Sequencing of PCR-amplified DNA

Ian S. Bevan, Ralph Rapley, and Matthew R. Walker

Molecular Biology Research Group, University Department of Clinical Biochemistry, Wolfson Research Laboratories, Queen Elizabeth Medical Centre, Edgbaston, Birmingham B15 2TH, UK

There is no doubt that since its inception the polymerase chain reaction (PCR)\(^{(1)}\) has revolutionized our ability to detect, analyze, and manipulate DNA and RNA sequences. The ability to analyze PCR amplification products is central to PCR's overall utility for increasing the rapidity of data acquisition, especially regarding genomic composition, organization, and regulation. Although there are many methods (reviewed in ref. 2) by which PCR products can be analyzed directly or indirectly (following cloning) by using restriction endonuclease mapping, hybridization, and so forth, nucleotide sequencing of PCR products has been problematic. Nucleotide sequencing of PCR products is essential to (i) confirm definitively the specificity of amplification, (ii) identify genetic variants (polymorphisms, rearrangements, translocations, etc.), (iii) identify hitherto uncharacterized genes, and (iv) map these genes within the organization of the genome. Thus, PCR product sequencing technology must evolve rapidly toward maximal efficiency, simplicity, reliability, and rapidity. Unfortunately, present methods for the sequencing of PCR products have not been adequate for the upsurge in the application of PCR methods to amplify diverse targets. However, this deficit is being met by a rapid development of sequencing protocols and technical advances.

In many ways the PCs' flexibility, as demonstrated by the variety of templates amplified, base analogues incorporated, and assay developed using PCR as the central technology, has allowed development of several PCR product sequencing strategies. And, as evidenced by the considerable diversity of sequencing protocols reported\(^{(3-14)}\) (see Table 1), sequencing of PCR products has not been a straightforward matter.

In practice, initial successes were achieved following the cloning of PCR products\(^{(15-20)}\). This is due in part to the desirability for cloning a PCR product other than for sequencing, including directed mutagenesis\(^{(21,22)}\) and protein expression. Furthermore, many PCR strategies intentionally or inadvertently generate a complex heterogeneous population (especially using degenerate gene family-related primers)\(^{(23)}\) and their analysis may only be undertaken by producing and analyzing a single clone. Cloning is also necessary in cases where, for example, there is need to determine unambiguously transcripts having alternative splicing, polyadenylation, or promoter start sites.

### Chemical Degradation Versus Chain-termination Sequencing

The choice of sequencing protocol based on either chemical degradation\(^{(24)}\) or enzymatic chain termina-

---

### Table 1: Sequencing Protocols

<table>
<thead>
<tr>
<th>Technique</th>
<th>Requirements</th>
<th>Preparation complexity(^{a})</th>
<th>Rapidity(^{b})</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Template preparation based</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymmetric PCR</td>
<td></td>
<td>++</td>
<td>++</td>
<td>6</td>
</tr>
<tr>
<td>Affinity capture</td>
<td>ligand modified primer</td>
<td>+++</td>
<td>+++</td>
<td>9</td>
</tr>
<tr>
<td>T7-Promoter tag</td>
<td>T7 promoter-primer</td>
<td>+++</td>
<td>++</td>
<td>10</td>
</tr>
<tr>
<td>(i) GAWTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) RAWTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ Exonuclease</td>
<td>kinased primer</td>
<td>++</td>
<td>++</td>
<td>12</td>
</tr>
<tr>
<td>SSCP/DGGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13 strand</td>
<td>cloned ss fragment</td>
<td>++++</td>
<td>+++</td>
<td>8</td>
</tr>
<tr>
<td>Indirect methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloning PCR products</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) M13 vectors</td>
<td></td>
<td>++++</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>(ii) pBluescript</td>
<td></td>
<td>++++</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>(iii) Plasmid vectors</td>
<td></td>
<td>++++</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>(b) Sequencing condition modification based</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helix destabilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) Temp (Taq)</td>
<td></td>
<td>+</td>
<td>+</td>
<td>14</td>
</tr>
<tr>
<td>(ii) DMSO</td>
<td></td>
<td>+</td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td>(iii) Detergents</td>
<td></td>
<td>+</td>
<td>+</td>
<td>46</td>
</tr>
<tr>
<td>(iv) SSB proteins</td>
<td></td>
<td>+</td>
<td>+</td>
<td>48</td>
</tr>
<tr>
<td>Cycle sequencing</td>
<td></td>
<td>+</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>Modified Sequenase</td>
<td></td>
<td>+</td>
<td>+</td>
<td>45</td>
</tr>
<tr>
<td>Coupled amplification</td>
<td>auxiliary set of</td>
<td>+++</td>
<td>+</td>
<td>52</td>
</tr>
<tr>
<td>and sequencing</td>
<td>labeled primers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)Empirically rates as simple (+) to complex (+++).

