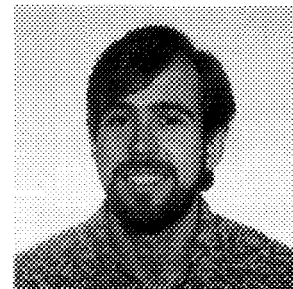


REVIEW

Revolutions in Rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends

Brian C. Schaefer

Division of Tumor Virology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115



Rapid amplification of cDNA ends (RACE) is a polymerase chain reaction (PCR)-based technique which was developed to facilitate the cloning of full-length cDNA 5'- and 3'-ends after a partial cDNA sequence has been obtained by other methods. While RACE can yield complete sequences of cDNA ends in only a few days, the RACE procedure frequently results in the exclusive amplification of truncated cDNA ends, undermining efforts to generate full-length clones. Many investigators have suggested modifications to the RACE protocol to improve the effectiveness of the technique. Based on first-hand experience with RACE, a critical review of numerous published variations of the key steps in the RACE method is presented. Also included is a detailed, effective protocol based on RNA ligase-mediated RACE/reverse ligation-mediated PCR, as well as a demonstration of its utility. © 1995 Academic Press, Inc.

Since the polymerase chain reaction (PCR)¹ was developed and the basic reaction conditions were opti-

¹ Abbreviations used: ACS-water, water meeting American Chemical Society specifications; AMV-RT, avian myeloblastosis virus reverse transcriptase; A-PCR, anchored polymerase chain reaction; BAP, bacterial alkaline phosphatase; BL, Burkitt's lymphoma; bp, base pair; cDNA, complementary DNA; CIP, calf intestinal alkaline phosphatase; dCTP, deoxycytidine triphosphate; ddA, dideoxy adenosine; ddTTP, dideoxythymidine triphosphate; DEPC-water, diethyl pyrocarbonate-treated water; dGTP, deoxyguanosine triphosphate; dNTP, a deoxynucleoside triphosphate; dsDNA, double-stranded DNA; dTTP, deoxythymidine triphosphate; EBV, Epstein-Barr virus; IPTG, isopropylthio- β -D-galactoside; kb, kilobase; LA-PCR, ligation-anchored polymerase chain reaction; 7-MeG, 7-methyl guanosine; 7-MeGppp, 7-methylguanosine triphosphate; MMLV-RT, Moloney murine leukemia virus reverse transcriptase; mRNA, messenger RNA; NPC, nasopharyngeal carcinoma; oligo, oligonucleotide; pCp, cytidine 3',5'-bisphosphate; PCR, polymerase chain reaction; poly(A) tail, messenger RNA polyadenosine tail; RACE, rapid ampli-

mized, the technique has been applied to many aspects of molecular biology research (1,2). One of the most useful applications of PCR is the amplification and subsequent cloning of low-abundance DNA sequences from complex mixtures, even using as little as a single cell as the source of genetic material. Previously, such cloning projects involved the arduous task of creating a phage or plasmid library and screening millions of plaques or colonies to obtain overlapping cDNAs representing the full-length sequence. This procedure often requires many months of work to isolate the desired sequence in entirety. The advent of PCR has made it possible to reduce this time to weeks or days.

This Review focuses on the application of PCR to the cloning of unknown flanking cDNA sequences when a small amount of cDNA sequence is already known. There were originally several independent descriptions of similar methods from researchers who named their techniques rapid amplification of cDNA ends (RACE) (3), anchored PCR (4), and one-sided PCR (5). There have since been several modifications of the original methods, including ligation anchored PCR (LA-PCR) (6,7), single-strand ligation to ss cDNA ends (SLIC) (8,9), RNA ligase-mediated RACE (RLM-RACE) (10), and reverse ligation-mediated PCR (RL-PCR) (11). By slightly different applications of the protocols, unknown sequences both 5' and 3' of the known sequence can be reverse transcribed from RNA, amplified by PCR, and cloned. Generally, the objective of investigators who use this method is to rapidly generate full-length cDNA clones after a partial cDNA has been ob-

fication of cDNA ends; RLM-RACE, RNA ligase-mediated rapid amplification of cDNA ends; RL-PCR, reverse ligation-mediated polymerase chain reaction; RT-PCR, reverse transcriptase polymerase chain reaction; SLIC, single-strand ligation to ss cDNA ends; ss, single-stranded; TAP, tobacco acid pyrophosphatase; TdT, terminal deoxynucleotidyltransferase; X-gal, 5-bromo-4-chloro-3-indolyl- β -D-galactoside.

0003-2697/95 \$6.00

Copyright © 1995 by Academic Press, Inc.

All rights of reproduction in any form reserved.

255

tained from a cDNA library. While the method is simple from a theoretical perspective, the actual execution is often fraught with technical difficulties, requiring substantial optimization before the desired result is obtained. Based on extensive experience with these methods, the various approaches described in published protocols will be evaluated, and a number of modifications which have recently been developed to attempt to overcome some of the more troublesome and frequently encountered technical problems will also be reviewed.

HISTORICAL PERSPECTIVE

The basic protocol for the PCR amplification and cloning of unknown cDNA sequences adjacent to known sequences was initially described in two independent reports. Frohman *et al.* (3) described the rapid amplification of cDNA ends (RACE) procedure for cloning unknown 5'- and 3'-cDNA ends, and Loh *et al.* (4) designed the anchored PCR (A-PCR) protocol to clone the 5'-ends of T cell receptor genes which have a constant 3'-region but variable 5'-sequences. The basic procedure described by both groups is nearly identical and is shown in Fig. 1a. In both protocols, to clone an unknown 5'-end sequence, cDNA is synthesized from an RNA sample using either oligo(dT) or an oligonucleotide primer complementary to a known sequence in the target mRNA. The 3'-end of the first strand cDNA is then modified by the addition of a homopolymer tail using terminal deoxynucleotidyltransferase (TdT). The cDNA second strand is then synthesized using a primer consisting of a cloning sites adaptor having a 3'-homopolymer tail (anchor(N)₁₅) which is complementary to the appended cDNA 3'-homopolymer. These double-stranded cDNAs are amplified by PCR, employing the cloning adaptor (anchor) as the 5'-end primer and an oligonucleotide complementary to a known sequence internal to the cDNA primer as the 3'-end PCR primer. An aliquot of the products of the PCR is then separated by electrophoresis through an agarose gel, and specific products are identified, cloned, and sequenced by established methods (see below).

While the A-PCR and RACE protocols are very similar, there are a few technical differences. First, Loh *et al.* (4) chose to TdT tail their cDNAs with poly(dG), whereas Frohman *et al.* use poly(dA), and the corresponding primers for second-strand synthesis are thus anchor(dC)₁₅ and anchor(dT)₁₅, respectively. Frohman *et al.* (3) chose to tail with poly(dA) so that the same anchor(dT)₁₅ could be used in both 5'- and 3'-RACE (see below). Also, they hypothesize that the A:T base pairing interaction, which is weaker than the G:C base pairing, should reduce the likelihood of anchor(N)₁₅ hybridization and priming at internal cDNA homopolymer sequences (see discussion below, under TdT Tailing and Second Strand Synthesis), rather than from

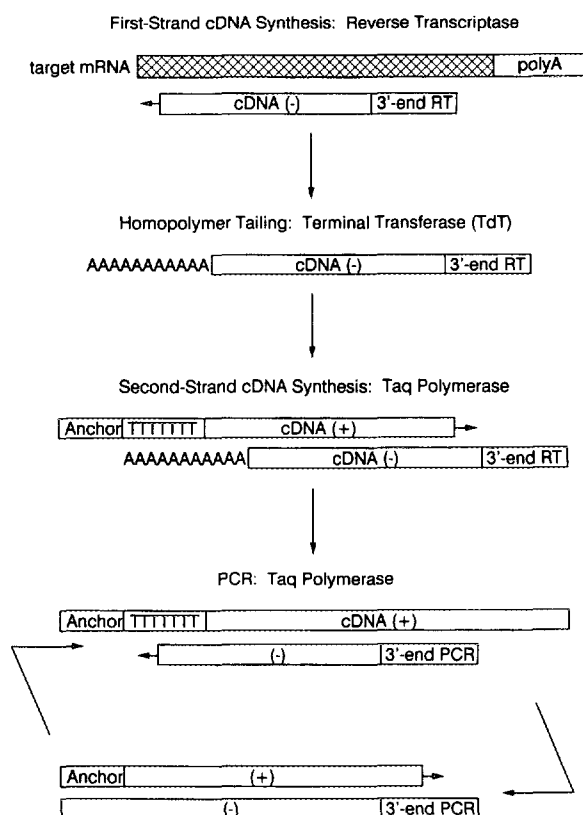


FIG. 1a. Major enzymatic steps in the standard RACE protocol. The target mRNA is first reverse transcribed using a specific 3'-end oligonucleotide primer (3'-end RT). TdT is then used to add a homopolymer tail (poly(dA), in this diagram) to the first-strand cDNA. Second-strand cDNA synthesis is achieved by the extension of an anchor primer having a poly(dT) tail, which has annealed to the first-strand cDNA poly(dA) tail, using *Taq* polymerase. The double-stranded cDNA is then amplified by PCR, using the anchor primer as the 5'-end primer and a 3'-end primer (3'-end PCR) complementary to known cDNA sequence internal to the 3'-end RT primer. Arrows at the ends of cDNA (open rectangles) indicate active DNA synthesis in the indicated direction. Bent arrows indicate the many cycles of amplification in the PCR.

the target homopolymer sequence added to the 3'-end of the first-strand cDNA (3).

To accomplish the synthesis of the cDNA second strand, both groups included the anchor(N)₁₅ primer in the PCR reaction, which allows second-strand synthesis by *Taq* polymerase to be followed immediately by PCR amplification. The RACE protocol specifies that the anchor and anchor(dT)₁₅ primers be combined at a 2.5 to 1 ratio and added, with the specific 3'-end primer, to a standard PCR mix. The authors found that the generation of full-length cDNAs was promoted by using a long extension step (up to 40 min) in the first PCR cycle, which presumably allowed more priming and synthesis of full-length second-strand cDNAs (3). In the A-PCR protocol, the anchor(dC)₁₅ and anchor primers

are present at a 1:9 ratio, and no special first cycle is included to promote full-length second-strand synthesis (4). Another small difference between the two protocols is that whereas the original description of the RACE method specifies only a single 40 cycle PCR (3), the A-PCR method includes two 25-cycle PCR amplifications employing two nested 3'-end primers (4). The second nested PCR amplification was included to increase specific product yield, while reducing background (see discussion below, under PCR).

The two groups also used slightly different approaches to screen for positive PCR products. Frohman *et al.* (3) Southern blotted an aliquot of the gel fractionated RACE product and used known cDNA sequences as a probe to detect the target transcript. They then digested with the appropriate enzymes to generate 5'- and 3'-sticky ends, gel-isolated fragments of the appropriate size, and cloned the fragments into a plasmid sequencing vector. Bacterial transformants were then screened by colony-lift hybridization, again using the known cDNA sequence as a probe. Plasmids from positive transformants were sequenced. Loh *et al.* (4) used the same approach, but they omitted Southern screening of their gel-fractionated A-PCR product, since they knew the approximate size of their expected products. Instead, they digested their A-PCR product and gel isolated fragments within the expected size range. Cloning, colony screening, and sequencing were performed as in the RACE method.

Frohman *et al.* (3) also include a protocol for amplifying unknown 3'-ends in their description of the RACE technique. Their method calls for the use of the anchor(dT)₁₅ to prime the synthesis of first-strand cDNAs via the annealing of the (dT)₁₅ stretch to the mRNA poly(A) tail. Since the anchor has already been incorporated during the first-strand synthesis, no TdT tailing or second-strand cDNA synthesis step is required. Instead, the first-strand cDNA is amplified by PCR using a specific primer complementary to known cDNA sequence as the 5'-end primer and the anchor primer as the 3'-end primer. Positive PCR products are identified as discussed above.

