BioTechniques

ISSN: 0736-6205 (Print) 1940-9818 (Online) Journal homepage: www.tandfonline.com/journals/ibtn20

Full-Length cDNA Cloning and Determination of mRNA 5′and 3′Ends by Amplification of Adaptor-Ligated cDNA

A. Chenchik, L. Diachenko, F. Moqadam, V. Tarabykin, S. Lukyanov & P.D. Siebert

To cite this article: A. Chenchik, L. Diachenko, F. Moqadam, V. Tarabykin, S. Lukyanov & P.D. Siebert (1996) Full-Length cDNA Cloning and Determination of mRNA 5'and 3'Ends by Amplification of Adaptor-Ligated cDNA, BioTechniques, 21:3, 526-534, DOI: 10.2144/96213pf02

To link to this article: https://doi.org/10.2144/96213pf02

© 2018 Author(s)

Published online: 02 Aug 2018.

Submit your article to this journal

Article views: 1689

View related articles

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ibtn20

Full-Length cDNA Cloning and Determination of mRNA 5′ and 3′ Ends by Amplification of Adaptor-Ligated cDNA

A. Chenchik, L. Diachenko, F. Moqadam, V. Tarabykin[1], S. Lukyanov[1] and P.D. Siebert

CLONTECH Laboratories, Inc., Palo Alto, CA, USA, and [1]Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

BioTechniques 21:526-534 (September 1996)

ABSTRACT

An efficient cDNA amplification procedure is described for determining of the 5′ and 3′ ends of mRNAs and cloning full-length cDNAs. In this approach, a double-stranded (ds) adaptor is ligated to both ends of a library of ds cDNA by T4 DNA ligase. This adaptor-ligated ds cDNA is then used to selectively amplify 5′- or 3′-cDNA fragments by PCR with a combination of gene-specific and adaptor-specific primers. This is a unified method for 5′ and 3′ rapid amplification of cDNA ends (RACE) from the same adaptor-ligated ds cDNA template. A specially designed adaptor combines features of "vectorette PCR" and "suppression PCR" technologies that significantly reduce background during amplification. The application of "long and accurate PCR" (LA PCR) technology makes possible the amplification of large RACE products and full-length cDNAs with high fidelity to the original mRNA. We investigated efficacy and limitations of this PCR-based approach for cDNA cloning by amplification of 5′- and 3′-RACE fragments and full-length cDNAs of three members of the abundant human actin gene family (1.3–1.9 kb), the medium abundance transferrin receptor mRNA (5.0 kb) and the low-medium abundance insulin-like growth factor II receptor mRNA (9.1 kb).

INTRODUCTION

Obtaining a full-length cDNA is one of the most important and often one of the most difficult tasks in characterizing a gene. Traditional methods usually produce only partial cDNA fragments; these include constructing and screening cDNA libraries, cloning closely related members of a gene family using polymerase chain reaction (PCR) and degenerate primers, identification of differentially expressed mRNAs by differential display (17) or RNA fingerprinting (27) and cloning of expressed sequence tags (ESTs) by single-pass sequencing method (1).

To facilitate recovery of the rest of the coding sequence, an in vitro method for the rapid amplification of cDNA ends (RACE) was proposed in 1988 (12). Various modifications and improvements of the RACE procedure have been developed and successfully used (4,5,9,11,15,19,27). However, several shortcomings remain: a high background of nonspecific PCR products is usually generated, often making it necessary to do a second round of PCR with nested primers to achieve greater specificity (11); inefficient homopolymer tailing (12) or single-stranded oligonucleotide anchor ligation to the 3′ end of the first-strand cDNA can lead to low efficiency 5′ RACE for cDNAs with strong secondary structure at the 3′ end (4).

To circumvent these drawbacks, we have developed a modified RACE procedure. This technology is based on the ligation of a specially designed double-stranded (ds) adaptor to both ends of the ds cDNA, followed by subsequent amplification of 5′- or 3′-RACE fragments using a combination of gene-specific and adaptor-specific primers and "long and accurate" (LA) PCR technology (3). We studied the utility of this strategy for cloning RACE cDNA fragments and full-sized cDNAs up to 9 kb and mapping of 5′ ends of several different human mRNAs expressed at different levels in poly(A)$^+$ RNA.

