Preface

Since its invention and subsequent development nearly 20 years ago, polymerase chain reaction (PCR) has been extensively utilized to identify numerous gene probes in vitro and in vivo. However, attempts to generate complete and full-length complementary cDNA libraries were, for the most part, fruitless and remained elusive until the last decade, when simple and rapid methods were developed. With current decoding and potential application of human genome information to genechips, there are urgent needs for identification of functional significance of these decoded gene sequences. Inherent in bringing these applications to fruition is the need to generate a complete and full-length cDNA library for potential functional assays of specific gene sequences.

Generation of cDNA Libraries: Methods and Protocols serves as a laboratory manual on the evolution of generation of cDNA libraries, covering both background information and step-by-step practical laboratory recipes for which protocols, reagents, operational tips, instrumentation, and other requirements are detailed. The first chapter of the book is an overview of the basics of generating cDNA libraries, which include the following: (a) the definition of a cDNA library, (b) different kinds of cDNA libraries, (c) differences between methods for cDNA library generation using conventional approaches and novel strategies, including reverse generation of RNA repertoires from cDNA libraries, and (d) the quality of cDNA libraries. In subsequent chapters, various methods are presented to provide the reader with a wide range of methodologies for enhancing the generation of complete and full-length libraries. Again, each method of cDNA library generation contains a balanced presentation of both background information and practical procedures. The remainder of this book explains how to confirm the quality of the cDNAs generated and some of the applications, including (a) electrophoresis, (b) Northern blotting, (c) microarray analysis, (d) subtractive hybridization, (e) subtractive cloning, (f) gene cloning, and (g) peptide library generation.

The final chapter of the book outlines the future use of full-length cDNA libraries in biomedical research, diagnostic utilization, drug development, and clinical therapy.

The authors contributing the various chapters are all experts in their fields, and they have either developed and/or routinely performed the methodologies

described herein. It is anticipated that the subject matter covered in *Generation of cDNA Libraries: Methods and Protocols* will be particularly useful for biologists, biochemists, molecular biologists, and clinicians, which would furnish them several ready-to-use methodologies for attacking the problems in their specific areas of interests.

Shao-Yao Ying

2

Rapid Amplification of cDNA Ends

Yue Zhang

1. Introduction

Rapid amplification of complementary DNA (cDNA) ends (RACE) is a powerful technique for obtaining the ends of cDNAs when only partial sequences are available. In essence, an adaptor with a defined sequence is attached to one end of the cDNA; then, the region between the adaptor and the known sequences is amplified by polymerase chain reaction (PCR). Since the initial publication in 1988 *(1)*, RACE has greatly facilitated the cloning of new genes. Currently, RACE remains the most effective method of cloning cDNAs ends. It is especially useful in the studies of temporal and spatial regulation of transcription initiation and differential splicing of mRNA. The methods described in this chapter are quite simple and efficient. A linker at the 3′ end and an adaptor at the 5′ end are added to the first strand of cDNA during reverse transcription; amplification of virtually any transcript to either end can then make use of this same pool of cDNAs. In addition to being simple, the efficiency of 5′-RACE is dramatically increased because the adaptor is added only to full-length cDNAs.

Since the initial description of RACE, many labs have developed significant improvements on the basic approach. The methods described here were developed from more recent reports from Frohman's and Roeder's groups. Among these, adaptor addition accompanying reverse transcription was developed from the CapFinding *(2–4)* technique of Clonetech (Palo Alto, CA): Moloney murine leukemia virus reverse transcriptase (MMLV RT) adds an extra two to four cytosines to the 3′ ends of newly synthesized cDNA strands upon reaching the cap structure at the 5′ end of mRNA templates. When an oligonucleotide with multiple G's at its 3′-most end is present in the reaction mixture, its terminal G nucleotides base pair with the C's of the

newly synthesized cDNA. Through a so-called "template switch" process, this oligonucleotide will serve as a continuing template for the RT. Thus, the reverse complement adaptor sequence can be easily incorporated into the 3′ end of the newly synthesized first strand of cDNA, which is at the beginning of the new cDNA (*see* **Fig. 1**). CapFinding obligates the addition of the adaptor in a Cap-dependent manner, resulting in adaptor attachment to full-length cDNA clones only. Therefore, because there is no additional enzymatic modification of the cDNAs after reverse transcription, this results in a simplified method with improved overall efficiency.