A somewhat later description of the RACE/A-PCR method was reported by Ohara *et al.* (5) who refer to their method as one-sided PCR. As in the RACE protocol, a poly(A) homopolymer tail is added to first-strand cDNAs having an unknown 5'-end. Rather than incorporating an anchor oligo via a second-strand synthesis step, the authors use the tailed cDNA directly in a PCR amplification, employing a (dT)₂₀ oligo as the 5'-end primer. The one-sided PCR protocol, like the A-PCR method (4), also includes a second PCR amplification in which an aliquot of the first PCR reaction is reamplified using the (dT)₂₀ primer and a nested specific primer, internal to the first. Since, in the one-sided PCR protocol, restriction sites are not included in the PCR oligo-

nucleotides, the PCR products are not cloned, but directly sequenced. The protocol for the amplification of 3'-ends is essentially identical to the RACE method (3), except that the (dT)₂₀ oligo is used for first-strand cDNA synthesis and as the 3'-end PCR primer, instead of the anchor(dT)₁₅ and anchor primers, respectively.

Although the various methods of using PCR to amplify and clone unknown regions of cDNA sequence have proven extremely valuable to many investigators, numerous groups (6,8,10,12,13) have observed that it can be extremely difficult to successfully apply these protocols. Generally speaking, there are two reasons for this high failure rate. The first is that there are three sequential enzymatic steps (reverse transcription, TdT tailing, and PCR), each of which is subject to partial or complete failure. The second is that even when the enzymatic syntheses proceed smoothly, there is often a high background of nonspecific or truncated product generated, often to the exclusion of specific, full-length product. In the following section, a number of proposed modifications to each of the major steps involved in the synthesis and amplification of unknown adjacent cDNA sequences will be reviewed. For simplicity, the basic method will be referred to as RACE for the remainder of the Review.

MODIFICATIONS AND OPTIMIZATION OF THE RACE PROTOCOL

Reverse Transcription

The most crucial step influencing the outcome of a RACE experiment is the generation of first-strand cDNA by reverse transcription of mRNA. This is especially true when attempting to clone the full-length 5'-ends of messages, since incomplete cDNAs will still be tailed by TdT and amplified by PCR. These truncated molecules have the same ends as the full-length molecules (i.e., both have the same specific 3'-end sequence internal to the cDNA primer and both also have an anchor sequence on their 5'-ends), thus both truncated cDNAs and full-length cDNAs will be amplified in the PCR reaction. The presence of these truncated cDNAs in the PCR mix can substantially reduce or even abolish the generation of PCR products representing full-length cDNAs, since smaller DNA molecules are often amplified preferentially in a mixture of heterogeneously sized PCR substrates which have identical ends (14).

The reverse transcription step should thus be optimized to avoid as much as possible the generation of incomplete cDNAs. It is essential, of course, to use high-quality mRNA which has not been degraded by ribonucleases. A thorough discussion of RNA preparation and handling is outside the scope of this Review. In short, we recommend the use of the single-step method (15) for preparation of RNA from tissues or

small quantities of tissue culture cells, or the cytoplasmic RNA protocol (16) for larger quantities of tissue-culture cells. We also recommend further purification of the RNA to eliminate nonpolyadenylated species by chromatography on oligo(dT) cellulose (17) or via the use of one of the many commercially available kits. The use of polyadenylated mRNA should prevent the generation and cloning of cDNAs which contain intronic sequences. If enough material is available, then an aliquot of the RNA sample to be used for RACE should be electrophoresed, Northern blotted, and probed with the known cDNA sequence to confirm that the message of interest is present and not detectably degraded.

However, even with good quality mRNA, many partial cDNAs are generated during reverse transcription. The truncation of cDNAs is generally thought to be the result of the reverse transcriptase enzyme being unable to continue polymerization in regions of the mRNA molecule which are involved in a stable secondary structure (18). Transcripts with a higher GC/AU ratio are more likely to contain regions of stable secondary structure, given that GC base pairs are substantially more stable than AU base pairs. Thus, mRNAs having the highest GC content are most likely to give rise to truncated cDNAs. An effective way to destabilize mRNA secondary structure in the reverse transcription reaction is to raise the temperature of the reaction mixture above the standard synthesis temperature (37 to 42°C). Mesophilic reverse transcriptase enzymes (AMV-RT, MMLV-RT) can only function at temperatures up to ca. 50°C, but there are now thermostable DNA polymerases which exhibit reverse transcriptase activity at substantially elevated temperatures (generally 60 to 70°C). At these temperatures, RNA secondary structure is destabilized, allowing reverse transcription to proceed more readily to the 5'-end of the molecule. Thermophilic DNA polymerases which can reverse transcribe mRNA under certain defined conditions (19–21) include, but are not limited to, rTth (Perkin-Elmer) and TetZ (Amersham). Even *Taq* polymerase has been reported to exhibit reverse transcriptase activity (22).

Although rTth is less susceptible to the effects of secondary structure than are the mesophilic reverse transcriptases (19,23), there are certain caveats which should be considered when using rTth and similar enzymes. First, rTth is primarily a DNA-directed DNA polymerase, and optimal enzyme activity occurs at a temperature (70°C) more favorable for DNA destabilization than the operational temperatures of mesophilic reverse transcriptases. Thus, contaminating genomic DNA is more likely to be primed and extended by rTth than by the mesophilic enzymes. The problem of DNA contamination can easily be avoided by pretreatment of mRNA with ribonuclease-free DNase (24).

There is a second potential difficulty related to the use of enzymes possessing thermophilic reverse transcriptase activity which is due to inherent limitations which do not affect mesophilic reverse transcriptases. Specifically, reverse transcription reactions longer than 15 min favor the production of truncated cDNAs due to rapid hydrolysis of RNA in the presence of divalent cations (Mn^{2+}) at the high temperatures required for optimal polymerase activity (19). Partially degraded RNA molecules can be primed and reverse transcribed, yielding a partial cDNA. This phenomenon may contribute to the reduced ability of rTth to generate cDNAs longer than 1 kb (20), given that longer extensions require longer periods of incubation, and that the probability of the RNA template being truncated by hydrolysis before completion of reverse transcription increases over time. Thus, the duration of the reverse transcription reaction should be limited to 15 min or less, and the region to be reverse transcribed should be kept to a minimal length. Thus, for a RACE amplification to clone a 5'-end, the reverse transcription should employ a specific primer rather than oligo(dT), and this primer should be positioned as far toward the 5'-end of the partial cDNA as possible while retaining enough sequence to specify two internal primers for PCR and a third region for Southern hybridization (25) to detect specific PCR products. Also, since full-length cDNAs are only generated efficiently when the ratio of enzyme molecules to template molecules is one or greater (19), it is important not to use excessive quantities of RNA or cDNA primer in the reverse transcription reaction.

What quantity of RNA constitutes too much template will generally have to be determined empirically, but there is evidence that primer concentrations can be reduced orders of magnitude below that which is called for in most protocols. Templeton *et al.* (26) have recently reported that when rTth is used to reverse transcribe the ϵ 1F-2 β message from 150 ng of mRNA, as little as a 1 pM concentration of specific primer was sufficient to generate the desired RACE amplification product after two rounds of nested PCR. In contrast, when the primer concentration was increased from 10 to 100 nM, specific product was no longer generated. The authors attribute the failure of RACE at the 100 nM primer concentration to an overabundance of the cDNA primer, which was then tailed by TdT and preferentially amplified in the subsequent PCR (26).

TdT Tailing and Second-Strand Synthesis

The standard RACE protocol specifies that the first-strand cDNA is tailed with a deoxyribonucleotide homopolymer, employing the TdT enzyme (3–5). This tail then serves as a substrate for a subsequent hybridization with an anchor polylinker which has a 3'-end homopolymer (anchor(N)₁₅) complementary to the cDNA

homopolymer tail. The anchor(N)₁₅ then primes second-strand cDNA synthesis. Although this approach can work effectively, it can also fail or promote the formation of undesired or artifactual products via various mechanisms.

In order for homopolymer tailing to be successful, free nucleotides and excess cDNA primer remaining after reverse transcription must be eliminated. If free nucleotides are not eliminated, they will be incorporated into the 3'-end cDNA tail by TdT, interfering with hybridization to the complementary homopolymer tail on the anchor primer. If excess cDNA primer is not removed, the 3'-end of this primer will compete with cDNA ends for tailing by TdT, lowering the efficiency of cDNA tailing. Tailed cDNA primers can also anneal to the tailed anchor primer, again competing with the cDNA, lowering the efficiency of the second-strand cDNA synthesis. These small molecules will in turn be a substrate in the PCR reaction for annealing and extension by the anchor primer, lowering PCR efficiency (26). To eliminate excess primers and free nucleotides, methods based on molecular weight exclusion such as Centricon microconcentrators (Amicon) can be used (23), but the simple and inexpensive method of two sequential precipitations with 1/4 vol of 10 M ammonium acetate and 3 vol of 100% ethanol (27) also works very well.

A major problem with the tailing strategy in 5'-end RACE is that *all* cDNAs are tailed by TdT regardless of whether they are full-length. Once the anchor has been added in the second-strand synthesis, these truncated cDNAs will compete in the PCR with full-length product for amplification (23). If the truncated products are much shorter than the full-length cDNAs, they may be preferentially amplified (14), preventing detectable levels of full-length product from accumulating in the PCR. The only way to reduce the generation of these truncated cDNAs is to optimize the reverse transcription (23), as discussed previously.

A related problem which leads to the production of truncated cDNAs can occur during the second-strand synthesis. If the target cDNA contains a homopolymer region of several nucleotides which is complementary to the anchor(N)₁₅ homopolymer tail, second-strand synthesis can be primed from that internal sequence, instead of from the terminal homopolymer region, generating a truncated second-strand. As discussed above, such truncated products will be amplified during PCR.

Internal priming by the anchor(N)₁₅ can be diminished in a number of ways. Since A:T base pairing involves a significantly weaker hydrogen bonding interaction than G:C base pairing (23,24), the use of A and T homopolymer tails tends to generate fewer internal priming events (23,24). However, in species which have A- and T-rich genomes, internal priming by anchor (dT or dA)₁₅ can be a frequent occurrence (10). Another way

to minimize nonspecific anchor(N)₁₅ priming events is to avoid using too much primer or too low a hybridization temperature during the second-strand cDNA synthesis. In this laboratory, reproducible success has been achieved using 0.5 pmol of an anchor(dT)₁₅ primer per microgram of mRNA starting material in a 20- μ l volume. Hybridization to the (dA)_x tailed cDNA is conducted at room temperature for 30 min prior to second-strand synthesis.

Alternatives to TdT tailing. Citing the numerous shortcomings inherent in the homopolymer tailing strategy, several groups have independently devised modifications of the RACE method which circumvent TdT-mediated tailing of the cDNA. These methods all incorporate the use of RNA ligase to covalently attach an anchor oligonucleotide to the 5'-end of the cDNA (6-9,12) or directly to the starting RNA population (10,11,13,28-30). Such strategies thereby eliminate both the TdT tailing step and the need for anchor(N)₁₅-primed second-strand cDNA synthesis. The possibility of internal priming by the anchor(N)₁₅ is thus entirely avoided (9,12). In the protocols which specify ligation of the anchor directly to RNA, the generation of any truncated cDNAs possessing a 3'-terminal anchor sequence can theoretically be avoided (10). These innovations thus represent a considerable improvement to the original RACE protocol.