Tecan Ex. 2043, Page 2 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

MATERIALS AND METHODS

Oligonucleotides

All oligonucleotides were synthesized on a MilligGen/Biosearch Model 8700 DNA Synthesizer (Burlington, MA, USA) using standard phosphoramidite chemistry and were purified by electrophoresis on 20% denaturing polyacrylamide gels (24). The following sequences of oligonucleotides were used in this study: cDNA synthesis primer (CDS): 5′-TCTAGAATTCAGCGGCCGC$(T)_{30}$VN-3′, where V = G or A or C; N = G or A or T or C. Adaptor: longer strand: 5′-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3′; shorter strand: 5′-PO_4-ACCTGCCC-NH_2-3′. Adaptor primer 1 (AP1): 5′-CCATCCTAATACGACTCACTATAGGGC-3′; nested adaptor primer 2 (AP2): 5′-ACTCACTATAGGGCTCGAGCGGC-3′. Human actin gene family PCR primers: ACT1: 5′-ACTCGTCATACTCCTGCTTGCTGATCCACATCTGC-3′; ACT2: 5′-ACCTGACTGACTACCTCATGAAGATCCTCA-3′. Human transferrin receptor cDNA PCR primers: TFR1: 5′-GTCAATGTCCCAAACGTCACCAGAGA-3′; TFR2: 5′-CTGCCAGCTTTACTGGAGAACTTGA-3′; 5′-TFR: 5′-GGCGGCTCGGGACGGAGGACGCGCTA-3′; 3′-TFR: 5′-ACCTCCAAAAGGCCCTACTCCTTAACGA-3′. Human insulin-like growth factor II receptor cDNA PCR primers: IGFR1: 5′-TTCAGTTTGGGTTTCTGCCTCACAGC-3′; IGFR2: 5′-TCCCATTCAGACAACGACGGATACAGACCA-3′; IGFR3: 5′-TGGAAGTGTGGATATTGTCCAGTGC-3′; IGFR4: 5′-ACAGGAAGGCAATGCTGCTCTGGA-3′; IGFR5: 5′-AGGTCGTACTCATTTCCAGCCAGGTCGGTGA-3′; IGFR6: 5′-AGGGACGAGAGGTCGAAGGAGTTGCCA-3′; 5′-IGFR: 5′-TCCCGCTCCGTCTCCACCTCCGC-3′; 3′-IGFR: 5′-CAGGGCGGTTTGCTTCTCAGCAATAGA-3′. Human insulin-like growth factor I receptor cDNA primers: 5′-iGFR: 5′-TCGCTGTGGGGGCTCCTGTTTCTCTCCGC-3′, 3′-iGFR: 5′-TCCCCACCTTCACTGAGAGAAAGACCTA-3′.

Preparation of Adaptor-Ligated ds cDNA

Adaptor-ligated cDNA was synthesized using reagents supplied in the Marathon™ cDNA Amplification Kit (CLONTECH Laboratories, Palo Alto, CA, USA) following the supplier's protocol. Alternatively, the following protocol was used. One picomol of cDNA synthesis primer was annealed to 1 µg of human placenta poly$(A)^+$ RNA or total RNA (CLONTECH), in a volume of 5 µL of deionized water, by heating the mixture for 2 min at 70°C, followed by cooling on ice for 2 min. First-strand cDNA synthesis was then initiated by mixing the annealed primer RNA with 200 U of Moloney murine leukemia virus (M-MLV) RNase H^- reverse transcriptase (SuperScript™ II reverse transcriptase; Life Technologies, Gaithersburg, MD, USA) in a final volume of 10 µL, containing 50 mM Tris-HCl (pH 8.3 at 22°C), 75 mM KCl, 6 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 1 mM of each dATP, dGTP, dCTP and dTTP. The first-strand cDNA synthesis reaction was incubated at 42°C for 1.5 h in an air incubator and then cooled on ice.