\(^{b}\)Rated as (+) for same day to (+++) more than 1 day required for completion.
relies on the individual laboratory’s experience or preference. However, there appears, for several reasons, from the literature to be a significant trend toward chain-termination-based methods. Chain-termination methods obviate the requirement for a primer label step, are available in simplified kit formats, avoid the use of toxic chemicals that the former method necessitates, and are more amenable to automation given the capability of using readily available labeled terminators. Sequencing of PCR products has been achieved using single-stranded and double-stranded products. All chain-termination sequencing reactions demand that the DNA template must be single stranded for the priming reaction to occur. Present methods essentially differ in the timing of the generation of the single-stranded DNA, i.e., during sequencing from double-stranded DNA templates or prior to sequencing by generating or separating single-stranded DNA templates. Furthermore, they differ in the methods used to generate or isolate the sequencing template, and consequently differ in their simplicity and potential rapidity.

Sequencing of PCR Products

Many methods have been described by which single-stranded templates from PCR-derived products may be generated, many having their origins in the fields of probe generation or product detection. Most require further manipulations to be undertaken before a single-stranded template of sufficient quality or quantity is produced, a potential drawback in terms of simplicity, reliability, and automation. However, there are exceptions where the single-stranded DNA template may be generated during amplification.

Templates Derived from Cloning Vectors

Cloning of PCR products into sequencing vectors such as M13 via cohesive or blunt-ended ligation or ligation-free methods allows the production of templates suitable for chain-termination sequencing and is subject to the same restraints as non-amplified fragments. Often PCR products are required for expression studies and must be cloned into a vector; therefore, it may be preferential to clone into a dual-purpose vector suitable for both sequencing and expression studies. Additional problems inherent with cloning PCR products do, however, decrease the desirability of performing cloning steps prior to sequencing (e.g., the low efficiency of digestion of primers containing certain restriction sites). Furthermore, rigorous screening of many clones may also be required to avoid artifacts such as shuffle clones, misprimed products, or misincorporation errors. Indeed, sequencing of numerous clones to gain consensus sequences initially was deemed imperative for taking into account possible amplification artifacts, especially the introduction of base substitution errors during amplification. Misincorporation has largely proven to be a less significant problem due to the availability of thermostable DNA polymerases with increased fidelity or proofreading activity.

There are, however, certain circumstances whereby laborious and time-consuming cloning of PCR products remains advantageous, for example, to discriminate between polymorphic or gene family-related products coamplified as a single-molecular-weight DNA species in a single tube (using degenerate or mixed allele-specific primers). More recently, short (up to 500 bp) heterogeneous PCR products have been separated as single strands in gel electrophoresis based on strand conformation differences (termed single-strand conformation polymorphism, SSCP).

Templates Derived from Asymmetric PCR

The fundamental amplification mechanism of the PCR can readily be modified to the preferential production of single-stranded product by limiting the availability of one of the pair of oligonucleotides primers; this process is termed asymmetric PCR (APCR). Primer ratios of 50:1 to 100:1 are most frequently used to generate single-stranded products. During the first 15–20 cycles, double-stranded DNA accumulates in an exponential fashion until limited by primer depletion. This continues to act as a template for linear copying of single-stranded DNA primed from the unexhausted oligonucleotide. However, single-stranded products generally require further separation prior to sequencing, which may be somewhat problematic given the range of mobilities that may be observed and the poor staining of single-stranded DNA by ethidium bromide. In practice, the difficulties in generating (and subsequently isolating) single-stranded DNA by APCR has led to alternative methods for single-stranded PCR product generation. Nevertheless, APCR has been successfully applied to the sequencing of many genes following its initial use for sequencing the HLA-DQA locus.

Template Derived from Affinity-capture

Sequencing of affinity-captured products has been used successfully in many situations, such as the generation of nucleotide sequence for mitochondrial DNA amplified from crude genomic DNA extracts, and to genotype individuals for human apolipoprotein E alleles. In the case of single base differences amongst coamplified alleles, full-length sequencing of the products may not be required. Indeed single base differences between alleles may be detected by single base addition of a labeled nucleotide to the sequencing primer.