Over the past 3 years, several groups have published protocols which describe the ligation of a DNA anchor oligonucleotide to the first-strand cDNA (6-9,12). This RACE variation is generally referred to as ligation-anchored PCR (LA-PCR) (6,7) or single-strand ligation to ss-cDNA ends (SLIC) (8,9) (Fig. 1b). The ligation reaction is catalyzed by the enzyme T4 RNA ligase, which has the ability to ligate single-stranded deoxyribonucleotides, as well as ribonucleotides (31,32). The anchor-ligated cDNA is then used directly in the PCR reaction, using a specific 3'-end primer and an anchor primer, as in the standard RACE protocols (6-9,12).

The anchor oligo must incorporate two modifications before it can be used for ligation. First, it must be phosphorylated at the 5'-end so that it is a substrate for the T4 RNA ligase enzyme (31). Phosphorylation can be accomplished either during oligonucleotide synthesis (8,12), or enzymatically by T4 polynucleotide kinase and ATP (6). Second, the anchor oligo must be blocked on its 3'-end to prevent the formation of oligo concatemers. This modification can also be incorporated during *in vitro* synthesis by using a blocked aminonucleotide analog as the final base (12), or a dideoxyribonucleotide can be added to the 3'-end in a TdT-mediated step, following oligo synthesis (8,9).

An additional requirement is that the oligo used for cDNA synthesis must have a 5'-hydroxyl group (i.e., the oligo has not been phosphorylated during or follow-

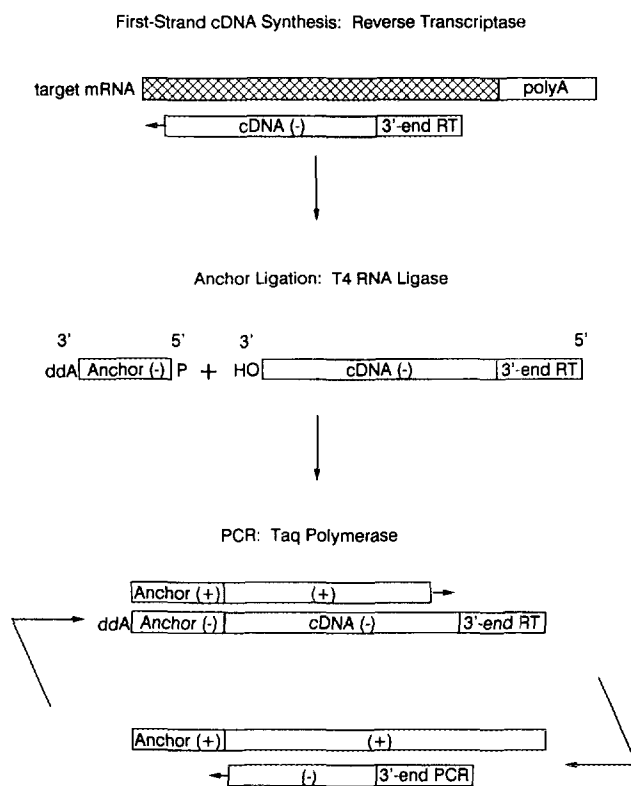


FIG. 1b. Major enzymatic steps in the ligation anchored PCR (LA-PCR) protocol. The target mRNA is first reverse transcribed using a specific 3'-end oligonucleotide primer (3'-end RT). A single-stranded DNA anchor oligo (Anchor(-)) is then ligated to the 3'-end of the cDNA using T4 RNA ligase. The Anchor(-) oligo has been modified by the addition of a phosphate group (P) to the 5'-end and a dideoxynucleotide (ddA) to the 3'-end. The anchor-ligated cDNA is then amplified by PCR, using a complementary-sequence anchor primer (Anchor(+)) as the 5'-end primer and a 3'-end primer (3'-end PCR) complementary to known cDNA sequence internal to the 3'-end RT primer. Arrows at the ends of cDNA (open rectangles) indicate active DNA synthesis in the indicated direction. Bent arrows indicate the many cycles of amplification in the PCR.

ing synthesis) so that the 5'-ends of cDNA molecules are not substrates for RNA ligase (6,8). Following cDNA synthesis, the RNA template is chemically degraded by treatment with sodium hydroxide at an elevated temperature. The base treatment hydrolyzes the RNA to ribonucleotides bearing 5'-hydroxyl and 3'- or 2'-phosphate groups, thus converting the RNA to a form that is not a substrate for T4 RNA ligase (7). After the above steps have been executed to remove or inactivate all competitive inhibitors of the desired ligation reaction, the only available substrates for the ligase enzyme are the phosphorylated 5'-ends of the anchor oligo and the 3'-hydroxyl groups of the cDNA molecules. The anchor-ligated cDNA is then used directly in a PCR containing the specific 3'-end primer and the anchor primer.

While the LA-PCR method avoids the generation of cDNAs which have been truncated during second-strand cDNA synthesis via the anchor(N)₁₅ annealing to an internal homopolymer in the first-strand cDNA, the other pathway for the generation of truncated 5'-ends remains uninhibited. Specifically, any incomplete first-strand cDNAs generated during cDNA synthesis will be substrates for T4 RNA ligase and will thus have the anchor oligo appended to their 3'-ends. PCR will then amplify these truncated cDNAs as discussed previously.

In contrast, RNA ligase-mediated RACE (RLM-

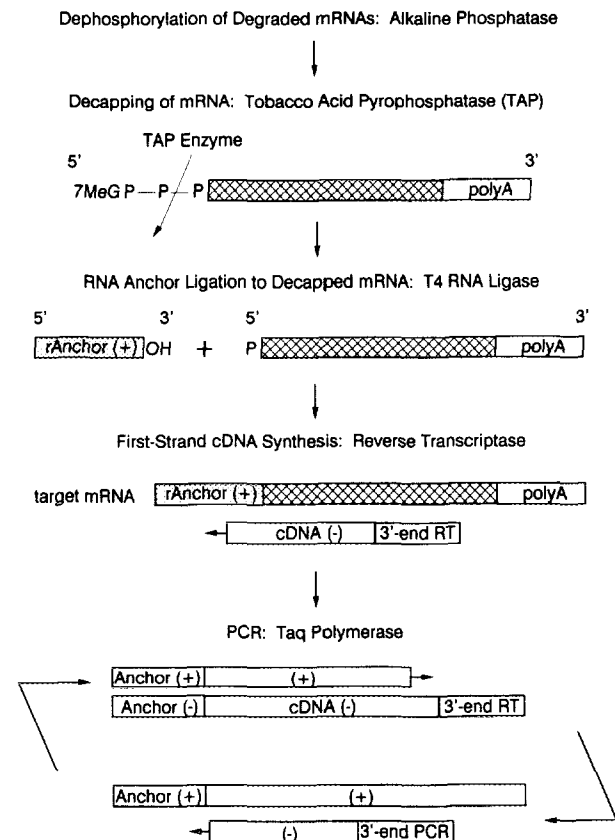


FIG. 1c. Major enzymatic steps in the RNA ligase-mediated RACE (RLM-RACE) protocol. Degraded mRNA and non-mRNA molecules are first dephosphorylated using an alkaline phosphatase (e.g., CIP). The 7-methyl guanosine triphosphate cap (7MeGP-P-P) is then removed by TAP, which leaves a phosphate (P) on the mRNA 5'-end. A single-stranded RNA anchor oligo (rAnchor(+)) is then ligated to the 5'-end of these mRNA molecules, using T4 RNA ligase. The anchor-ligated mRNA target molecules are reverse transcribed using a specific 3'-end oligonucleotide primer (3'-end RT). The resulting single-stranded cDNA is then amplified by PCR, using a DNA anchor primer (Anchor(+)) as the 5'-end primer and a 3'-end primer (3'-end PCR) complementary to known cDNA sequence internal to the 3'-end RT primer. The diagonal arrow from the TAP enzyme indicates the bond hydrolyzed. Arrows at the ends of cDNA (open rectangles) indicate active DNA synthesis in the indicated direction. Bent arrows indicate the many cycles of amplification in the PCR.

RACE) (10), also called reverse ligation-mediated PCR (RLPCR) (11) (Fig. 1c), can circumvent all mechanisms by which truncated 5'-ends are generated. In this method, an RNA oligonucleotide anchor sequence is ligated to the 5'-ends of the mRNA. The anchor-modified mRNA is reverse transcribed to generate first-strand cDNA which is then amplified by PCR, using a 3'-end specific primer and a (DNA) anchor primer (10,11,13,28,30). Since only those cDNAs which are extended to full-length will have the anchor sequence at their 3'-ends, truncated cDNAs will not be amplified in the PCR (10). The only way that truncated products could be generated would be via nonspecific internal annealing of the DNA anchor primer during PCR. Such artifacts can be avoided by careful design of primers and adjustment of PCR parameters (see section on PCR below).

In order to prepare the mRNA for RLM-RACE, the mRNA must be treated chemically (10) or enzymatically (11,13,28,30,33) to remove the 7-MeGppp cap which is covalently attached to the 5'-end of all mRNA transcripts (34). The chemical cap removal protocol utilizes the β -elimination method (35) in which the RNA is oxidized by sequential treatment with a sodium periodate/sodium dodecyl sulfate solution and glycerol. Oxidized mRNA is then treated with cyclohexylamine to remove the 7-MeG cap. The pyrophosphates which remain on the 5'-end following decapping are removed by treatment with CIP, and the mRNA is then rephosphorylated by T4 polynucleotide kinase and ATP, to make the 5'-end a substrate for polynucleotide kinase (5'-pyrophosphates are not a substrate for T4 RNA ligase) (31). The major shortcoming of the chemical decapping method is that degraded RNAs and non-mRNAs can be phosphorylated by polynucleotide kinase, creating unwanted substrates for the ligase reaction (10). The non-mRNAs will compete for the RNA anchor, potentially decreasing the efficiency of ligation to mRNAs. More detrimental, however, is the potential for degraded mRNAs to have an anchor ligated to their 5'-ends. When reverse transcribed, these truncated mRNAs will give rise to truncated cDNAs having a 3'-terminal anchor which will be amplified in the PCR (10).

The more elegant enzymatic method of decapping employs fewer steps and avoids the shortcomings of the chemical protocol (10). The RNA sample is first treated with an alkaline phosphatase to remove the 5'-phosphates from any degraded RNAs and noncapped (non-messenger) RNA species (11,13). After removal of phosphatase activity by heat inactivation and/or phenol/chloroform extraction (13), the mRNAs are decapped by treatment with the enzyme tobacco acid pyrophosphatase (TAP) which hydrolyzes the phosphoric acid anhydride bonds in the triphosphate bridge of the mRNA cap (33). The TAP enzyme thus leaves a 5'-

phosphate on decapped mRNAs. Thus, only 5'-mRNA ends which have a 7-MeGppp cap are converted to substrates for T4 RNA ligase. Truncated mRNAs are left with a 5'-hydroxyl group to which an anchor cannot be ligated. Thus, when cDNA is synthesized, only cDNAs which are extended to full-length using a nondegraded mRNA as a template will have an anchor sequence at the 3'-end (13).

The RNA anchor oligo can either be synthesized chemically (13,30) or transcribed *in vitro* from a linearized plasmid template having an appropriate promoter (e.g., T7, T3, or SP6) (10,28,29,36). In either case, the 5'-ends of the RNA oligos do not need to be modified following synthesis, since the *in vitro* transcription product has a 5'-triphosphate and the chemically synthesized oligo has a 5'-hydroxyl group, neither of which is a substrate for RNA ligase. The formation of oligo concatemers is thus prevented. The RNA anchor oligo can theoretically be any length which includes at least the minimal number of nucleotides to ensure specificity during the PCR reaction. A length of 30 nucleotides or less is suggested simply to limit possible secondary structures within the oligo and annealing of the oligo to partially complementary sequences in the mRNA population, which could contribute to decreased efficiency in the ligation reaction. The parameters for the design of PCR primers (see PCR below) should also be followed when designing the RNA anchor.