This protocol also utilizes biotin–streptavidin interactions to facilitate the elimination of excessive adaptors before carrying out PCR (*see* **Fig. 1**). The importance of adaptor elimination has been documented since the initial description of RACE *(1)*. The presence of extra adaptors is detrimental to the following PCR reaction because their sequence or complement sequence is present in ALL cDNAs in the reaction mixture, resulting in heavy background amplification and failure to amplify the specific product if not removed.

2. Materials

1. 5X Reverse transcription buffer: 250 mM Tris-HCl pH 8.3 (at 45°C), 30 mM MgCl$_2$, 10 mM MnCl$_2$, 50 mM dithiothreitol, 1 mg/mL bovine serum albumin (BSA).
2. Biotin-labeled primer P$_{total}$ (biotin-labeled primers can be ordered from Invitrogen). The sequences of P$_{total}$ and the following primers are listed in **Fig. 2**.
3. CapFinder adaptor.
4. RNasin (Promega Biotech).
5. dNTPs: 10 mM solutions (PL-Biochemicals/Pharmacia or Roche).
6. SuperScript II RNase H$^-$ Reverse Transcriptase (Invitrogen/Life Technologies).
7. Streptavidin MagneSphere Particles and MagneSphere Magnetic Separation Stand (components of PolyATract mRNA Isolation System from Promega).
8. TE: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA.
9. PCR cocktail: hot start polymerase systems are recommended (e.g., Stratagene Hercules, or Roche Expand™ High Fidelity). Assemble the PCR cocktail according to the manufacturer's instructions. One should also use the extension temperature specified. For simplicity, 72°C is used in the following methods.
10. Gene specific primer 1 (GSP1), GSP2, and P$_o$ and P$_i$ primers for 3′-RACE or reverse gene specific primer 1 (RGSP1), RGSP2, and U$_o$ and U$_i$ primers for 5′-RACE.

3. Methods

3.1. Reverse Transcription to Generate cDNA Templates (see *Notes 1–7*)

1. Assemble reverse transcription components on ice: 4 μL of 5X reverse transcription buffer, 2 μL of dNTPs, 1 μL of CapFinder adaptor (10 μM), and 0.25 μL (10 U) of RNasin.
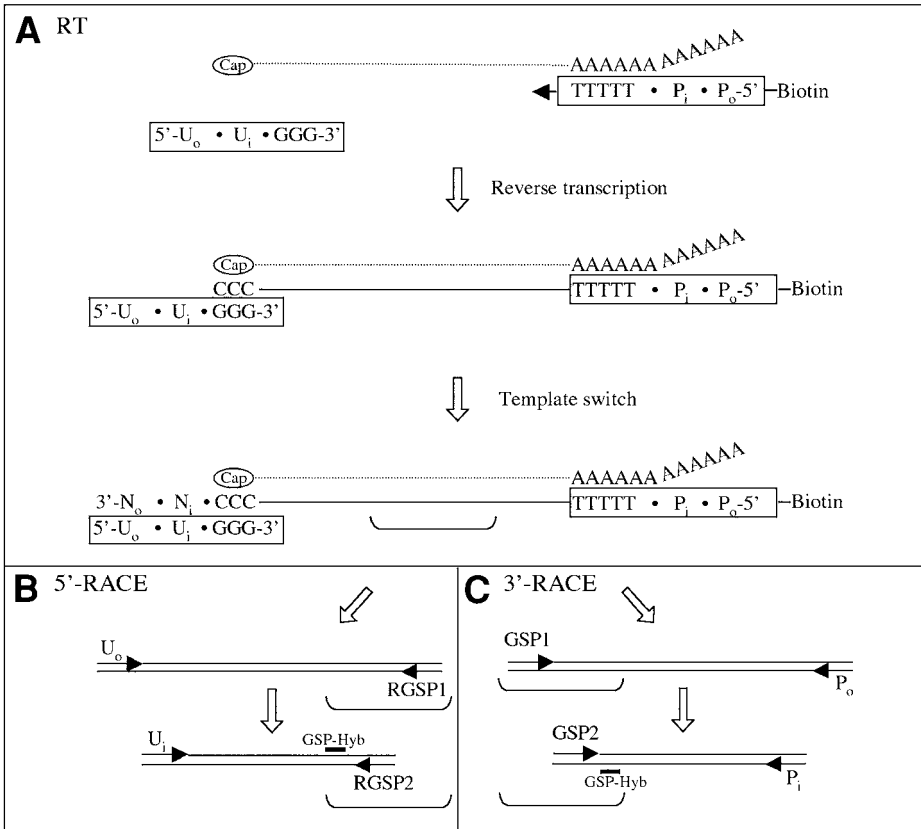
A RT

