mixture. Random priming is facilitated by the interaction of a few nucleotides at the 3′ end of the sequencing primer with a short stretch of complementary sequence at a random position on the DNA (Rao 1994). Such complexes, although inherently unstable, are rapidly stabilized by the addition of a few nucleotides by the highly active DNA polymerase (the in vitro synthetic rate of *Taq* DNA polymerase is on the order of 5000 nucleotides per minute). Random priming events are also facilitated by the ends of PCR-amplified DNA. PCR-amplified DNA, because it is a linear molecule, consists of numerous 3′ ends, each having a free hydroxyl group. These 3′ ends, especially the ends of short partial products formed during PCR, are capable of priming DNA synthesis in a random fashion. As a result, a mixture of random chain-terminated products are generated in addition to the specific chain-terminated products. The random products appear as background bands in all lanes when [α-^{32}P]dATP is used for labeling DNA, either by the Sequenase strategy or by the cycle sequencing strategy.

The random priming events can be minimized by the following considerations. First, use of a 5′-labeled primer rather than [α-^{32}P]dATP in the cycle sequencing protocol eliminates this background. This is because the random chain-terminated products are not radioactively labeled and therefore do not appear on the final autoradiogram. However, even under these conditions, extensive random priming should be avoided, since the random events, by competing with the specific events for polymerase and dNTPs, reduce the intensity of the specific bands.

Random priming can also be minimized by choosing a stringent annealing temperature. In general, a very low temperature of annealing (snap cooling [Kusukawa et al. 1990] as in the Sequenase protocol), or a high temperature of annealing (close to the estimated T_m value of the primers as in the cycle sequencing protocol), improves the specificity of priming. However, a high annealing temperature (50–60°C) as in the cycle sequencing protocol is preferred over the snap cooling procedure because a temperature shift from 94°C (denaturation temperature) to 50–60°C (annealing temperature) can

be accomplished by a thermal cycler in less than 30 seconds. This minimizes random priming events as well as the strand-annealing events (see below).

Strand Annealing

PCR-amplified DNA is a linear, double-stranded molecule several hundred nucleotides in length. However, the DNA polymerases used for in vitro DNA synthesis cannot replicate a double-stranded template because they lack the accessory proteins such as single-stranded DNA-binding protein, helicase, etc. Therefore, the double-stranded PCR product must first be converted to a single-stranded form. This is accomplished by heating PCR-amplified DNA for several minutes at 94°C. The denatured single strands tend to reassociate, reconstituting the double-stranded form, which has two consequences.

First, if this happens during the annealing step, binding of the primer to the template and priming of DNA synthesis are inefficient, leading to the appearance of very low intensity sequence ladders. Therefore, the two protocols described above are designed to minimize the time required for the annealing step. This favors the kinetics of annealing of a short primer to the template rather than that of the opposite strand that is several hundred nucleotides long. The presence of detergents such as NP-40 and Tween 20, and possibly also dimethyl sulfoxide, may further inhibit strand annealing and improve the quality of the sequence ladders generated (Bachmann et al. 1990).

A second problem associated with strand annealing involves the extension step. Even when the annealing of primer to template occurs, extension by polymerase can be hindered due to strand annealing in the regions downstream from the replication fork. Therefore, DNA synthesis is terminated randomly, resulting in the dissociation of polymerase from the replication complex. These random stops, resulting from ddNTP-independent terminations, appear as background bands in all four lanes. Indeed, this is one of the primary reasons for the appearance of high background in the sequence ladders generated from PCR-amplified DNA. In addition to the random stops, stops can also occur at a high frequency at specific points on the template. For instance, GC-rich regions of the template tend to form double-stranded complexes readily and generate a specific terminated product at a much higher frequency than a random product. These "strong-stop" products appear as intense bands in all four lanes at specific positions on the autoradiogram.

These strand-annealing problems can be overcome by the following modifications: (1) Sequencing reactions should be performed under high-stringency conditions, for example, the use of high annealing and extension temperatures for short time periods (clearly, the cycle

sequencing protocol is best suited for maintaining high-stringency conditions during sequencing reactions); (2) the concentrations of ddNTPs required for chain termination should be optimized to generate predominantly ddNTP-specific terminations that would far outweigh the nonspecific terminations; and (3) because the products of ddNTP-independent terminations have a free 3′-hydroxyl group, these can be further extended by terminal dideoxynucleotidyl transferase in a template-independent manner; this enzyme converts the background bands to high-molecular-weight products that are retained at the top of the sequencing gel, thereby reducing the background (Fawcett and Bartlett 1990). Finally, strand-annealing problems can be eliminated by generating a single-stranded DNA template either by asymmetric PCR (Gyllensten and Erlich 1988) or by converting the double-stranded PCR-amplified DNA into single-stranded half-templates using phage T7 gene *6* exonuclease (USB/Amersham) (Fuller 1989).