Once an appropriate oligo has been designed and synthesized, it is ligated to the phosphorylated mRNA using T4 RNA ligase (10,13,28,30,36). Maruyama and Sugano (13) and others (30,37) report that ligation efficiency can be enhanced by the addition of polyethylene glycol 8000 to a 25% concentration in the ligation reaction. Also, to block circularization of mRNAs, pCp can be ligated to the 3'-ends of the mRNA molecules before decapping (13,38), or RNA can be treated with sodium periodate to remove the 3'-hydroxyl group (28). Others have used the RLM-RACE protocol without taking this precaution (10,30,36).

Gorovsky and Liu (10) and Volloch *et al.* (29) have also applied the RLM-RACE method to the amplification and cloning of 3'-ends of transcripts. In this application of the RLM-RACE protocol, a DNA (10) or RNA (29) anchor oligonucleotide is used for ligation to the 3'-ends of mRNA molecules. Although DNA oligonucleotides are not efficient acceptors in RNA ligase-catalyzed reactions, they are good donors (31), and thus the more easily handled DNA anchor can be used instead of an oligoribonucleotide anchor. The DNA anchor oligo is simply phosphorylated by T4 polynucleotide kinase and ligated to the 3'-mRNA ends using T4 RNA ligase. Reverse transcription is primed by another DNA oligonucleotide which is exactly complementary to the ligated anchor, and PCR is performed using this anchor-complement as the 3'-end primer with a specific

5'-end primer which is complementary to known cDNA sequence (10,29).

Because the 3'-anchor is ligated to the mRNA tail, the entire mRNA tail is preserved in the resulting RACE product. Thus, in addition to being able to unambiguously determine that the entire 3'-end of the transcript has been cloned, biochemical information is obtained concerning the 3'-end of the transcript. The interested investigator can determine the length of the poly(A) tail from individual clones and thus the variance in that length within the population. Also, one can determine whether canonical polyadenylation signals are used in a given transcript (10,29).

PCR Amplification

While an exhaustive exploration of techniques to optimize the parameters of the PCR is well outside the scope of this Review, certain parameters which simplify and enhance the reproducibility of the PCR will be discussed. Also, a small amount of literature will be reviewed which describes methods by which some of the more commonly encountered problems in PCR may be overcome. For further information, see the review on PCR strategy by Arnheim and Erlich (1), general discussions of PCR optimization by Innis and Gelfand (2), and PCR-related laboratory technique by Kwok (39), and a report discussing decontamination of PCR reagents by Sarkar and Sommer (40).

Fidelity concerns. One very important issue which should be considered in the PCR amplification of cDNAs is polymerase fidelity. Since the RACE technique is designed for the determination of unknown cDNA sequences, it is important that precautions are taken to minimize the introduction of mutations during the PCR. There are now commercially available thermophilic polymerases which have 3' → 5' proofreading exonuclease activity, such as pfu (Stratagene) and Vent (New England Biolabs). Of these two enzymes, pfu has the highest fidelity, measured at 1.6×10^{-6} mutations per nucleotide (21). In contrast, the frequency of nucleotide misincorporation by *Taq* polymerase has been reported to be 2.5×10^{-5} or 12-fold higher than pfu (21). The pfu enzyme is also characterized by extreme thermostability, remaining 95% active after 60 min at 95°C (21). Thus, longer and higher temperature denaturation cycles can be performed without losing significant polymerase activity.

In a recent paper by Barnes (41), a mixture of pfu and KlenTaq1 (a genetically modified *Taq* polymerase) at a 1:15 ratio was found to have at least as high a fidelity as pfu by itself, and much longer amplifications were possible than with either enzyme alone. Barnes speculates that the quantity of pfu in the reaction mix is sufficient to remove mismatched bases added by the *Taq* derivative. He further speculates that PCR ampli-

fications are normally limited to templates of less than 10 kb because *Taq* polymerase misincorporations prematurely terminate primer extensions. Thus, when molecules of 10 kb and greater are the targets of the PCR, such misincorporations occur too frequently for the desired product to be efficiently amplified. The addition of the proofreading activity of pfu polymerase is sufficient to overcome that barrier (41,42).

In order to enhance the product yield and specificity of PCR, protocols have been designed in which template (43), template and polymerase (44), or biochemically active enzyme (45) are not added to the other PCR components until the reagents have been prewarmed to a temperature above the primer annealing temperature. This technique, called "hot start," is assumed to improve the specificity of the primer:template interaction by preventing the nonspecific annealing of primer to previously denatured templates during the room-temperature assembly of reagents (43,45). *Taq* polymerase can extend templates at the slow but significant rate of approximately 15 nucleotides per minute at room temperature (21,46), and thus nonspecifically annealed primer:template or primer:primer pairs can be extended during assembly of reagents and further amplified during PCR, resulting in a high background of nonspecific products (43,45).

Since the pfu enzyme has negligible polymerase and exonuclease activity at low temperatures (21), all reaction components can be assembled on ice (or at 4°C in the thermocycler) without worry of extension of nonspecifically annealed primers or degradation of primer and template by the 3' → 5' proofreading exonuclease activity of the enzyme. Due to the extreme thermostability of pfu, an initial denaturation of 98°C for 5 min can be performed without significantly diminishing enzyme activity (unpublished observations), although long denaturations may result in damage to the PCR template (see Thermocycling Parameters, below). The entire reaction mix can thus be denatured under very stringent conditions, then immediately cooled to the appropriate temperature to allow primer annealing followed by primer extension. Thus renaturation of the template and nonspecific primer annealing should both be minimized. The method of denaturation in the presence of the pfu polymerase allows the generation of specific products with minimal contamination by nonspecific template extension products and primer-dimers (unpublished observations).

Primer design. Primer design is another parameter which has a large impact on the outcome of the PCR. There are many "rules of thumb" followed by users of the PCR technique which have no basis in experimental data (14). There are, however, a few basic rules and simple calculations which allow the investigator to successfully design primers which will yield the desired

PCR product with little to no background, as is thoroughly discussed by Rychlik (14) and briefly summarized here.

The need for well-designed primers is somewhat dependent on the length of the target molecule to be amplified. Targets of 200 to 400 bp in length are amplified most efficiently in PCR, and thus many poorly designed primers will still generate the desired product within this size range (14). Since the size of the desired full-length RACE product is very often not known before PCR, primers should be designed to conform to as many parameters of optimization as possible. First, primer pairs with 3'-terminal complementarity must be avoided to prevent the formation of primer-dimers. Even a few complementary bases at the 3'-ends of the primers can be enough to generate primer-dimers, since *Taq* and related polymerases extend primers rapidly (high processivity). Also, primers with self-complementarity involving their 3'-ends must be avoided, since such primers will be extended using themselves as a template, instead of the target molecule. Additionally, the 3'-ends of all PCR primers should be designed to be *unstable* to minimize the probability of priming from nontarget sites. At the target site, hybridization stability is provided by the 5'- and midregions of the primer, and the primer is successfully extended. In contrast, where the primer has short regions of complementarity with nontarget sequences, the unstable 3'-end is not likely to anneal nonspecifically and be extended before the brief hybrid formation event has ended. Rychlik has constructed a table in which free energy values for DNA duplex formation based on nearest neighbor nucleotide analysis (14,47) are given. These values can be used to estimate DNA duplex stability for any given sequence. Rychlik suggests that the five 3'-terminal bases of a given primer not have a stability greater than -9 kcal/mol, which generally means that there are no more than two G or C residues in the last five bases (14).

Rychlik also suggests that primer length be kept short (16–18 nucleotides) for targets less than 500 bp, while longer targets (5 kb) should be amplified by longer primers of about 24 nucleotides (14). While these guidelines may be difficult to follow when primers with cloning adaptors are used, the region of primer complementarity to the target molecule can be kept within the 16–18 nucleotide range. Primer GC/AT ratio should be as close to the GC/AT ratio of the template as possible (14). While the majority of the base pair composition of a RACE target may be unknown, the GC/AT ratio can be estimated from the known cDNA sequence or from the GC/AT ratio in the target genome, if such statistics are available.

Thermocycling parameters. Thermocycling parameters also have a dramatic influence on the degree of

success achieved in any given PCR amplification (1), particularly as the target size enters the several kilobase range (41,48). The denaturing cycle must be hot enough and long enough to successfully denature at least a subset of target molecules. Theoretically, maximally efficient PCR can only be achieved if all target molecules are denatured in each denaturation cycle. However, an important consideration is that the rate of DNA depurination increases dramatically as temperature is increased and pH is decreased (49). During PCR, temperatures of 94°C or more are reached while a Tris-based PCR buffer which is pH 8.3 at 25°C will drop to pH 6.1 at 95°C. *Taq* polymerase is not able to continue primer extension when a depurinated site is encountered (41). By the calculations of Barnes (41), there is reason to believe that depurination occurs at a rate which is significant for the PCR amplification of targets in the kilobases size range (41) or perhaps even smaller targets (50). Thus, denaturation times should be limited to what is minimally required for successful PCR. Denaturation cycles of as short as 10 s at 94°C have been used to successfully amplify targets of 20 kb and greater (41,48), although the inclusion of cosolvents such as glycerol and DMSO improved yields in some of these experiments, presumably by lowering the strand-separation temperature (48). Cheng *et al.* (48) also note that the detrimental effects of high temperature in the PCR reaction may be reduced by using a PCR buffer which suffers less dramatic changes in pH during the denaturation cycle. If a buffer having a temperature coefficient lower than that of Tris is used (tricine, bicine, or EPPS, for example) (51), the pH will be held higher during denaturation, and depurination should be reduced (48).

There are numerous formulas available to calculate the ideal hybridization temperature for given PCR oligo pairs (2,25), and the calculation of ideal hybridization temperature reportedly requires nearest neighbor nucleotide analysis of the template, as well (52). However, Rychlik states (14) that if the primers are designed properly, incorporating 3'-terminal instability (as discussed above), the optimal temperature range for primer annealing is quite broad. We have found that even the most basic estimate for annealing temperature, $\{2*(A/T) + 4*(G/C)\} - 5 = T_{\text{hyb}}$, works well with most primer pairs. In this formula, the primer annealing temperature is calculated by assigning 2°C for each A or T and 4°C for each G or C. The hybridization temperature is then calculated as 5° below the annealing temperature of the primer with the lowest annealing temperature (2). It is best to use primers with matching annealing temperatures whenever possible. Cheng *et al.* (48) report that when primers with an optimal calculated annealing temperature of 68°C are used, a two-temperature thermal cycling profile with a 94°C denaturation and a 68°C annealing/exten-

sion works well. The extension step should generally include 1 min per kilobase in the amplification target (1), although for targets over a few kilobases, longer extensions may be required (41,48). For RACE amplifications in which the target size is unknown, the maximal extension time required can be estimated from the size of the mRNA transcript as determined by Northern blotting.

Increasing specificity. Unless the target cDNA for RACE cloning is abundant, a single PCR amplification may be insufficient to generate the desired product at a yield sufficient for cloning. In the RACE technique, there are many nontarget cDNAs generated which have an anchor sequence at the 5'-end. Thus, compared to PCR using two primers with unique specificity, the RACE PCR amplification has a reduced efficiency because a considerable quantity of anchor primer is consumed by annealing to and extension of these nontarget cDNA sequences. The accumulation of nontarget single-stranded cDNA increases the probability that the specific (3'-end) primer will anneal nonspecifically and be extended by polymerase, thereby generating a substrate which will compete in future cycles for nucleotides, polymerase molecules, and both primers.