We also synthesized first-strand cDNA using random $d(N)_6$ primers (500 ng) or gene-specific primers IGFR4 (0.5 pmol) or IGRF5 (0.5 pmol) instead of the cDNA synthesis

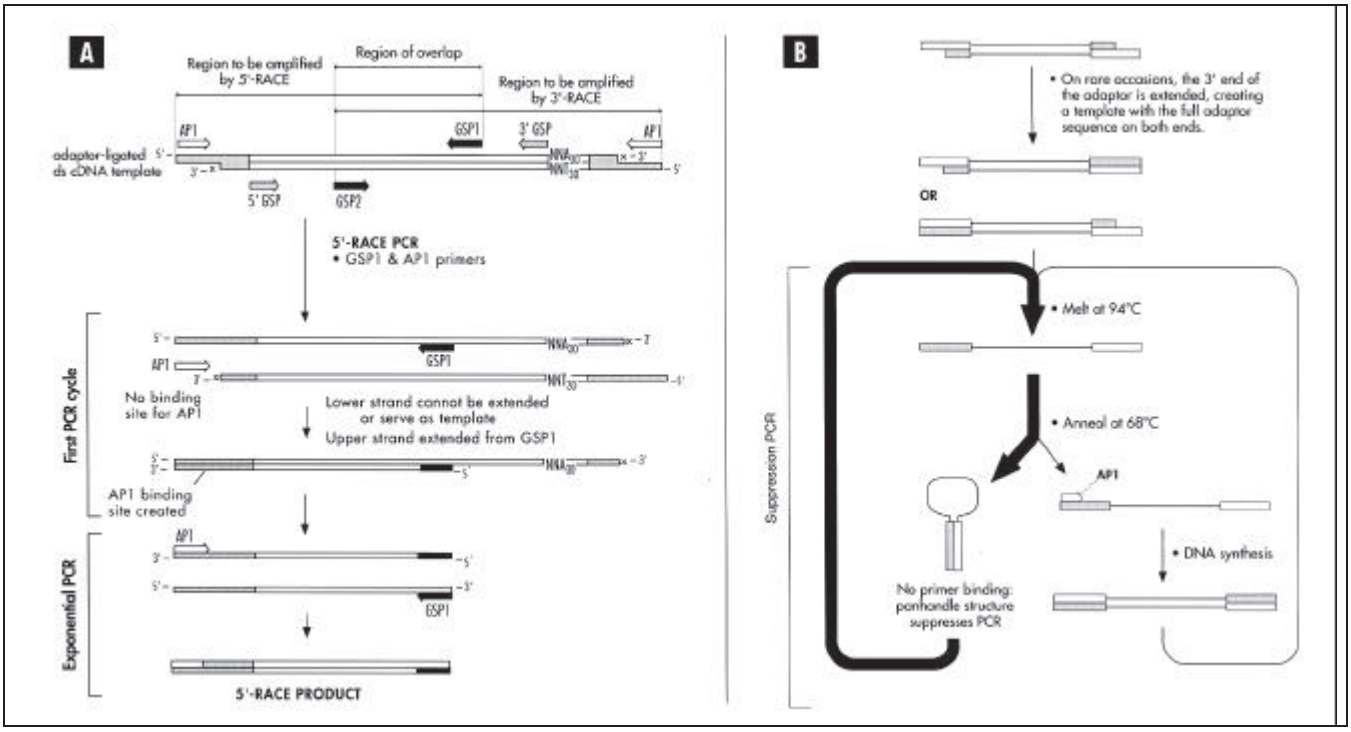

Figure 1. Strategy for RACE cDNA amplification of the 5′ and 3′ ends of specific cDNAs. cDNA synthesis from poly$(A)^+$ RNA and subsequent adaptor ligation creates a population of cDNAs with the structure depicted above. (A) 5′-RACE amplification primed by extension of GSP1 using the upper strand as a template. This generates an AP1 primer binding site at the 3′ end of the newly synthesized strand, allowing subsequent exponential amplification of the cDNA fragment flanked with AP1 and GSP1. 3′-RACE amplification is primed by extension of GSP2 using the bottom strand as a template. This generates an AP1 primer binding site and allows subsequent exponential amplification of the cDNA fragment flanked with AP1 and GSP2. White lines, cDNA; gray lines, adaptor; NNT_{30}, cDNA synthesis primer; arrows, PCR primers. (B) Mechanism of the suppression PCR effect. cDNA molecules that contain the adaptor sequence at both ends cannot be efficiently extended by the shorter AP1 primer as they will form "panhandle-like" structures after each denaturation/annealing cycle.