Errors in the DNA Sequence Generated from PCR-amplified DNA

It is known that the *Taq* DNA polymerase, which is the most widely used enzyme for PCR, lacks proofreading 3′ → 5′ exonuclease activity and incorporates errors at a high frequency. Estimates of error frequencies under PCR conditions range from 1 error in 4000 nucleotides synthesized to 1 error in 400 nucleotides synthesized (Saiki et al. 1988; Ho et al. 1989; Eckert and Kunkel 1990). Therefore, it is possible that almost every molecule in a 1-kb size PCR-amplified DNA could have an error. The question then is whether direct sequencing of the mutant PCR-amplified DNA results in the generation of an inaccurate DNA sequence. The answer to this question is clear when one considers a worst-case scenario in which a hypothetical target DNA is amplified starting from a single DNA molecule and assuming that an error is incorporated in the very first cycle. Consequently, after the first cycle, one of the four DNA strands has a mutation. Upon further amplification of these four strands for about 25 cycles, the mutant sequence constitutes about 25% of the final amplified product. When this DNA is used as a template for sequencing, the final autoradiogram shows the mutant band only at one-third the intensity of the correct nucleotide band. Because the DNA sequence is read by subtracting the background, the nucleotide sequence deduced from such a sequence ladder is that of the correct nucleotide.

The above scenario is one that is encountered rarely because, in most PCR experiments, the copy number of starting DNA is on the order of 10^3–10^5 molecules. Any errors incorporated in the initial cycles are randomized, and therefore, any specific mutant sequence constitutes only a minuscule fraction of the total product. The DNA sequence generated is a consensus sequence from millions of template

DNA molecules, thus these errors are not accounted for in the final sequence. Therefore, in practice, the direct sequencing of PCR-amplified DNA does not result in the incorporation of errors. The sequence generated is accurate, despite the low fidelity of *Taq* DNA polymerase.

Indeed, it is desirable to generate the DNA sequence by directly sequencing the PCR-amplified DNA rather than from a clone of the amplified DNA, because during cloning, a single DNA molecule out of billions of amplified molecules is selected. It is highly probable that the cloned molecule will contain a mutation as a result of PCR. Therefore, the sequence generated from a cloned DNA should be confirmed by comparing it with the sequence generated directly from the PCR-amplified product to ascertain the accuracy of the cloned sequence.

Other thermostable DNA polymerases such as the Vent polymerase (New England Biolabs) and the *Pfu* polymerase (Stratagene) exhibit proofreading exonuclease activity and reportedly incorporate errors 15- to 30-fold less frequently than the *Taq* DNA polymerase (Lundberg et al. 1991; Mattila et al. 1991). Therefore, these polymerases should generate a better-quality PCR product, particularly if the cloned DNA is also used for the expression and functional characterization of gene products.

There are, however, situations in which the direct sequencing of PCR-amplified DNA is not the answer. For example, if the starting sample contains one normal allele and one deleted allele, the deleted allele is masked and cannot be deduced from the sequence generated by direct sequencing. Similarly, if the sample contains multiple alleles, direct sequencing generates a composite sequence ladder. It would be hard, if not impossible, to decipher the sequence of individual alleles from such a composite ladder. In such situations, the PCR-amplified DNA should be cloned first and the DNA sequence of a number of clones should be determined to generate the DNA sequence of individual alleles.

CONCLUSIONS

The direct sequencing of PCR-amplified DNA bypasses the time-consuming cloning steps and rapidly generates accurate DNA sequence information from small quantities of precious biological samples. Although this approach generates considerable background, in most cases it does not preclude the researcher from generating complete and accurate DNA sequencing information. As discussed above, this background can be minimized by performing both the PCR and DNA sequencing reactions under stringent conditions. The two strategies described above allow the determination of the DNA sequence from almost any type of PCR-amplified DNA template. Although both protocols are expected to generate high-intensity se-

quence ladders, the cycle sequencing strategy is preferred over the Sequenase strategy because it is very convenient to set up, especially when it is necessary to generate DNA sequences routinely from a large number of templates. In addition, reaction parameters, such as random priming and strand annealing, can be controlled stringently using the cycle sequencing strategy. This is particularly useful for generating sequences from templates having a high GC content, because the extension and chain-termination reactions can be performed at an elevated temperature, which destabilizes secondary structures in the DNA template.