However, the products of the first PCR can be purified away from remaining primers (see below) and subjected to another PCR amplification employing the anchor primer and a second specific 3'-end primer which is complementary to a known cDNA sequence within the region amplified by the first 3'-end PCR primer (4,5,23). Frohman (23) reports that if the number of cycles in the two PCR amplifications totals 60, the desired RACE product should appear as a specific ethidium staining band on an agarose gel.

Another method which has potential for enhancement of specific PCR product yield is referred to as affinity capture (53). There are a number of variations of this protocol, but the basic method involves the use of an appropriately tagged oligo to capture the desired products on an affinity matrix which specifically binds the tag. For example, a biotinylated oligo complementary to a target cDNA sequence is hybridized in solution to denatured cDNA, and hybrids are captured on a streptavidin affinity matrix. Appropriately stringent washes (53) are then used to remove nonspecifically annealed hybrids. The cDNA, now significantly enriched for the specific target sequence, is eluted from the column and amplified by PCR (53). Although RACE protocols which incorporate affinity capture were not found during the preparation of this Review, the removal of a substantial amount of contaminating sequences by affinity capture could dramatically enhance the yield of the RACE PCR.

Cloning

Once product is obtained at a detectable level, purification, recovery, and cloning must be achieved. Since

a considerable percentage of PCR product is often lost during recovery via phenol/chloroform extraction followed by ethanol precipitation (27), it is best to use kits specifically designed for the recovery of PCR products. The Wizard PCR Preps (Promega) system is rapid and gives high yields of PCR products of at least 200 bp in length, while effectively removing primers, primer-dimers, and free nucleotides (54). The purified product can then be cut with the appropriate restriction endonucleases, size selected in low-melting point agarose, and ligated directly in the agarose (25) to a sequencing vector.

Although protocols for the direct sequencing of RACE products have been described (5,23), there is often heterogeneity at the 5'-ends of cDNAs, due to alternative transcription start sites (10,13) and non-template-directed addition of nucleotides to the 3'-ends of first-strand cDNAs by reverse transcriptase enzymes (23,24,55). RACE products should thus be cloned, and several individual clones should be sequenced.

The cloning of RACE products is facilitated by incorporating restriction endonuclease sites into the sequences of the anchor and specific 3'-end oligos. Since the anchor primer is not complementary to specific transcript sequences, the anchor sequence can easily be designed to incorporate three or four 6-bp restriction endonuclease recognition sequences. In contrast, the 3'-end PCR primer must be complementary to a known cDNA sequence, so an extragenomic endonuclease recognition sequence usually must be included on the 5'-end of the oligo. The choice of restriction sites is extremely important and will dramatically influence the efficiency of the subsequent ligation to the sequencing vector. Enzymes which generate sticky ends with a 4-bp overhang are ligated much more efficiently than blunt ends or ends with shorter overhangs (56). Restriction sites for enzymes which demonstrate marked site preferences (56) or which are recut inefficiently following ligation (56) should be avoided entirely. Additionally, several restriction endonucleases cut poorly or not at all when their restriction recognition sites are placed close to the end of a DNA molecule (56). As a general rule, the restriction endonuclease site should be preceded by a minimum of four bases at the 5'-end of the PCR oligo. Even with four or more extra 5'-end bases, however, some enzymes (*NotI*, *NdeI*, *BstXI*, etc.) still cut poorly. Upon further limiting the choice of enzyme recognition sites to those which have compatible sites for ligation in common sequencing vectors, the list of remaining sites is actually quite small. *BamHI*, *BglII*, *EcoRI*, *KpnI*, *PstI*, *SacI*, and *XbaI* (New England Biolabs) work well for digestion and subsequent ligation of PCR products. If other sites are necessary or desired in a particular cloning strategy, the method of Jung *et al.* (57) can enhance the cleavage efficiency of PCR-primer cloning sites. In this method, sites at the ends

of PCR primers are converted into internal restriction sites by end-to-end ligase-mediated concatemerization of the PCR products, thereby making the restriction site a better substrate for enzyme cleavage.

A final but crucial concern is that the restriction sites chosen for cloning may be present within the sequence of the RACE product as well. When designing primers, it can be helpful to consider the relative frequency of candidate recognition sequences in the target genome, if this information is available. Once the primers have been designed and the RACE product has been generated, cut and uncut product can be electrophoresed side by side on an agarose gel to determine if there are internal restriction sites in the RACE product. If there is an internal site corresponding to the enzyme used to cut the anchor primer, the product can instead be cut with an enzyme specific for one of the two or three other restriction sites encoded in the anchor. If, however, there is an internal site corresponding to the unique enzyme site incorporated in the specific 3'-end PCR oligo, the simplest alternative would be to cut the RACE product with an enzyme which recognizes a rare site within the previously known sequence included in the RACE product. In the very likely event that there are no rare restriction sites in the known region of cDNA sequence, the 3'-end PCR primer can be redesigned to incorporate a different restriction endonuclease site. Aliquots of the RACE product can be digested with candidate endonucleases until one is found which does not cut, and the new primer can be designed to include this site. Then, an aliquot of purified RACE product can be reamplified with the anchor primer and the new 3'-end primer for a small number of cycles to incorporate the new oligo with the new restriction site. The RACE product may then be cloned as discussed above.

Methods for the cloning of PCR products have also been developed which do not require the use of restriction endonucleases. Many, if not all, DNA polymerases have an evolutionarily conserved enzymatic activity which results in the non-template-directed addition of one or several nucleotides to the 3'-end of the DNA molecule, beyond the final nucleotide of the template (55). Specifically, *Taq* polymerase adds a single non-template-directed adenosine to 3'-DNA ends at very high frequency (55). This property was exploited by several different groups who described essentially identical cloning strategies in which *Taq*-synthesized PCR products are directly ligated into linearized vectors having a single thymidine overhang at both 3'-ends (58–60). The site of PCR product ligation interrupts the coding sequence of the β -galactosidase gene, allowing transformants with insert-containing plasmids to be distinguished from empty vector (there is a background of vector religation, despite the fact that the thymidine overhangs on the vector ends are not

compatible cohesive ends) by blue/white color selection of colonies plated on media containing IPTG and X-gal. Vectors with restriction sites engineered to generate the thymidine single base-pair overhangs have been described (58) and are available commercially [Invitrogen]. Alternatively, pBluescript [Stratagene] (or any other cloning vector with a unique restriction site for a blunt end-generating endonuclease within the lacZ gene) can be linearized with *Sma*I or *Eco*RV and then modified by incubation with *Taq* polymerase and dTTP at 70°C for 2 h (59), or with terminal transferase and ddTTP at 37°C for 1 h (60), thereby generating a blue/white-selectable vector having the appropriate thymidine overhangs.

Taking a slightly different approach, Liu and Schwartz (61) have described a high-efficiency method for the blunt-end cloning of PCR products. Following PCR with *Taq* polymerase, the PCR product was incubated with the Klenow fragment of DNA polymerase I at 22°C for 30 min, resulting in the blunting of the PCR product ends (via the 3' to 5' exonuclease activity of the Klenow enzyme). The blunted PCR product was then mixed with *Sma*I-digested pBluescript [Stratagene] in the presence of T4 DNA ligase and the *Sma*I enzyme. The *Sma*I enzyme thus recuts any vector molecule which recircularizes without an insert, whereas vectors containing inserts are protected from *Sma*I digestion (as long as the primer ends do not specify *Sma*I half-sites). The only potential caveat with this method is that a PCR product containing a *Sma*I site(s) will also be cut during ligation, and will thus not be clonable by this protocol. A kit employing an improved version of this method is now commercially available [Stratagene] which, in place of *Sma*I, utilizes the blunt-cutting endonuclease *Srf*I which has an infrequently encountered 8-bp recognition sequence, dramatically decreasing the probability that the PCR product will be cut in the ligation reaction.

It has been shown that thermostable DNA polymerase Vent [New England Biolabs], which has a 3' to 5' proofreading exonuclease activity, generates PCR products with greater than 95% blunt ends (56). This is presumably also true of pfu [Stratagene] and other polymerases with 3' to 5' proofreading activity, and thus the PCR products of such polymerases are excellent substrates for blunt cloning, without further modification (62). The blunt cloning strategy should be the most efficient restriction endonuclease-independent method of cloning the PCR products of proofreading polymerases (62), although it has been reported that cloning into vectors with thymidine overhangs (see above) also works well with these enzymes (58).

Several groups have reported similar methods for the cloning of PCR products independent of both restriction endonucleases and DNA ligase, via the generation of PCR products and vectors having long (ca. 12 bp) com-

plementary single-stranded ends (63–66). However, to achieve the small advantage of avoiding an enzymatic ligation step, the authors employ longer PCR primers, an additional set of PCR primers, and at least one additional PCR amplification. Because these protocols do not seem to offer any clear advantage to the cloning methods outlined above, they will not be discussed further. The interested investigator is referred to the references cited above, particularly the chapter by Shuldiner and Tanner (65), which includes a detailed protocol.

Once the RACE product has been cloned into the desired sequencing vector and transformed into an *Escherichia coli* host, colonies can be screened for positives by colony hybridization (25) using the previously known region of the RACE product as a probe. If the RACE product has been amplified in two rounds of PCR using nested 3'-end primers, colony screening is not generally necessary since the majority of clones should contain a specific RACE insert. Ideally, the sequencing of several RACE clones will reveal both the nucleotide sequence of the 5'-end (or 3'-end) of the target mRNA and give convincing evidence that the entire end sequence has been obtained.

A major problem encountered when using standard RACE or LA-PCR for cloning 5'-end sequences is that truncated RACE products cannot generally be distinguished from full-length RACE clones, since truncated first-strand cDNAs are terminated in an anchor(N)₁₅ (standard RACE) or ligated anchor (LA-PCR), as are the full-length cDNAs (discussed above). There have been reports that the 5'-terminal 7-MeGppp cap can be reverse transcribed fairly efficiently during first-strand cDNA synthesis (67,68). The authors of these reports assert that the presence of a nongenomic cytidine at the 3'-end of a first-strand cDNA can be interpreted as evidence that the cDNA is full-length (67,68). Thus, when using the standard RACE protocol, a nucleotide other than dCTP should be used in the TdT tailing step so that the nongenomic first-strand 3'-terminal cytidine can be recognized (67). In addition to the obvious limitation that the genomic sequence in the vicinity of the putative transcription initiation site must be known to determine whether the 3'-terminal cytidine is nongenomic, reverse transcriptases often add one or several bases at the 3'-end of cDNAs in a template-independent fashion (23,24,55). In fact, AMV reverse transcriptase has been shown to add, with equal efficiency, a single template-independent adenosine or cytidine (but not a guanine or thymidine) to the 3'-end of DNA molecules (55). It may thus be unwise to assume that a nongenomic cytidine added to the 3'-end of the first-strand cDNA is the result of reverse transcription of the cap structure. As discussed above, we recommend using the RLM-RACE method and isolating several independent clones from at least two sepa-

rate PCR amplifications to assess whether the complete 5'-end of the transcript has been cloned.

OTHER APPLICATIONS OF RACE

Although most RACE protocols have been designed with the cloning of full-length cDNA ends as the goal, some investigators have modified the basic protocol to clone neighboring internal sequences. Confronted with the problem of obtaining additional 3' sequence from the 21-kb titin mRNA, Fritz *et al.* (69) devised a RACE technique to clone unknown neighboring internal sequences for use when the poly(A) tail is too distant to allow cloning by standard 3' RACE methodology (24). Their modification is simply to use an anchor oligo terminating in six random nucleotides (Anchor(N)₆) for cDNA synthesis, instead of the Anchor(dT)₁₅. The cDNA is thus randomly primed. The specific target is amplified using a specific 5'-end primer and the anchor primer, as in the standard 3' RACE protocol. The resulting product contains additional target transcript sequences 3' to the original clone. This process can be repeated with new 5'-end primers to "walk" the length of the transcript.