$5'-U_o \cdot U_i \cdot GGG-3'$

Cap .... AAAAAA AAAAAA
TTTTT $\cdot$ P_i $\cdot$ P_o-5' —Biotin

⇩ Reverse transcription

Cap .... AAAAAA AAAAAA
CCC ———————————
$5'-U_o \cdot U_i \cdot GGG-3'$
TTTTT $\cdot$ P_i $\cdot$ P_o-5' —Biotin

⇩ Template switch

$3'-N_o \cdot N_i \cdot CCC$ ——— Cap .... AAAAAA AAAAAA
$5'-U_o \cdot U_i \cdot GGG-3'$
TTTTT $\cdot$ P_i $\cdot$ P_o-5' —Biotin

B 5'-RACE

U_o ▶
RGSP1 ◀

⇩

U_i ▶ GSP-Hyb
RGSP2 ◀

C 3'-RACE

GSP1 ▶
◀ P_o

⇩

GSP2 ▶
GSP-Hyb ◀ P_i

Fig. 1. Schematic representation of RACE. (**A**) Reverse transcription, template switch, and incorporation of adaptor sequences at the 3′ end of the first strand of cDNA. Biotin-labeled primer P_{total} is used to initiate reverse transcription through hybridization of the poly(dT) tract with the mRNA polyA tail. After reaching the 5′ end of the mRNA, oligo(dC) is added by the reverse transcriptase in a Cap-dependent manner. Then, through template switching via base pairing between the oligo(dC) and the oligo(dG) at the end of CapFinder Adaptor, the reverse complementary sequence of the CapFinder oligo is incorporated to the first strand of the cDNA. The dotted line indicates mRNA, the solid line indicates cDNA, and the rectangle indicates the primer. The brace indicates the known region. (**B**) 5'-RACE. The first round of PCR uses primer U_o and RGSP1 (reverse gene-specific primer 1); the second round uses U_i and RGSP2. GSP-Hyb is also within the known region; it can be used to confirm the authenticity of the RACE product. (**C**) 3′-RACE. Similar to 5′-RACE, but note that GSP1 and GSP2 are in the same sequences as the gene, whereas the P_o and P_i are the reverse complement.

```
CapFinder: 5'-TGGTTGCCATAAGCGGATCATCGGGAGGAGAAACGGG-3'
U_o:          5'-TGGTTGCCATAAGCGGATC-3'
U_i:                              5'-TCATCGGGAGGAGAAACGG-3':


P_total: 3'-G/A/C(T)_17CTATCGCTCCGCTCGCAAGGTTTGGGTCAGGTTGGTTT-5'
P_i:             3'-CTATCGCTCCGCTCGCAAG-5'
P_o:                              3'-GGTTTGGGTCAGGTTGGTTT-5
```

```
For convenience: The 5'->3' sequences of the P primers:
```

```
P_total: 5'-TTTGGTTGGACTGGGTTTGGAACGCTCGCCTCGCTATC(T)_17C/A/G-3'
P_o:     5'-TTTGGTTGGACTGGGTTTGG-3'
P_i:                          5'-GGAACGCTCGCCTCGCTATC -3'
```

Fig. 2. Primer sequences and their relationship. GSPs or RGSPs are not included. The CapFinder sequence was selected from the *Yersinia pestis* Genome sequence. BLASTN search results using it can be retrieved with request ID (RID) 1009985097-16592-24994. P_{total} is biotin labeled at the 5′ end. This is a "lock-docking" degenerate primer that actually consists of three primers with different (A, G, or C) nucleotides at its 3′ end. Its RID is 1009987520-16411-27741.

2. Heat 1 μg of polyA$^+$ RNA and 10 pmol P_{total} primer in 11.75 μL of water at 80°C for 3 min, cool rapidly on ice, and spin for 5 s in a microcentrifuge. Combine with the components from **step 1**.
3. Add 1 μL (200 U) of SuperScript II reverse transcriptase to the above mixture and incubate for 5 min at room temperature, 30 min at 42°C, 30 min at 45°C, and 10 min at 50°C.
4. Incubate at 70°C for 15 min to inactivate the reverse transcriptase. Add in magnetic streptavidin beads; use about five times the binding capacity required to complex the amount of biotinylated primer used. Wash with TE at 50°C three times to eliminate the free CapFinder adaptors.
5. Dilute the reaction mixture to 0.5 mL with TE and store at 4°C (cDNA pool).