Tecan Ex. 2043, Page 3 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

primer. Second-strand cDNA synthesis was carried out in a total volume of 80 µL, containing 10 µL of the first-strand cDNA synthesis reaction mixture, 20 mM Tris-HCl (pH 7.5 at 22°C), 100 mM KCl, 10 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 0.15 mM β-nicotinamide-adenine dinucleotide (NAD), 50 µg/mL bovine serum albumin (BSA), 0.2 mM of each dATP, dGTP, dCTP and dTTP, 300 U/mL *E. coli* DNA polymerase I, 12 U/mL *E. coli* RNase H and 60 U/mL *E. coli* DNA ligase. The reaction mixture was incubated at 16°C for 1.5 h, then 2 µL of 3 U/µL T4 DNA polymerase (New England Biolabs, Beverly, MA, USA) were added and additionally incubated at 16°C for 30 min and stopped by addition of 4 µL of 2 mg/mL glycogen, 0.2 M EDTA. The ds cDNA was extracted once with phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol), once with chloroform/isoamyl alcohol (24:1, vol/vol) and then precipitated by addition of one-half volume of 4 M ammonium acetate (about 35 µL) and 3.7 vol of 95% ethanol (about 260 µL).

After vortex mixing, the tube was immediately centrifuged at 14 000 rpm (16 000× *g*) in an Eppendorf® Model 5415C microcentrifuge (Brinkmann Instruments, Westbury, NY, USA) for 20 min. The pellet was washed with 80% ethanol, centrifuged as above for 10 min, air-dried and dissolved in 10 µL of deionized water. The ds cDNA was then ligated to an adaptor overnight at 16°C under the following conditions: 5 µL of ds cDNA solution, 50 mM Tris-HCl (pH 7.8 at 22°C), 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 5% polyethylene glycol (mol wt 8000), 2 µM of adaptor and 1 U of T4 DNA ligase (Life Technologies) in a total volume of 10 µL. The ligation mixture was then diluted 100-fold by addition of 1 mL of 10 mM Tricine-KOH (pH 8.5 at 22°C), 0.1 mM EDTA, incubated at 94°C for 2 min, cooled on ice and stored at -20°C.

5′, 3′ RACE and Full-Length cDNA Amplification

PCR amplification was performed using the Advantage™ cDNA PCR Kit (CLONTECH). This kit contains a mixture of KlenTaq-1 (3) and Deep $Vent_R$ ® DNA Polymerases (New England Biolabs) and TaqStart™ antibody (CLONTECH). The TaqStart antibody provides automatic "hot-start" PCR (16). Amplification was conducted in a 50-µL volume containing 5 µL of diluted adaptor-ligated ds cDNA, 40 mM Tricine-KOH (pH 9.2 at 22°C), 3.5 mM $Mg(OAc)_2$, 10 mM KOAc, 75 µg/mL BSA, 200 µM of each dATP, dGTP, dCTP and dTTP, 0.2 µM of each adaptor primer 1 (AP1) and gene-specific primer (GSP1 for 5′ RACE or GSP2 for 3′ RACE; see Figure 1) and 1 µL of 50× Advantage KlenTaq Polymerase Mix (CLONTECH). Temperature parameters of the PCRs were as follows: 1 min at 94°C followed by 23–30 cycles of 94°C for 30 s and 68°C for 5 min; followed by a 10-min final extension at 68°C. For 5′ and 3′ RACE of rare cDNAs (insulin-like growth factor I receptor), we used a "touchdown PCR" program: 1 min at 94°C followed by 5 cycles of 94°C for 30 s and 72°C for 5 min; then 5 cycles of 94°C for 30 s and 70°C for 5 min; then 25 cycles of 94°C for 30 s and 68°C for 5 min; followed by a 10-min final extension at 68°C. Full-length cDNAs were amplified directly from adaptor-ligated cDNA using a combination of flanking 5′- and 3′-gene-specific primers (corresponding to the 5′ and 3′ ends of each cDNA) and the same conditions and cycle parameters used for 5′- and 3′-RACE PCRs, except a 10-min extension time. PCR products were examined on 1.2% agarose/EtdBr gels in 1× TBE buffer containing 90 mM Tris-borate, 2 mM EDTA (pH 8.0 at 22°C). As a DNA size marker, we used a 1-kb DNA Ladder (Life Technologies).