Struck and Collins (70) have used a similar method to clone unknown neighboring 5' sequences which involves using a specific 3'-end primer and a spurious (i.e., any) 5'-end primer. Even though the 5'-end primer is not complementary to target transcript sequences, a small number of amplification products will be generated by nonspecific priming of the target cDNA by the spurious primer. The first PCR thus results in the incorporation of the spurious primer in a few specific amplification products. The products of this initial reaction are amplified in a second PCR, employing the same spurious 5'-end primer, and using a second specific 3'-end primer, internal to the sequence of the first. This PCR results in very efficient amplification of the few target molecules from the first PCR which have incorporated the spurious primer. The amplification products contain novel sequence 5' of the original partial cDNA terminus.

Whitcomb *et al.* (71) have devised a method which they call random-primed/anchored-PCR (RPA-PCR) to generate nested deletions from a target sequence which has been cloned into a plasmid vector. Denatured plasmid DNA is hybridized with an anchor(N)₁₀ random primer, and these primers are extended by T7 polymerase. After excess primers and short extension products have been removed, the ssDNA is amplified with the anchor primer and a primer specific for the constant end of the desired nested deletion products. The PCR products are then separated on a low-melting-point agarose gel, and fragments of the desired size are isolated and cloned into a plasmid vector. In this manner, clones with nested deletions at one end, which are

evenly spread over at least 2 kb, can be isolated from a single PCR reaction (71).

Although most investigators use RACE to avoid further screening of a cDNA library, some have found the basic methodology useful for the generation of general cDNA libraries from very small numbers of cells. Belyavsky *et al.* (72) generated a library containing 10^6 clones from total RNA isolated from only 10 myeloma cells. First-strand cDNA was primed with an anchor(dT)₁₆ primer and homopolymer tailed with dGTP using terminal transferase. The cDNA was then amplified with an anchor 2(dC)₁₃ primer and the anchor(dT)₁₆ primer to generate the library. Inserts cloned and recovered were up to 1.5 kb in length and representative of high-, moderate-, and low-abundance mRNA species. Domec *et al.* (73) used the same approach to achieve equally good results. The SLIC/LA-PCR method has also been successfully used to generate cDNA libraries (8,9). Others have followed conventional cDNA synthesis and linker ligation with PCR amplification to generate libraries from small amounts of starting material (74).

An alternative to generating anchor(dT)_N-primed libraries is to prime the cDNA with random hexamers. One can then add a homopolymer tail to the cDNA using TdT (75), or ligate a single-stranded DNA anchor to the 3'-ends using T4 RNA ligase (12). The resulting single-stranded material is a library of 5'-ends, from which any target 5'-cDNA end can be amplified using standard RACE or LA-PCR/SLIC, respectively. Although amplification of random-primed cDNA may generate more background than PCR of cDNA made by specific priming, random-primed libraries have been shown to be a useful source of specific RACE products (12,75).

RLM-RACE CLONING OF 705-bp GC-RICH 5'-cDNA END

In order to evaluate the utility of the RLM-RACE/RL-PCR method, the 5'-end of a GC-rich (59.4%), low-abundance Epstein-Barr virus (EBV) mRNA was cloned from the Burkitt's lymphoma cell line, Rael, using a protocol adapted principally from Maruyama and Sugano (13) (see Appendix). The RNA anchor oligo was designed to have no stable internal secondary structures at 37°C, as assessed by the MFold (76) sequence analysis program. The final three base pairs (AAA) are a "tag" sequence not included in the RLM-RACE Anchor 2 oligo. Thus, unwanted PCR products which have been derived from spurious hybridization of the RLM-RACE Anchor 2 to a nonanchor sequence will not include this tag, and can be identified during sequencing (Table 1). The oligo was chemically synthesized (77), deprotected (77), size-fractionated by PAGE (25), and purified by reverse-phase chromatography on a Sep-

Pak C18 column [Waters]. DNA anchor oligos (Table 1) were designed according to the specifications of Rychlik (14). The use of two nested DNA anchors for the RACE PCR amplifications results in enhanced specificity (4,5,23).

The target mRNA is a highly spliced 4.8-kb transcript of unknown function which appears to be expressed in all cells latently infected by EBV. The 5'-end of this mRNA was previously identified by screening a nasopharyngeal carcinoma (NPC) cDNA library (78), but no information regarding the 5'-end structure of the 4.8-kb mRNA in Burkitt's lymphoma (BL) cell lines has been published. The strategy was to generate and PCR amplify the 5'-end of the cDNA using three nested primers located ca. 80–100 bp downstream of the mRNA exon 5 splice acceptor (Fig. 2a, Table 1). Details of the RLM-RACE procedure used to amplify the 5'-end can be found in the protocol (see Appendix).

Initially, the primary and secondary RLM-RACE products were analyzed by Southern blotting (25) with a ³²P-labeled exon 5 probe (Fig. 2b, lanes A and B). Numerous discrete products distributed over a range extending up to ca. 1100 bp were observed in the primary PCR, but amplification of the smallest products was favored at the expense of larger species in the secondary PCR. Since there was a prominent hybridization signal at ca. 750 bp (band 4; the approximate size anticipated if the 4.8-kb Rael BL transcript has the same structure as the NPC transcript), primary RLM-RACE material was size-selected for 700- to 800-bp species and reamplified according to the secondary PCR parameters (see Appendix). This amplification resulted in a prominent ethidium-staining 734-bp product (Fig. 2b, lane E), which was cloned, sequenced, and found to have an exon structure (Fig. 2a) identical to that reported for the major NPC 4.8-kb transcript (78). The 5'-ends of 19 clones were sequenced, identifying two discreet start sites (Fig. 2c). These sites have been confirmed to be the major start sites by S1 nuclease analysis (data not shown). As a comparison, standard RACE methodology was used to attempt to clone this same 5'-end region. The longest specific products identified by Southern blotting in either the primary or secondary PCR were truncated cDNAs which extended only into the middle of exon 4, as assessed by cloning and sequencing (data not shown).

A control RT-PCR was performed using the Ex5-cDNA oligo for reverse transcription and the Ex5PCR 1 oligo paired with a specific 5'-end exon 1 oligo (Table 1). The products of this control PCR and the RLM-RACE products were analyzed by Southern blotting using an exon 1 probe (Fig. 2b, lane G). Eight discreet species were detected in this analysis and designated 1–8, from highest to lowest molecular weight, and these have been correlated with RLM-RACE products having a similar size (the corresponding RLM-RACE

TABLE 1
RLM-RACE Oligonucleotides

Oligo ^a	Oligo Sequence	T _{hyb} (°C)
rAnchor (RNA)	AGAGUCUUGACAAGCUUAGGUACCAGUAAA	N/A
RLM-RACE Anchor1 ^b	TGATCGTAAGAGTCTTGACAAGCT	41
RLM-RACE Anchor2 ^b	<u>AGATGAATT</u> <u>CAAGCTT</u> <u>AGGTACCA</u> GT	45
Ex5cDNA	CTTCGATATCGAGTGTCTG	56
Ex5PCR 1	ACCAACGAGGCTGACCTGA	60
Ex5PCR 2	ACCTGATCCTGCAGATATC	56
Ex1PCR (5'-end)	AGCAACATGGAGACCTCGAA	60

^a Oligos were synthesized by the Dana-Farber Cancer Institute Molecular Biology Core Facility.

^b T_{hyb} calculated for nucleotides homologous to rAnchor sequences are underlined.

products are expected to be 69 bp larger). Most of the larger species appear to be represented in the primary PCR at a relative abundance similar to that seen in the control (Fig. 2b, lanes C and G). The very abundant low-molecular-weight species of the unselected secondary RLM-RACE amplification, which gave intense hybridization signals with the exon 5 probe, were not detected by the exon 1 probe (Fig. 2b, lanes B and D). These species were cloned and found to be partial cDNA clones resulting from RNA anchor ligation to partially degraded RNA molecules, demonstrating that CIP pre-treatment of mRNA is not 100% effective. However, only a few distinct species were identified by cloning (i.e., in screening many clones, few unique 5'-ends were identified), suggesting that the population of truncated cDNA species having a terminal anchor sequence is small compared to the population of full-length species. Titration of cDNA to high dilution might thus be ex-

pected to eliminate truncated species from the PCR (see protocol notes in Appendix).

The RLM-RACE/RL-PCR method is thus an effective method of cloning full-length cDNA ends. Although

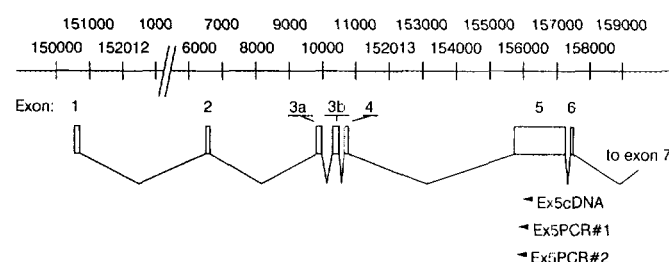


FIG. 2a. Exon structure of the 5'-end of the major BL 4.8-kb RNA as determined by RLM-RACE cloning. Clones containing exons 1, 2, 3a, 3b, 4, and the 5'-end of exon 5 were obtained from the BL cell line Rael using the RLM-RACE method. EBV genome coordinates are indicated along the top of the figure (the idiosyncratic numbering of the EBV genome results in the interruption of nucleotides 152,012 and 152,013 by nucleotides 1–11,835). Nucleotides between 1000 and 6000 have been deleted from the figure for aesthetic reasons. The location of cDNA and 3'-end PCR primers are indicated below exon 5. The exon nomenclature reflects that used for the identical NPC transcript (78).

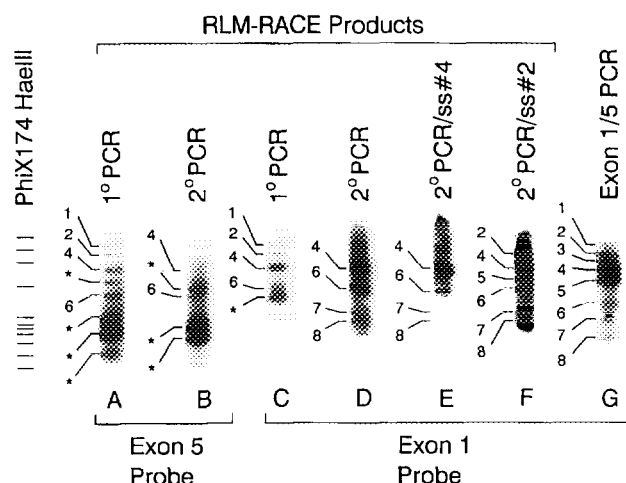


FIG. 2b. Southern hybridization analysis of RLM-RACE amplification products. Products of the RLM-RACE primary PCR amplification (lanes A and C), secondary PCR amplification (lanes B and D), secondary PCR amplifications of primary PCR product which had been size-selected for band 4 (lane E) or band 2 (lane F), and a control exon 1/exon 5 PCR amplification (lane G), were analyzed by hybridization with an exon 5 (lanes A and B) or exon 1 (lanes C–G) probe. Distinct PCR products in the control amplification (lane G) are numbered consecutively (1–8) from highest to lowest molecular weight. RLM-RACE products (lanes A–F) were assigned the same number if they migrated with a corresponding mobility and hybridized to the exon 1 probe. Products marked with an asterisk (*) are truncated cDNAs (see text) which hybridize only with the exon 5 probe (lanes A and B) or a product of unknown origin (lane C) which was not amplified in the secondary PCR (lane D). PCR products were separated by electrophoresis through 1.5% LMPA agarose, transferred to a nylon membrane, and hybridized with ³²P-labeled dsDNA probes corresponding to exon 5 or exon 1 sequences. The migration of the molecular weight standard (PhiX174/HaeIII digest) is indicated at the extreme left.