The use of affinity-capture of single-stranded DNA for sequencing relies on the incorporation of a ligand (such as biotin) in one of the amplimers, generally at the 5' terminus of the primer, (see Fig. 1). Following amplification, the double-stranded product is passed through a column or over a matrix containing a ligand-binding, such as streptavidin, to which the biotinylated primer-derived strand will bind. This allows subsequent denaturation and elution of the unbiotinylated strand with NaOH. Following this procedure, it is possible to use either the eluted, unlabeled strand of DNA for sequencing or, alternatively, the affinity matrix-bound strand. The attached moiety may in itself facilitate the subsequent sequencing of the single-stranded DNA and increases the potential for automation of the process; in particular, the improved handling properties of magnetic microspheres

PCR Methods and Applications 223
appear to offer many advantages. Magnetic beads have also facilitated sequencing by walking in any direction along genomic DNA starting from a known sequence using both exponential and linear PCR.\(^{(34)}\)

The affinity-capture approach to single-stranded DNA PCR product separation prior to sequencing allows flexibility or universal applicability in terms of matrices (e.g., cellulose, magnetic microspheres), ligands, and receptor molecules. A potential drawback to this and several other approaches is, however, that the original PCR reaction must not generate any nonspecific amplification products because these will also be captured, resulting in "ghost sequences" overlaying the true target sequence generated. Many of these problems may be eliminated by optimization of amplification specificity\(^{(35-38)}\) or the use of a nested sequencing primer. Perhaps more important to genome mapping studies and clinical diagnostics, these methods lend themselves to the development of rapid, automated systems.

**Automated Systems**

Automated workstations are being developed that incorporate PCR to generate sequencing templates and chain-extension termination products. These workstations detect the products generated by using fluorescent dye-labeled amplimer/sequencing primer or fluorescent dye-labeled dideoxy nucleotide analogues.\(^{(26)}\) This avoids the use of potentially hazardous radioactive labels and, in combination with advanced detectors and signal compensating software packages, increases the lengths of readable sequence obtained. Although some methods utilize affinity capture for separating the double-stranded DNA products, more elegant approaches have been developed based on linear generation of terminated products in "cycle" sequencing. Subsequent to initial cycles of PCR, labeled primers and ddNTPs are included in the reaction to enable generation of chain termination products. Discussion of the relative merits of various automated sequencers is not appropriate here, because most of these considerations are based on technological differences in terms of cost, throughput, and rapidity rather than the efficiency of sequencing protocols.

**Templates Derived from an Amplimer Containing T7 Promoter Sequence**

A relatively simple method for generating single-stranded DNA from double-stranded DNA amplification products was originally developed to facilitate in vitro probe generation. This procedure relies upon the addition of a 5' T7 promoter sequence onto one of the amplimers\(^{(39)}\) (see Fig. 2). Thus, following amplification, T7 DNA polymerase may be used in a second reaction to generate single-stranded DNA. This approach is applicable to both RNA (RNA amplification with transcript sequencing, RAWTS)\(^{(10)}\) and genomic DNA (genomic amplification with transcript sequencing, GAWTS)\(^{(7)}\) sequencing. These techniques are based on the generation of single-stranded RNA from the double-stranded DNA product. The RNA may then be used as a sequencing template using standard protocols. Unfortunately, this approach is suscep-
FIGURE 2 Strategy for T7 promoter-tagged primer sequencing of RNA (RAWTS). In RAWTS, a first-strand cDNA is generated by reverse transcription, which then serves as template for PCR amplification using one amplimer containing the bacteriophage T7 promoter sequence. The PCR strand generated from the T7 promoter sequence-tagged amplimer can subsequently be used as template for RNA transcription using T7 RNA polymerase to generate a single-stranded RNA product. This can be used directly for sequencing by the dideoxy chain-termination method. Promoter-tagged sequencing of DNA (GAWTS) may also be similarly performed by excluding the reverse transcription step.

Sequencing from Double-stranded PCR Products

An obvious advantage of sequencing a double-stranded PCR product is that either DNA strand may be used. Most published methods for sequencing double-stranded DNA PCR products require that the specific PCR product must be purified from any remaining primers, incomplete products, or misprimed artifacts. This step is usually required whether an amplimer or nested (internal) oligonucleotide is to be used as the sequencing primer. Purification is generally achieved by either agarose or acrylamide gel electrophoretic separation followed by gel excision and elution. Methods of gel elution vary (often empirically) and range from electroelution, elution-capture, or digestion of the gel matrix, for example, by agarose; their relative merits appear to vary among laboratories in terms of purity and yield of product recovered. The use of spin columns and centrifuge ultrafiltration methods have also been important in allowing recovery and concentration of the specific PCR products, as has the application of more sophisticated chromatographic techniques such as high-performance liquid chromatography (HPLC).