Analysis of RACE Products

Aliquots (10-µL) of RACE reaction products were separated by electrophoresis as described above, transferred to Maximum Strength Nytran® Nylon (Schleicher & Schuell, Keene, NH, USA) hybridized with a ^{32}P-labeled oligonucleotide probe in ExpressHyb™ Hybridization Solution (CLONTECH) in accordance with the supplier's recommendations. Bands corresponding to gene-specific RACE products were purified from agarose gel slices and cloned directly into a TA Cloning® vector (Invitrogen, San Diego, CA, USA) following the supplier's recommendations. Plasmids with gene-specific RACE product inserts were identified by low-density (ca. 100 colonies/plate) colony lift hybridization (24) with a ^{32}P-labeled oligonucleotide probe. Plasmid DNAs were isolated using QIAwell™ 8 Plus Plasmid Kit (Qiagen, Chatsworth, CA, USA), and inserts were sequenced with the *fmol*® DNA Sequencing System (Promega, Madison, WI, USA), using the supplier's recommendations.

Single-Stranded (ss)DNA Secondary Structure Prediction

ssDNA secondary structure prediction and energy calculation for 5′ end of IGFRII cDNA (1–300 bases) were carried out using DNASIS™ (Version 7.13) software program (Hitachi Software Engineering America, San Bruno, CA, USA).

RESULTS AND DISCUSSION

RACE cDNA Amplification Protocol

In our attempts to improve the efficiency and reliability of

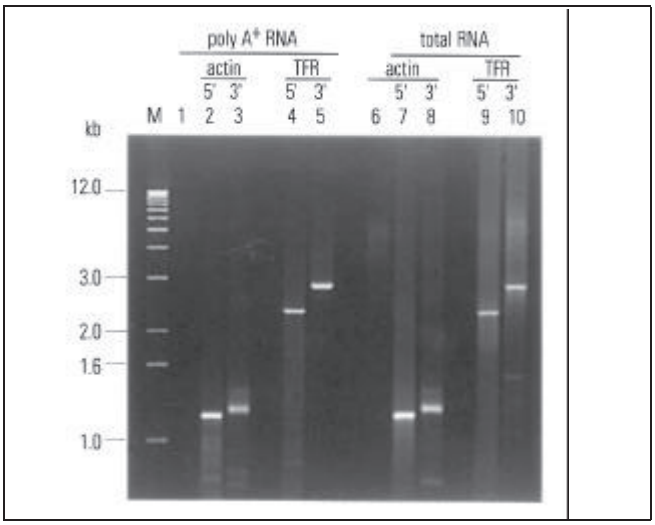

Figure 2. 5′- and 3′-RACE amplification of actin and transferrin receptor cDNAs using human placenta poly(A)$^+$ or total RNA as a template for cDNA synthesis. Lanes 1–5: poly(A)$^+$ RNA template and 25 PCR cycles. Lanes 6–10: total RNA template and 27 PCR cycles. Lanes 1 and 6: control RACE reactions primed with AP1 alone. Lanes 2 and 7: 1.2-kb 5′-RACE actin fragment generated with ACT1 and AP1 primers. Lanes 3 and 8: 1.3-kb 3′-RACE actin fragment amplified with ACT2 and AP1 primers. Lanes 4 and 9: 2.5-kb 5′-RACE TFR product generated with TFR1 and AP1 primers. Lanes 5 and 10: 2.8-kb 3′-RACE TFR fragment amplified with TFR2 and AP1 primers. Lane M: DNA size marker.

Tecan Ex. 2043, Page 4 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

5′-RACE amplification independent of the sequence, abundance and size of the target cDNA, we developed a method that combines aspects of the standard adaptor-based technology for cDNA library construction (14). Our cDNA amplification protocol begins with first-strand cDNA synthesis using M-MLV reverse transcriptase and either poly(A)$^+$ or total RNA and a modified lock-docking oligo(dT) primer (6). Following second-strand cDNA synthesis using a combination of *E. coli* DNA polymerase I, *E. coli* RNase H and *E. coli* DNA ligase (14), blunt ends are created by T4 DNA polymerase, and a ds adaptor is ligated at both ends of the ds cDNA using T4 DNA ligase (Figure 1A). At this step, the pool of adaptor-ligated ds cDNA is essentially an uncloned cDNA library from which many genes can be amplified using different sets of primers. Both 5′- and 3′-RACE PCRs are primed with a gene-specific internal primer (GSP1 for 5′ RACE or GSP2 for 3′ RACE) and an adaptor primer 1 (AP1) (Figure 1A).