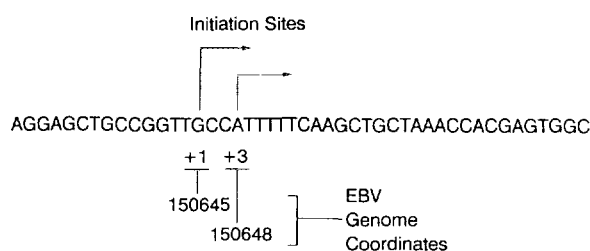


FIG. 2c. Transcription initiation sites of the major BL 4.8-kb RNA as determined by RLM-RACE cloning. Two sites of transcription initiation were indicated by sequencing 19 RLM-RACE clones which have the structure shown in Fig. 2a. The +1 initiation site was observed in 13 clones, and the +3 initiation site was observed in the remaining 6. These start sites were confirmed to be the major sites of initiation by S1 nuclease mapping (data not shown). The same sites of initiation were identified for the NPC transcript (78), as well as a third site at -4.

clones having truncated ends can be generated, full-length cDNA templates probably form the majority of the preamplification population. Empirical optimization of PCR conditions (see Appendix) should thus be expected to result in exclusive generation of full-length species in most cases.

APPENDIX

RLM-RACE Protocol for Cloning of cDNA 5'-Ends

The protocol can easily be adapted to the cloning of 3'-ends using the modifications described by Liu and Gorovsky (10).

RLM-RACE Protocol

A. Dephosphorylation of degraded RNAs and non-mRNAs

1. Resuspend 1–5 μ g of poly(A)-selected RNA in 20 μ l of DEPC-water. Heat to 55°C for 5 min and resuspend thoroughly with a pipetman to ensure that all RNA enters solution.

2. Add:

10 $\times$ CIP buffer	5.0 μ l
CIP enzyme (2 U)	0.2 μ l
RNAsin (50 U)	1.25 μ l
DEPC-water	23.5 μ l

Gently pipet mixture up and down several times to mix.

3. Incubate at 37°C for at least 1 h (see notes).

4. Add:

DEPC-water	250 μ l
10 M ammonium acetate	75 μ l
Phenol:chloroform, 1:1	375 μ l

Vortex at high speed for 15–30 s. Spin at 12,000g for 5 min.

5. Recover aqueous phase and repeat extraction with an equal volume of phenol:chloroform, 1:1.

6. Recover aqueous phase and add 900 μ l of absolute ethanol. Mix by inverting several times, then spin at 12,000g for 20 min.

7. Pour off supernatant, add 400 μ l of 70% ethanol, vortex briefly, and spin at 12,000g for 5 min.

8. Pour off supernatant. Remove residual liquid with a pipetman, taking care not to dislodge pellet. Air dry briefly on bench.

B. Enzymatic decapping of mRNA molecules.

1. Resuspend RNA pellet in 20 μ l of DEPC-water, as described above (step A.1).

2. Add:

10 $\times$ TAP buffer	5.0 μ l
TAP enzyme (2.5 U)	0.5 μ l
RNAsin (50 U)	1.25 μ l
DEPC-water	23.5 μ l

Gently pipet mixture up and down several times to mix.

3. Incubate at 37°C for 1 h.

4. Follow step A.4 for phenol:chloroform extraction.

5. Follow steps A.6 through A.8 for ethanol precipitation (precipitate only 1 μ g of decapped RNA. The remainder can be stored at -70°C for further experiments).

C. Ligation of oligoribonucleotide to mRNA 5'-ends.

1. Resuspend RNA pellet in 4 μ l of DEPC-water, as described above (step A.1).

2. Add:

10 $\times$ RNA ligase buffer	5.0 μ l
RNA oligo (2 μ g)	10.0 μ l
T4 RNA ligase (100 U)	5.0 μ l
RNAsin (50 U)	1.25 μ l
40% polyethylene glycol	25.0 μ l

Gently pipet mixture up and down several times to mix.

3. Pulse reaction briefly in microfuge to force liquid to bottom of tube and eliminate large bubbles. Overlay with mineral oil.

4. Incubate at 17°C for 16 h.

5. Transfer ligation mix to a fresh tube, taking care not to transfer any mineral oil.

6. Follow steps A.4 and A.5 for phenol:chloroform extraction.

7. Follow steps A.6 through A.8 for ethanol precipitation.

D. Reverse transcription.

1. Resuspend RNA in 11 μ l of DEPC-water, following the procedure described in step A.1.

2. Add:

5 $\times$ Superscript buffer	4.0 μ l
cDNA primer (50 pmol)	1.0 μ l

3. Denature mixture at 68°C for 10 min, then immediately cool in ice for 5 min.

4. Add:

10 mM dNTP mix	1.0 μ l
0.1 M DTT	2.0 μ l
Superscript enzyme (200 U)	1.0 μ l

Gently pipet mixture up and down several times to mix.

5. Incubate at 42°C for 1 h.

6. Add:

ACS-water	280 μ l
10M ammonium acetate	75 μ l
Phenol:chloroform, 1:1	375 μ l

Vortex at high speed for 15–30 s. Spin at 12,000g for 5 min.

7. Recover aqueous phase and repeat extraction with an equal volume of chloroform.

8. Follow steps A.6 through A.8 for ethanol precipitation.

E. RNA removal.

1. Resuspend RNA/cDNA pellet in 25 μ l of ACS-water, following the procedure described in step A.1.

2. Add:

0.5 M EDTA, pH 8.0	1.0 μ l
0.15 N sodium hydroxide	12.5 μ l

3. Incubate at 68°C for 15 min.

4. Add:

1 M Tris, pH 8.0	12.5 μ l
1 N hydrochloric acid	12.5 μ l
10 M ammonium acetate	13.0 μ l
Absolute ethanol	190.0 μ l

5. Mix by inverting several times, then spin at 12,000g for 20 min.

6. Follow steps A.7 and A.8 for 70% ethanol wash.

7. Resuspend cDNA pellet in 100 μ l of ACS-water, following the procedure described in step A.1.

F. Primary PCR.

1. Aliquot 5 μ l of cDNA into a 500- μ l-capacity thin-walled PCR tube. Place tube on ice.

2. Add:

10 $\times$ pfu buffer	5.0 μ l
2 mM dNTP mix	5.0 μ l
RLM-RACE Anchor 1 (40 pmol)	4.0 μ l
pfu enzyme (1.25 U)	0.5 μ l
ACS-water	29.5 μ l

Gently pipet mixture up and down several times to mix.

3. Cover mixture with light mineral oil and place tube in thermal cycler.

4. PCR

1 cycle:	98°C, 1–5 min
	43°C, 1 min
	72°C, 2.5 min
4 cycles:	97°C, 1 min
	43°C, 1 min
	72°C, 2 min
Soak:	85°C

5. As soon as the soak file begins, add 1 μ l (40 pmol)

of 3'-PCR primer 1. Make sure to add *under* the mineral oil. Gently mix by pipetting up and down a few times, taking care not to make air bubbles in the sample.

6. PCR

30 cycles:	97°C, 1 min
	55°C, 1 min
	72°C, 2 min

Soak: 4°C

7. Remove 15 μ l of primary PCR product and transfer to a fresh 1.5-ml microfuge tube, taking care not to transfer any mineral oil.

8. Add 15 μ l of ACS-water. Purify PCR product using the Wizard PCR Preps system (Promega), following the instructions of the manufacturer. Use 50 μ l of ACS-water for the final elution.

G. Secondary PCR.

1. Aliquot 5 μ l of purified primary PCR product into a 500- μ l-capacity thin-walled PCR tube. Place tube on ice.

2. Add:

10 $\times$ pfu buffer	5.0 μ l
2 mM dNTP mix	5.0 μ l
3'-PCR primer 2 (40 pmol)	4.0 μ l
pfu enzyme (1.25 U)	0.5 μ l
ACS-water	29.5 μ l

Gently pipet mixture up and down several times to mix.

3. Cover mixture with light mineral oil and place tube in thermal cycler.

4. PCR

1 cycle:	98°C, 1–5 min
	51°C, 1 min
	72°C, 2.5 min
9 cycles:	97°C, 1 min
	51°C, 1 min
	72°C, 2 min
Soak:	85°C

5. As soon as the soak file begins, add 1 μ l (40 pmol) of RLM-RACE Anchor 2. Make sure to add *under* the mineral oil. Gently mix by pipetting up and down a few times, taking care not to make air bubbles in the sample.

6. PCR

4 cycles:	97°C, 1 min
	45°C, 1 min
	72°C, 2 min
26 cycles:	97°C, 1 min
	51°C, 1 min
	72°C, 2 min
Soak:	4°C

7. Purify 15 μ l of secondary PCR product as described in step F.8.

H. Analysis and cloning of RLM-RACE products.

1. Digest 10 μ l of purified secondary PCR product with restriction enzymes which will create "sticky" 5'- and 3'-ends. (Alternatively, one of the restriction endonuclease-independent cloning methods can be used, as discussed under Cloning in the text.)

2. Run the following on a 1.5% agarose gel with the appropriate molecular weight markers: 2.5 μ l of *unpurified* primary PCR reaction, 2.5 μ l of *unpurified* secondary PCR reaction, 5.0 μ l of *purified* secondary PCR reaction, and 25% of *purified, digested* secondary PCR product.

3. Photograph (ethidium bromide-stained) gel with ruler next to molecular weight markers.

4. Southern blot gel and probe blot with the specific 3'-end probe. Visualize positive PCR products by autoradiography.

5. Compare the autoradiogram with the photograph of the ethidium-stained gel to ascertain the following:

a. Are all ethidium-staining bands in secondary PCR specific?

b. Were any specific products of the primary PCR lost in the secondary PCR?

c. Are all specific products of the secondary PCR visible as ethidium-staining bands?

d. Were secondary PCR products recovered with high yield in the purification step?

e. Does digestion of secondary PCR products reveal cutting at internal sites by the selected restriction enzymes?

6. Provided that the RACE product has been successfully amplified, purified, and digested according to the criteria listed above, the remainder of the digested product can be ligated into an appropriately digested sequencing vector and transformed into *E. coli*.

7. Grow overnight cultures of individual colonies, purify plasmid DNA, and digest with appropriate restriction enzymes to determine size of inserts. Sequence plasmids containing desired inserts.

Notes

General. DEPC-water and Tris-equilibrated phenol should be prepared as described in Sambrook *et al.* (25). ACS-water can be purchased from Aldrich. Solutions used with RNA should be prepared with DEPC-water and autoclaved. Solutions used with cDNA and for PCR should be prepared with ACS-water and autoclaved. Divide solutions into small aliquots to guard against contamination. The use of tips with aerosol-blocking plugs greatly reduces the chance of DNA contamination from pipetting devices.

A. CIP should be used instead of BAP, because BAP activity is very difficult to remove (25). CIP is obtained from New England Biolabs at a concentration of (10 U/ μ l) with a 10 $\times$ buffer. A small quantity of degraded,

phosphorylated RNA can survive the 1 h CIP treatment, so longer incubations may yield improved results in certain cases. CIP should not be heat-inactivated, since this will promote RNA hydrolysis, leading to the production of more degraded RNA molecules with 5'-terminal phosphates. RNasin is obtained from Promega at a concentration of (40 U/ μ l).

B. TAP is obtained from Epicentre at a concentration of (5 U/ μ l) with a 10 $\times$ buffer. The 3'-hydroxyl group can be removed (28) or blocked (13,38) to prevent RNA self-ligation, but this step is not necessary.