In summary, high-intensity, low-background sequence ladders can be consistently generated directly from PCR-amplified DNA by (1) performing PCR under stringent conditions, (2) purifying the PCR-amplified DNA from unused dNTPs and primers by QIAGEN chromatography, (3) performing cycle sequencing under stringent conditions, and (4) using a high-specific-activity, 5′-labeled sequencing primer for amplifying the chain-terminated products.

ACKNOWLEDGMENT

The author thanks Mark Miller from the Center for Advanced Training in Cell and Molecular Biology, Department of Biology, The Catholic University of America, for his expert assistance in the preparation of schematics.

REFERENCES

Arnheim, N. and H. Erlich. 1992. Polymerase chain reaction strategy. *Annu. Rev. Biochem.* **61:** 131–156.

Bachmann, B., W. Luke, and G. Hunsmann. 1990. Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.* **18:** 1309.

Barr, P.J., R.M. Thayer, P. Laybourn, R.C. Najarian, F. Sela, and D.R. Tolan. 1986. 7-Deaza-2′- deoxyguanosine-5′-triphosphate: Enhanced resolution in M13 dideoxy sequencing. *BioTechniques* **4:** 428–432.

Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18:** 3739–3744.

Fawcett, T.W. and S.G. Bartlett. 1990. An effective method for eliminating "artifact banding" when sequencing double-stranded DNA templates. *BioTechniques* **8:** 46–48.

Fuller, C.W. 1989. Using T7 gene 6 exonuclease to prepare single stranded templates for sequencing. *U.S. Biochemical Editorial Comments* **16:** 1–8.

Gyllensten, U.B. 1989. PCR and DNA sequencing. *BioTechniques* **7:** 700–708.

Gyllensten, U.B. and H.A. Erlich. 1988. Generation of single stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQα locus. *Proc. Natl. Acad. Sci.* **85:** 7652–7656.

Ho, S.N., H.D. Hunt, R. Horton, J.K. Pullen, and L.R. Pease. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77:** 51–59.

Innis, M.A., K.B. Myambo, D.H. Gelfand, and M.A.-D. Brow. 1988. DNA sequencing with *Thermus aquaticus* polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci.* **85:** 9436–9440.

Kusukawa, N., T. Uemori, K. Asada, and I. Kato. 1990. Rapid and reliable protocol for direct sequencing of material amplified by the polymerase chain reaction. *BioTechniques* **9:** 66–71.

Lee, J.-S. 1991. Alternative dideoxy sequencing of double-stranded DNA by cyclic reactions using *Taq* polymerase. *DNA Cell Biol.* **10:** 67–73.

Lundberg, K.S., D.D. Shoemaker, M.W.W. Aetams, J.M. Short, J.A. Sorge, and E.J. Mathur. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108:** 1–6.

Mattila, P., J. Korpela, T. Tenkanen, and P. Pitkanen. 1991. Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase—An extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.* **19:** 4967–4973.

Mullis, K. 1991. The polymerase chain reaction in an anemic mode: How to avoid cold oligodeoxyribonuclear fusion. *PCR Methods Appl.* **1:** 1–5.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.* **51:** 263–273.

Murray, V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17:** 8889.

Rao, V.B. 1994. Direct sequencing of polymerase chain reaction amplified DNA. *Anal. Biochem.* **216:** 1–14.

Rao, V.B. and N.B. Saunders. 1992. A rapid polymerase chain reaction-directed sequencing strategy using a thermostable DNA polymerase from *Thermus flavus*. *Gene* **115:** 17–23.

Ruano, G. and K.K. Kidd. 1991. Coupled amplification and sequencing of genomic DNA. *Proc. Natl. Acad. Sci.* **88:** 2815–2819.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.F. Scharf, R. Higuchi, R.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.

Saunders, N., W. Zollinger, and V.B. Rao. 1993. A rapid and sensitive strategy employed for amplification and sequencing of *por A* from a single colony forming unit of *Neisseria meningitidis*. *Gene* **157:** 153–162.

Tabor, S. and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci.* **84:** 4767–4771.

——. 1989. Effect of manganese ions on the incorporation of dideoxy-nucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci.* **86:** 4076–4080.