Our cDNA amplification procedure uses blunt-end ligation of a ds adaptor to the 5′ end of ds cDNA by T4 DNA ligase. This method appears to be more reproducible and more efficient than ss anchor ligation to 3′ end of ss cDNA or 5′ end of mRNA by T4 RNA ligase or homopolymeric tailing using terminal transferase [see also Bertling et al. (4)].

T4 DNA ligase has a lower substrate K_m [5×10^{-5} M (20)] than T4 RNA ligase [$K_m > 10^{-3}$ M (7)]. Thus, at the typical levels of substrate concentration [10^{-7}–10^{-6} M cDNA or poly(A)$^+$ RNA], T4 DNA ligase gives much higher efficiency ligation of ds adaptor to ds cDNA than T4 RNA ligase-mediated ligation of ss anchor to the 3′ end of ss cDNA (9), ss anchor ligation to the 5′ end of decapped mRNA (5,27) or head-to-tail ligation of ss cDNA (15).

In addition, blunt-end ds ligation efficiency is less influenced by the sequence of a particular cDNA. In contrast, strong secondary structure at the 3′ end of some ss cDNAs (the 5′ end of the mRNA) can prevent efficient anchor ligation by T4 RNA ligase or tailing by terminal transferase.

The special design of the ds adaptor significantly reduces nonspecific amplification by combining "vectorette PCR" (2) with a newly developed method termed "suppression PCR" (18,25), which is based on a previously described phenome-