3.2. Amplification of the cDNA (see Notes 8–13)

3.2.1. First Round

1. Add an aliquot (1 μL) of the cDNA pool (resuspend well) and primers (25 pmol each of GSP1 and P_o for 3′-RACE, or RGSP1 and U_o for 5′-RACE) to 50 μL of PCR cocktail in a 0.5-mL PCR tube.
2. Heat the mixture in the thermal cycler at 95°C for 5 min to denature the first-strand products and the streptavidin; add 2.5 U *Taq* polymerase and mix well (hot start). Incubate at appropriate annealing temperature for 2 min. Extend the cDNAs at 72°C for 40 min. It is not necessary to keep the magnetic streptavidin
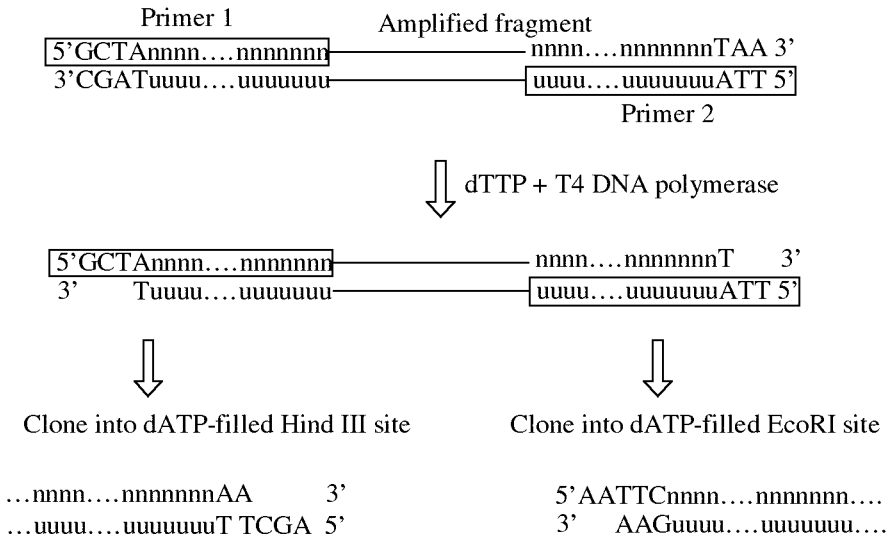
Primer 1   Amplified fragment

5'GCTAnnnn….nnnnnnn━━━━━━━━━ nnnn….nnnnnnnTAA 3'
3'CGATuuuu….uuuuuuu━━━━━━━━━ uuuu….uuuuuuuATT 5'
                                              Primer 2

⇓ dTTP + T4 DNA polymerase

5'GCTAnnnn….nnnnnnn━━━━━━━━━ nnnn….nnnnnnnT   3'
3'   Tuuuu….uuuuuuu━━━━━━━━━ uuuu….uuuuuuuATT 5'

⇓   ⇓

Clone into dATP-filled Hind III site   Clone into dATP-filled EcoRI site

…nnnn….nnnnnnnAA   3'   5'AATTCnnnn….nnnnnnn….
…uuuu….uuuuuuuT TCGA 5'   3'   AAGuuuu….uuuuuuu….

Fig. 3. A safe and easy cloning method. Details are described in **Subheading 3.3.**

beads resuspended during these incubations because the biotin should not interact with the denatured streptavidin. It has been reported that styrene beads are smaller, stay without agitation in solution, and are potentially better *(3)*. The respective performances have not been compared in the author's hands.

3. Carry out 30 cycles of amplification using a step program (94°C, 1 min; 52–68°C, 1 min; 72°C, 3 min), followed by a 15-min final extension at 72°C. Cool to room temperature. The extension time at 72°C needs to be adjusted according to the length of the product expected and the speed of the polymerase used.

3.2.2. Second Round (If Necessary)

1. Dilute 1 µL of the amplification products from the first round into 20 µL of TE.
2. Amplify 1 µL of the diluted material with primers GSP2 and P_i for 3′-RACE, or RGSP2 and U_i for 5′-RACE, using the first-round procedure, but eliminate the initial 2-min annealing step and the 72°C, 40-min extension step.