Helix Destabilization Methods

The essential problem with sequencing individual strands of duplexed PCR products appears to be maintaining dissociation of the two rapidly reannealing complementary strands sufficiently while allowing annealing and extension of the sequencing primer. The result of reannealing of the duplexes during the sequencing reaction is usually premature termination and/or high background terminations (often seen as four-tracking; see Fig. 3). Therefore, it is unsurprising that successful protocols have commonly involved reagents or procedures designed to maximize duplex dissociation. Many of these protocols were first employed for sequencing double-stranded plasmids, where an alkali denaturation step is sufficient to denature the two strands, allowing sequencing primer annealing and extension. Generally, however, these plasmid sequencing protocols require further manipulation to allow sequencing of PCR products.

Use of Denaturants or Single-stranded DNA Binding Proteins

The first step of the sequencing reaction requires the denaturation of the double-stranded product such that a sequencing primer may bind. This is generally achieved by heating to 100°C for several minutes. Although slow cooling will favor reassociation of the denatured DNA, very rapid cooling will maintain the dissociation, allowing the sequencing reaction to continue by favoring the annealing of the shorter sequencing primer. Much investigation has been carried out regarding the temperatures required to prevent template reannealing, with the coldest temperatures (e.g., snap-freezing in liquid nitrogen) having greater efficacy. The sequencing reaction must occur at the optimal temperature for the DNA polymerase used, usually at 37°C, and therefore, reagents that will promote the continued dissociation of the PCR
FIGURE 3 Examples of frequently encountered artifacts generated during nucleotide sequencing of double-stranded PCR products. (a) Multiple tracking, where more than one equally intense band is visible at the same position on the autoradiograph. In some cases, this may be due to runoff products from other fragments in the mixture. However, this particular problem may be overcome by PCR product purification prior to sequencing. (b) "Ghosting," where, although more than one band is visible at the same position, it is possible to differentiate the correct base termination based on the specific bands' greater intensity. (c) Compression, where correct termination products are out of sequence or compacted within the gel. These artifacts are not unique to PCR product sequencing; however, they are particularly prevalent unless sequencing protocol modifications are adopted to destabilize duplexes or secondary structures. (d) Results that may be obtained using double-stranded templates with modified sequencing protocols. (i) Multiple tracking; (ii) ghosting; (iii) region of compression (correct sequence reads CTCCA).

Use of Taq Polymerase

Alternatively, thermostable DNA polymerases may also be used in the sequencing reactions, allowing the reaction to be performed at around 70°C and reducing the risk of target DNA annealing and the sequencing of templates with significant secondary structure. Modifications of the linear PCR using thermostable DNA polymerase in direct sequencing produce reduced background signals, increased rapidity, and permit automation. Sequencing reaction conditions must, however, be varied according to template and reaction volumes. Of particular importance is the ddNTP concentration because some polymerases (such as Taq) incorporate ddNTPs less efficiently than dNTPs, although Taq polymerase has high processivity and more readily allows the incorporation of helix destabilizing nucleotide analogues, such as 7-deaza-2'-dGTP, permitting increased resolution of band compressions (Fig. 3). However, certain analogues, such as inosine, should be avoided when using Taq polymerase.

Summary

Alternatives for sequencing of PCR products essentially fall into one of two categories; generation of single-stranded DNA for sequencing or the direct sequencing of double-stranded product. Of the two alternatives, sequencing of double-stranded PCR products is likely to be of greatest immediate significance in terms of general applicability and rapidity. Double-stranded sequencing allows the use of the PCR product for other purposes either prior to or subsequent to generation of sequence data. The single-stranded sequencing methods generally require some prior decision regarding sequencing of the product. Assisted by automated workstation development, sequencing of single-stranded DNA PCR products generated either during thermal cycling or following affinity-capture strand separation may have significant future utility, particularly in genome mapping and routine clinical diagnosis. Despite template type and protocol differences, in all situations the purity and concentration of PCR-amplified DNA template used remains the most critical factor determining the efficiency and reliability of nucleotide sequencing methods.

References

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

40. Murphy, G. and T. Kavanagh. 1988. Speeding up the sequencing of double-stranded DNA. *Nucleic Acids Res.* **16:** 5198.


Sequencing of PCR-amplified DNA.

I S Bevan, R Rapley and M R Walker

*Genome Res.* 1992 1: 222-228
Access the most recent version at doi:10.1101/gr.1.4.222

References
This article cites 49 articles, 11 of which can be accessed free at:
[http://genome.cshlp.org/content/1/4/222.full.html#ref-list-1](http://genome.cshlp.org/content/1/4/222.full.html#ref-list-1)

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](http://genome.cshlp.org/subscriptions)