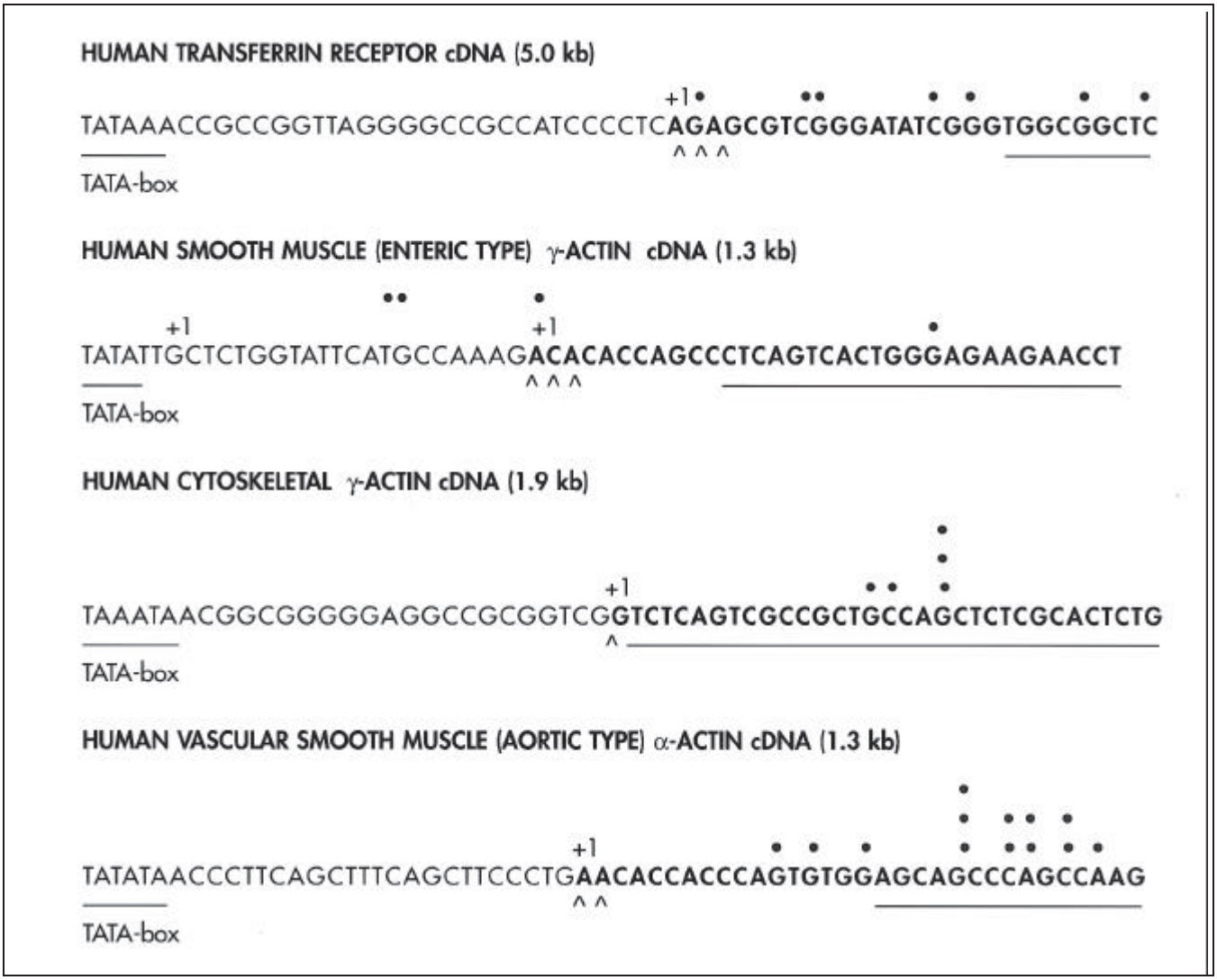

Figure 3. Mapping the 5′ ends of 5′-RACE fragments of transferrin receptor and actin gene family cDNAs. Bold letters show mRNA sequence, and light letters show genomic sequence based on published data (10,21,23,26). The TATA promoter elements and the longest cDNAs cloned from cDNA libraries are underlined. 5′ ends of randomly cloned 5′-RACE fragments are marked by dots. The cap-site position (+1) based on primer extension assay (10) or S1 nuclease protection assay (21) are marked by carats (^). For vascular smooth muscle γ-actin gene, +1 (at -21G) is an alternative cap-site identified by anchor ligation RACE technology.

Tecan Ex. 2043, Page 5 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

non of inefficient amplification of DNA fragments flanked by inverted terminal repeats (13,28).

The vectorette feature of the adaptor is the absence of an AP1 primer binding site. The AP1 primer is complementary to the missing strand of the single-stranded portion of the adaptor (Figure 1A). Furthermore, the 3′ end of the shorter strand of the adaptor is blocked with an amine group and therefore cannot be extended to create AP1 binding sites in the general cDNA population. The absence of the AP1 binding site in the general cDNA population significantly reduces nonspecific linear amplification by AP1 alone from the distal ends of the cDNA population, which is a common problem for the most other RACE protocols (11). During the first round of thermal cycling, the inner, gene-specific primer (GSP) is extended, creating an AP1 binding site at the 5′ (or 3′) terminus of the cDNA(s), which contains a GSP-binding site (Figure 1A). In subsequent cycles, both AP1 and GSP can bind, allowing exponential amplification of the cDNA of interest.

Nonspecific, end-to-end amplification is reduced by the suppression PCR effect. Rare molecules can appear in the general cDNA population that contain extended adaptor sequence at one end (Figure 1B). These molecules contain long inverted terminal repeats at both ends and will form a stable "panhandle-like" structure during each denaturation-annealing PCR cycle. The essential feature of suppression PCR is the use of an adaptor primer (AP1) that is about two times shorter than the adaptor (44 nucleotides [nt]) itself, so intramolecular annealing of the longer adaptor sequences is both highly favored and more stable than intermolecular annealing to the much shorter AP1. The resulting panhandle-like structure cannot serve as a template for exponential nonspecific cDNA amplification using AP1 alone. However, when an internal gene-specific primer extends a DNA strand through the adaptor (Figure 1A), the extension product will contain the adaptor sequence only at one end and thus cannot form the panhandle structure and will be amplified exponentially.

Additionally, a nested AP2 primer can be used in a second round of PCR (13) as a means of increasing sensitivity and reducing background.

The cDNA amplification procedure described here requires optimized LA PCR technology (3). LA PCR, which uses a combination of thermostable DNA polymerases, allows amplification of longer RACE products, including full-length cDNAs (in our experiments up to 9 kb) with an error rate that has been reported to be about 10% lower than that obtained with conventional PCR (3).