3.3. Safe and Easy Cloning Protocol (see Note 14)

1. Insert preparation: select a pair of restriction enzymes for which you can synthesize half-sites appended to PCR primers that can be chewed back to form appropriate overhangs, as shown for *Hin*dIII and *Eco*RI in **Fig. 3**. For example, add "TTA" to the 5′ end of P_i and add "GCTA" to the 5′ end of GSP2. Carry out PCR as usual.
2. After PCR, clean up using Qiagen PCR cleanup spin columns.

3. On ice, add the selected dNTP(s) (e.g., dTTP) to a final concentration of 0.2 m*M*, 1/10 vol of 10X T4 DNA polymerase buffer, and 1–2 U T4 DNA polymerase.

4. Incubate at 12°C for 15 min and then 75°C for 10 min to heat inactivate the T4 DNA polymerase. (Optional: gel-isolate DNA fragment of interest, depending on degree of success of PCR amplification.)

5. Vector preparation: digest vector (e.g., pGem-7ZF (Promega)) using the selected enzymes (e.g., *Hin*dIII and *Eco*RI) under optimal conditions, in a volume of 10 μL.

6. Add a 10-μL mixture containing the selected dNTP(s) (e.g., dATP) at a final concentration of 0.4 m*M*, 1 μL of the restriction buffer used for digestion, 0.5 μL Klenow, and 0.25 μL Sequenase.

7. Incubate at 37°C for 15 min and then 75°C for 10 min to heat inactivate the polymerases.

8. Gel-isolate the linearized vector fragment.

9. For ligation, use equal molar amounts of vector and insert.

Other manipulations of RACE PCR products are discussed in **Note 15**.

4. Notes

1. PolyA RNA is preferentially used for reverse transcription to decrease background, although total RNA can be used as well. An important factor in the generation of full-length cDNAs concerns the stringency of the reverse transcription reaction. Reverse transcription reactions were historically carried out at relatively low temperatures (37–42°C) using a vast excess of primer (approximately one-half the mass of the mRNA). Under these low-stringency conditions, a stretch of A residues as short as six to eight nucleotides will suffice as a binding site for an oligo(dT)-tailed primer. This may result in cDNA synthesis being initiated at sites upstream of the polyA tail, leading to truncation of the desired amplification product. One should be suspicious that this has occurred if a canonical polyadenylation signal sequence is not found near the 3′ end of the cDNAs generated. This low-stringency problem can be minimized by controlling two parameters: primer concentration and reaction temperature. The primer concentration can be reduced dramatically without decreasing the amount of cDNA synthesized significantly and will begin to bind preferentially to the longest A-rich stretches present (i.e., the polyA tail). The recommended quantity in **Subheading 3.1.** represents a good starting point. It can be reduced fivefold further if significant truncation is observed.

2. The efficiency of cDNA extension is important, especially for 5′-RACE. In the described protocol, the incubation temperature is raised slowly to encourage reverse transcription to proceed through regions of difficult secondary structure. Synthesis of cDNAs at elevated temperatures should diminish the amount of secondary structure encountered in GC-rich regions of the mRNA. Because the half-life of reverse transcriptase rapidly decreases as the incubation temperature increases, the reaction cannot be carried out at elevated temperatures in its

entirety. Alternatively, the problem of difficult secondary structure (and non-specific reverse transcription) can be approached using heat-stable reverse transcriptases, which are now available from several suppliers (Perkin-Elmer-Cetus, Amersham, Epicentre Technologies, and others). Like PCR reactions, the stringency of reverse transcription can be controlled by adjusting the temperature at which the primer is annealed to the mRNA. Optimal temperature depends on the specific reaction buffer and reverse transcriptase used and should be determined empirically, but will usually be in the range 48–56°C for a primer terminated by a 17-nt oligo(dT).

3. In addition to synthesis of cDNAs at elevated temperature, there are several other approaches that encourage cDNA extension. First, use clean, intact RNA. Second, select a gene-specific primer for reverse transcription (GSP-RT) that is close to the 5′ end within the known sequences, thus minimizing difficult regions. A random hexamer (50 ng) can also be used to create a universal 5′-end cDNA pool. If using random hexamers, then a room-temperature 10-min incubation period is needed after mixing everything together.