C. T4 RNA ligase is obtained from New England BioLabs at a concentration of (20 U/ μ l) with a 10 $\times$ buffer. To improve the efficiency of RNA oligo ligation, the oligo should be at high molar excess relative to the decapped RNA. The RNA oligo must be PAGE-purified in order to remove molecules shorter than full-length.

D. Superscript reverse transcriptase is obtained from Gibco-BRL at a concentration of (200 U/ μ l) with a 5 $\times$ buffer. This enzyme can effectively generate full-length cDNA molecules from GC-rich templates at 42°C.

F and G. The quantity of cDNA required to generate the desired PCR product will need to be empirically determined in each case, but should be no more than 5% of the final product. Since a small quantity of partially degraded molecules may survive CIP treatment and ultimately appear in the final cDNA population as partial clones with a terminal anchor, we recommend titrating the amount of input cDNA to the lowest possible quantity. If good quality RNA was used, the full-length molecules should be present in excess, and thus, at high dilution, there should only be full-length templates present.

Recombinant pfu enzyme is obtained from Stratagene at a concentration of (2.5 U/ μ l) with a 10 $\times$ buffer. The denaturation times and temperatures have been chosen to maximize template denaturation. Since DNA is clearly damaged at high temperatures (41,49,50), it is probably advisable to limit the time and temperature of denaturation to 15 s and 95°C, respectively, when the target of amplification is over a few hundred base pairs. Primers are specified in Table 1, and annealing temperatures are specific for the primers listed. Extension time of 2 min/kb of target sequence works well when using pfu. The PCR profile is a modification of the two-step protocol described by Barnard *et al.* (79). Several cycles of asymmetric amplification with one primer are carried out in order to achieve amplification of a desired subset of target molecules. The template mixture is thereby enriched for a desired target before addition of the second primer. A lower number of cycles in one or both amplifications may be sufficient for many targets.

H. If certain products of the primary PCR appear to be lost due to competition with smaller or more abundant species in the secondary PCR, the primary PCR product can be size-selected by fractionation in low-melting-point agarose, purified, and amplified by the secondary PCR protocol. Other suggestions for optimizing cloning are contained within the text.

ACKNOWLEDGMENTS

The author thanks Dr. Erik Flemington for suggesting that this Review be written and for critical reading of and commentary on the manuscript. The author also thanks Drs. S. H. Speck and J. L. Strominger for guidance and financial support. This work was made possible by National Institutes of Health grants CA47554 to J.L.S. and CA43143 to S.H.S., and an Office of Naval Research Graduate Research Fellowship to B.C.S.

REFERENCES

1. Arnheim, N., and Erlich, H. (1992) *Annu. Rev. Biochem.* **61**, 131–156.
2. Innis, M. A., and Gelfand, D. H. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.) pp. 3–12, Academic Press, San Diego.
3. Frohman, M. A., Dush, M. K., and Martin, G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
4. Loh, E. Y., Elliot, J. F., Cwirla, S., Lanier, L. L., and Davis, M. M. (1989) *Science* **243**, 217–220.
5. Ohara, O., Dorit, R., and Gilbert, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5673–5677.
6. Troutt, A. B., McHeyzer-Williams, M. G., Pulendran, B., and Nossal, G. J. V. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9823–9825.
7. Troutt, A. B. (1994) in *PCR Technology: Current Innovations* (Griffin, H. G., and Griffin, A. M., Eds.) pp. 141–145, CRC Press, Boca Raton, FL.
8. Dumas Milne Edwards, J. B., Delort, J., and Mallet, J. (1991) *Nucleic Acids Res.* **19**, 5227–5232.
9. Dumas Milne Edwards, J. B., Delort, J., and Mallet, J. (1993) in *Methods in Molecular Biology* (White, B. A., Ed.) Vol. 15, pp. 365–385, Humana Press, Totowa, NJ.
10. Gorovsky, M. A., and Liu, X. (1993) *Nucleic Acids Res.* **21**, 4954–4960.
11. Fromont-Racine, M., Bertrand, E., Pictet, R., and Grange, T. (1993) *Nucleic Acids Res.* **21**, 1683–1684.
12. Apte, A., and Siebert, P. D. (1993) *BioTechniques* **15**, 890–893.
13. Maruyama, K., and Sugano, S. (1994) *Gene* **138**, 171–174.
14. Rychlik, W. (1993) in *Methods in Molecular Biology* (White, B. A., Ed.) Vol. 15, pp. 31–40, Humana Press, Totowa, NJ.
15. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
16. Favaloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749.
17. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
18. Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M., and Gerard, G. F. (1988) *Nucleic Acids Res.* **16**, 265–277.
19. Meyers, T. W., and Gelfand, D. H. (1991) *Biochemistry* **30**, 7661–7666.
20. Technical Bulletin for rTth DNA Polymerase (1991) Perkin-Elmer Cetus Inc., Norwalk, CT.
21. Bej, A. K., and Mahbubani, M. M. (1994) in *PCR Technology: Current Innovations* (Griffin, H. G., and Griffin, A. M., Eds.) pp. 219–237, CRC Press, Boca Raton, FL.
22. Jones, M. D. (1993) *Methods Enzymol.* **218**, 413–419.
23. Frohman, M. A. (1993) *Methods Enzymol.* **218**, 340–356.
24. Frohman, M. A. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.) pp. 28–38, Academic Press, San Diego.
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26. Templeton, N. S., Urselay, E., and Safer, B. (1993) *BioTechniques* **15**, 48–51.
27. Treco, D. A. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., Eds.) pp. 2.1.1–2.1.7, Green Publishing Associates and Wiley-Interscience, New York.
28. Volloch, V., Schweitzer, B., and Rits, S. (1994) *Nucleic Acids Res.* **22**, 2507–2511.
29. Volloch, V., Schweitzer, B., Zhang, X., and Rits, S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10671–10675.
30. Sallie, R., *PCR Methods Appl.* **3**, 54–56.
31. Brennan, C. A., Manthey, A. E., and Gumpert, R. I. (1983) *Methods Enzymol.* **100**, 38–52.
32. Tessier, D. C., Brousseau, R., and Vernet, T. (1986) *Biochemistry* **158**, 171–178.
33. Mandl, C. W., Heinz, F. X., Puchhammer-Stockl, E., and Kunz, C. (1991) *BioTechniques* **10**, 484–486.
34. Lewin, B. (1990) *Genes IV*, Cell Press, Cambridge, MA.
35. Greenburg, R., and Burn, V. E. (1988) *Nucleic Acids Res.* **16**, 3437–3454.
36. Bertrand, E., Fromont-Racine, M., Pictet, R., and Grange, T. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3496–3500.
37. Harrison, B., and Zimmerman, S. B. (1984) *Nucleic Acids Res.* **12**, 8235–8251.
38. Romaniuk, P. J., and Uhlenbeck, O. C. (1983) *Methods Enzymol.* **100**, 52–59.
39. Kwok, S. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.) pp. 142–145, Academic Press, San Diego.
40. Sarkar, G., and Sommer, S. (1993) *Methods Enzymol.* **218**, 381–389.
41. Barnes, W. M. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2216–2220.
42. Roberts, S. S. (1994) *J. NIH Res.* **6**, 46–48.
43. D'Aquila, R. T., Bechtel, L. J., Videler, J. A., Eron, J. J., Gorczyca, P., and Kaplan, J. C. (1991) *Nucleic Acids Res.* **19**, 3749.
44. Technical Bulletin for AmpliWax PCR Gem 100 (1991) Perkin-Elmer Cetus Inc., Norwalk, CT.
45. Kellogg, D. E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P. D., and Chenchik, A. (1994) *BioTechniques* **16**, 1134–1137.
46. Gelfand, D. H., and White, T. J. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.) pp. 129–141, Academic Press, San Diego.
47. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3746–3750.
48. Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5695–5699.

49. Lindhal, T., and Nyberg, B. (1972) *Biochemistry* **11**, 3610–3623.
50. Douglas, A., and Atchison, B. (1993) *PCR Methods Appl.* **3**, 133–134.
51. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1986) *Data for Biochemical Research*, 3rd ed., Oxford Univ. Press, Oxford, England.
52. Rychlik, W., Spencer, W. J., and Rhoads, R. E. (1990) *Nucleic Acids Res.* **18**, 6409–6412.
53. Syvanen, A., and Soderlund, H. (1993) *Methods Enzymol.* **218**, 474–490.
54. Technical Bulletin No. 118 (1993) Wizard PCR Preps DNA Purification System, Promega Corp., Madison, WI.
55. Clark, J. M. (1988) *Nucleic Acids. Res.* **16**, 9677–9686.
56. New England Biolabs Catalog (1993) New England Biolabs, Inc., Beverly, MA.
57. Jung, V., Pestka, S. B., and Pestka, S. (1993) *Methods Enzymol.* **218**, 357–362.
58. Mead, D. A., Pey, N. K., Herrnsstadt, C., Marcil, R. A., and Smith, L. M. (1991) *BioTechnology* **9**, 657–663.
59. Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1991) *Nucleic Acids Res.* **19**, 1154.
60. Holton, T. A., and Graham, M. W. (1991) *Nucleic Acids. Res.* **19**, 1156.
61. Liu, Z., and Schwartz, L. M. (1992) *BioTechniques* **12**, 28–30.
62. Lohff, C. J., and Cease, K. B. (1992) *Nucleic Acids Res.* **20**, 144.
63. Aslandis, C., and de Jong, P. J. (1990) *Nucleic Acids Res.* **18**, 6069–6074.
64. Shuldiner, A. R., Tanner, K., Scott, L. A., Moore, C. A., and Roth, J. (1991) *Anal. Biochem.* **194**, 9–15.
65. Shuldiner, A. R., and Tanner, K. (1993) in *Methods in Molecular Biology* (White, B. A., Ed.) Vol. 15, pp. 229–239, Humana Press Inc., Totowa, NJ.
66. Jones, D. H., Sakamoto, K., Vorce, R. L., and Howard, B. H. (1990) *Nature* **344**, 793–794.
67. Bahring, S., Sandig, V., Lieber, A., and Strauss, M. (1994) *BioTechniques* **16**, 807–808.
68. Hirzman, J., Hahnen, J., and Hobom, G. (1993) *Nucleic Acids. Res.* **21**, 3597–3598.
69. Fritz, J. D., Greaser, M. L., and Wolff, J. A. (1991) *Nucleic Acids Res.* **19**, 3747.
70. Struck, F., and Collins, J. (1994) *Nucleic Acids Res.* **22**, 1923–1924.
71. Whitcomb, J. M., Rashtchian, A., and Hughes, S. H. (1993) *Nucleic Acids Res.* **21**, 4143–4146.
72. Belyavavsky, A., Vinogradova, T., and Rajewsky, K. (1989) *Nucleic Acids Res.* **17**, 2919–2932.
73. Domec, C., Garaby, B., Fournier, M., and Bonnet, J. (1990) *Anal. Biochem.* **188**, 422–426.
74. Akawitz, A., and Manuelidis, L. (1989) *Gene* **31**, 295–306.
75. Harvey, R. J., and Darlison, M. G. (1991) *Nucleic Acids Res.* **19**, 4002.
76. Zuker, M., and Jaeger, J. A. (1991) in *GCG Package, Version 7*, Genetics Computer Group, Madison, WI; distributed by Molecular Biology Computer Research Resource, Dana-Farber Cancer Institute, Boston.
77. Ellington, A., and Green, R. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., Eds.) pp. 2.11.1–2.11.18, Green Publishing Associates and Wiley-Interscience, New York.
78. Smith, P. R., Gao, Y., Karren, L., Jones, M. D., Snudden, D., and Griffin, B. E. (1993) *J. Virol.* **67**, 3217–3225.
79. Barnard, R., Southard, J. N., and Talamantes, F. (1994) *BioTechniques* **16**, 251–252.