4. The successfulness of 5′-RACE relies on the incorporation of the CapFinder Adaptor sequence at the beginning of the cDNA. As mentioned, this step depends on the addition of extra oligo(dC) at the end of the first strand of cDNA. It has been shown that the Mn^{2+} ion in the reverse transcription buffer greatly increases the percentage of oligo(dC) added to the end of the first strands of cDNAs *(5)*.

5. The presence of excess P_{total} and CapFinder adaptors during amplification will be detrimental to the reaction. Virtually all of the cDNA produced will contain the primer P_{total} at the 3′ end and (hopefully) the CapFinder sequence at the 5′ end. The physical presence of both primers will cause heavy background and failure of RACE. This phenomenon has been described even in the original RACE article *(1)*. Here, a semisolid-phase cDNA synthesizing protocol is adapted to deal with this problem. The primer P_{total} is biotin labeled, and after the reverse transcription reaction, streptavidin beads are used to separate to the cDNAs from unincorporated CapFinder adaptors.

6. The following discusses some issues regarding potential problems with the reverse transcription steps.

    a. Damaged RNA: Electrophorese RNA in a 1% formaldehyde minigel and examine the integrity of the 18*S* and 28*S* ribosomal bands. Discard the RNA preparation if ribosomal bands are not sharp.

    b. Contaminants: Ensure that the RNA preparation is free of agents that inhibit reverse transcription (e.g., lithium chloride and sodium dodecyl sulfate) *(6)*.

    c. Bad reagents: To monitor reverse transcription of the RNA, add 20 µCi of 32p-dCTP to the reaction, separate the newly created cDNAs using gel electrophoresis, wrap the gel in Saran Wrap™, and expose it to X-ray film. Accurate estimates of cDNA size can best be determined using alkaline agarose gels, but a simple 1% agarose minigel will suffice to confirm that reverse transcription took place and that cDNAs of reasonable length were

generated. Note that adding 32p-dCTP to the reverse transcription reaction results in the detection of cDNAs synthesized both through the specific priming of mRNA and through RNA self-priming. When a gene-specific primer is used to prime transcription (5′-end RACE) or when total RNA is used as a template, the majority of the labeled cDNA will actually have been generated from RNA self-priming. To monitor extension of the primer used for reverse transcription, label the primer using T4 DNA kinase and 32p-γATP prior to reverse transcription. A much longer exposure time will be required to detect the labeled primer-extension products than when 32p-dCTP is added to the reaction.

7. To monitor reverse transcription of the gene of interest, one may attempt to amplify an internal fragment of the gene containing a region derived from two or more exons, if sufficient sequence information is available.

8. For 3′-end amplification, it is important to add the *Taq* polymerase after heating the mixture to a temperature above the T_m of the primers ("hot-start" PCR). Addition of the enzyme prior to this point allows one "cycle" to take place at room temperature, promoting the synthesis of nonspecific background products dependent on low-stringency interactions.

9. An annealing temperature close to the effective T_m of the primers should be used. Computer programs to assist in the selection of primers are widely available and should be used. An extension time of 1-min/kb expected product should be allowed during the amplification cycles. If the expected length of product is unknown, try 3–4 min initially.

10. Very little substrate is required for the PCR reaction. One microgram of polyA$^+$ RNA typically contains approx 5×10^7 copies of each low-abundance transcript. The PCR reaction described here works optimally when 10^3–10^5 templates (of the desired cDNA) are present in the starting mixture; therefore, as little as 0.002% of the reverse transcription mixture suffices for the PCR reaction! The addition of too much starting material to the amplification reaction will lead to production of large amounts of nonspecific product and should be avoided. The RACE technique is particularly sensitive to this problem, as every cDNA in the mixture, desired and undesired, contains a binding site for the P_i and P_o primers.

11. It was found empirically that allowing extra extension time (40 min) during the first amplification round (when the second strand of cDNA is created) sometimes resulted in increased yields of the specific product relative to background amplification and, in particular, increased the yields of long cDNAs versus short cDNAs when specific cDNA ends of multiple lengths were present *(1)*. Prior treatment of cDNA templates with RNA hydrolysis or a combination of RNase H and RNase A infrequently improves the efficiency of amplification of specific cDNAs.

12. Some potential amplification problems are as follows:

    a. No product: If no products are observed for the first set of amplifications after 30 cycles, add fresh *Taq* polymerase and carry out an additional 15 rounds of

amplification (extra enzyme is not necessary if the entire set of 45 cycles is carried out without interruption at cycle 30). Product is always observed after a total of 45 cycles if efficient amplification is taking place. If no product is observed, carry out a PCR reaction using control templates and primers to ensure the integrity of the reagents.

b. Smeared product from the bottom of the gel to the loading well: There are too many cycles or too much starting material.

c. Nonspecific amplification, but no specific amplification: Check sequence of cDNA and primers. If all are correct, examine primers (using computer program) for secondary structure and self-annealing problems. Consider ordering new primers. Determine whether too much template is being added or if the choice of annealing temperatures could be improved. Alternatively, secondary structure in the template may block amplification. Consider adding formamide *(7)* or [7]aza-GTP (in a 1:3 ratio with dGTP) to the reaction to assist polymerization. [7]aza-GTP can also be added to the reverse transcription reaction.

d. Inappropriate templates: To determine whether the amplification products observed are being generated from cDNA or whether they derive from residual genomic DNA or contaminating plasmids, pretreat an aliquot of the RNA with RNase A.

13. The following describes the analysis of the quality of the RACE PCR products:

a. The production of specific partial cDNAs by the RACE protocol is assessed using Southern blot hybridization analysis. After the second set of amplification cycles, the first- and second-set reaction products are electrophoresed in a 1% agarose gel, stained with ethidium bromide, denatured, and transferred to a nylon membrane. After hybridization with a labeled oligomer or gene fragment derived from a region contained within the amplified fragment (e.g., GSP-Hyb in **Fig. 1B** or **1C**), gene-specific partial cDNA ends should be detected easily. Yields of the desired product relative to nonspecific amplified cDNA in the first-round products should vary from <1% of the amplified material to nearly 100%, depending largely on the stringency of the amplification reaction, the amplification efficiency of the specific cDNA end, and the relative abundance of the specific transcript within the mRNA source. In the second set of amplification cycles, approx 100% of the cDNA detected by ethidium bromide staining should represent specific product. If specific hybridization is not observed, then troubleshooting steps should be initiated.

b. Information gained from this analysis should be used to optimize the RT procedure. If low yields of specific product are observed because nonspecific products are being amplified efficiently, then annealing temperatures can be raised gradually (approx 2°C at a time) and sequentially in each stage of the procedure until nonspecific products are no longer observed. Alternatively, some investigators have reported success using the "touchdown PCR" procedure to optimize the annealing temperature without trial and error *(8)*. Optimizing the annealing temperature is also recommended if multiple species of specific products are observed, which could indicate that truncation of

specific products is occurring. If multiple species of specific products are observed after the reverse transcription and amplification reactions have been fully optimized, then the possibility should be entertained that alternate splicing or promoter use is occurring.

c. Look for TATA, CCAAT, and initiator element (Inr) sites at or around the candidate transcription site in the genomic DNA sequence if it is available. One should usually be able to find either TATA or an Inr.

14. Cloning of RACE products like any other PCR products.

a. Option 1: To clone the cDNA ends directly from the amplification reaction (or after gel purification, which is recommended), ligate an aliquot of the products to plasmid vector encoding a one-nucleotide 3′ overhang consisting of a "T" on both strands. Such vector DNA is available commercially (Invitrogen's "TA Kit") or can be easily and inexpensively prepared (e.g., **ref. *9***).

b. Option 2: A safer and very effective approach is to modify the ends of the primers to allow the creation of overhanging ends using T4 DNA polymerase to chew back a few nucleotides from the amplified product in a controlled manner and Klenow enzyme (or Sequenase) to fill in partially restriction-enzyme-digested overhanging ends on the vector, as shown in **Fig. 3** and discussed in **Subheading 3.4.** (adapted from **refs. *10*** and ***11***).

This approach has many advantages. It eliminates the possibility that the restriction enzymes chosen for the cloning step will cleave the cDNA end in the unknown region. In addition, vector dephosphorylation is not required because vector self-ligation is no longer possible, insert kinasing (and polishing) is not necessary, and insert multimerization and fusion clones are not observed. Overall, the procedure is more reliable than "TA" cloning.

15. Other manipulation of RACE PCR products:

a. Sequencing: RACE products can be sequenced directly on a population level using a variety of protocols, including cycle sequencing, from the end at which the gene-specific primers are located. Note that 3′-RACE products cannot be sequenced on a population level using the P_i primer at the unknown end, because individual cDNAs contain different numbers of A residues in their polyA tails and, as a consequence, the sequencing ladder falls out of register after reading through the tail. Using the set of primers TTTTTTTTTTTTTTTTTTA/G/C, 3′-end products can be sequenced from their unknown end. The non-T nucleotide at the 3′ end of the primer forces the appropriate primer to bind to the inner end of the polyA tail ***(12)***. The other two primers do not participate in the sequencing reaction. Individual cDNA ends, once cloned into a plasmid vector, can be sequenced from either end using gene-specific or vector primers.

b. Hybridization probes: RACE products are generally pure enough that they can be used as probes for RNA and DNA blot analyses. It should be kept in mind that small amounts of contaminating nonspecific cDNAs will always

be present. It is also possible to include a T7 RNA polymerase promoter in one or both primer sequences and to use the RACE products with in vitro transcription reactions to produce RNA probes. Primers encoding the T7 RNA polymerase promoter sequence do not appear to function as amplification primers as efficiently as the others listed in **Fig. 2** (personal observation). Therefore, the T7 RNA polymerase promoter sequence should not be incorporated into RACE primers as a general rule.

c. Construction of full-length cDNAs: It is possible to use the RACE protocol to create overlapping 5′ and 3′ cDNA ends that can later, through judicious choice of restriction enzyme sites, be joined together through subcloning to form a full-length cDNA. It is also possible to use the sequence information gained from acquisition of the 5′ and 3′ cDNA ends to make new primers representing the extreme 5′ and 3′ ends of the cDNA and to employ them to amplify a *de novo* copy of a full-length cDNA directly from the cDNA pool. Despite the added expense of making two more primers, there are several reasons why the second approach is preferred. First, a relatively high error rate can be associated with the PCR conditions for which efficient RACE amplification takes place (depending on the conditions used) and numerous clones may have to be sequenced to identify one without mutations. In contrast, two specific primers from the extreme ends of the cDNA can be used under less efficient but low-error-rate conditions *(13)* for a minimum of cycles to amplify a new cDNA that is likely to be free of mutations. Second, convenient restriction sites are often not available, making the subcloning project difficult. Third, by using the second approach, the synthetic polyA tail (if present) can be removed from the 5′ end of the cDNA. Homopolymer tails appended to the 5′ ends of cDNAs have, in some cases, been reported to inhibit translation. Finally, if alternate promoters, splicing, and polyadenylation signal sequences are being used and result in multiple 5′ and 3′ ends, it is possible that one might join two cDNA halves that are never actually found together in vivo. Employing primers from the extreme ends of the cDNA as described confirms that the resulting amplified cDNA represents a mRNA actually present in the starting population.

Acknowledgments

References

1. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85(23),** 8998–9002.
2. Zhang, Y. and Frohman, M. A. (1997) Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs, in *Methods in Molecular Biology* (Cowell, I. G. and Austin, C. A., eds.), Vol. 69, pp. 61–87. Humana, Totowa, NJ.
3. Schramm, G., Bruchhaus, I., and Roeder, T. (2000) A simple and reliable 5′-RACE approach. *Nucleic Acids Res.* **28(22),** E96.
4. Clontech Laboratories. (1996) CapFinder™ PCR cDNA Library Construction Kit. *Clontechniques* **11,** 1.
5. Schmidt, W. M. and Mueller, M. W. (1999) CapSelect: a highly sensitive method for 5′ CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res.* **27(21),** e31.
6. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Sarkar, G., Kapelner, S., and Sommer, S. S. (1990) Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* **18(24),** 7465.
8. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991). "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19(14),** 4008.
9. Mead, D. A., Pey, N. K., Herrnstadt, C., Marcil, R. A., and Smith, L. M. (1991) A universal method for the direct cloning of PCR amplified nucleic acid. *Bio/ Technology* **9(7),** 657–663.
10. Stoker, A. W. (1990) Cloning of PCR products after defined cohesive termini are created with T4 DNA polymerase. *Nucleic Acids Res.* **18(14),** 4290.
11. Iwahana, H., Mizusawa, N., Ii, S., Yoshimoto, K., and Itakura, M. (1994) An end-trimming method to amplify adjacent cDNA fragments by PCR. *Biotechniques* **16(1),** 94–98.
12. Thweatt, R., Goldstein, S., and Shmookler Reis, R. J. (1990) A universal primer mixture for sequence determination at the 3′ ends of cDNAs. *Anal. Biochem.* **190(2),** 314–316.
13. Eckert, K. A. and Kunkel, T. A. (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res.* **18(13),** 3739–3744.