Amplification of 5′- and 3′-cDNA Fragments

To assess the efficiency of our RACE cDNA amplification procedure, we carried out separate 5′- and 3′-RACE amplifications using adaptor-ligated ds cDNA synthesized from human placenta poly(A)$^+$ RNA as a template and a combination of AP1 and GSP1 (for 5′ RACE) or AP1 and GSP2 (for 3′ RACE). Gene-specific primers were designed approximately in the middle of a conserved coding region of abundant (ca. 1%–2%) actin gene family cDNAs (1.3–1.9-kb-long) (21), medium abundance (ca. 0.1%) transferrin receptor cDNA (5.0 kb) (26) and low abundance (ca. 0.01%) insulin-like growth factor II receptor cDNA (9.1 kb) (22).

5′- and 3′-RACE amplification of actin and transferrin receptor cDNAs synthesized from poly(A)$^+$ RNA generated single major bands of predicted size (Figure 2, lanes 2–5) with very little nonspecific amplification (lane 1). When total

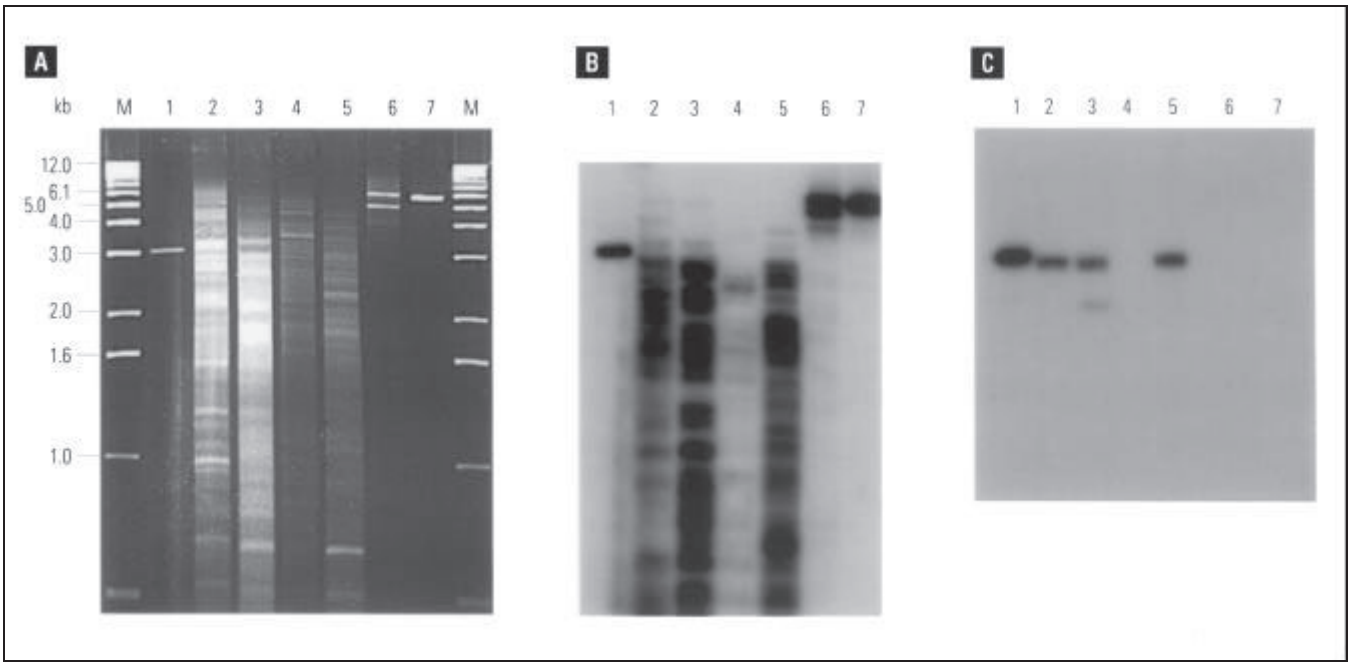

Figure 4. 5′- and 3′-RACE amplification of IGFRII receptor cDNA. For map and position of primers see Figure 5. (A) Different samples of cDNAs were synthesized using CDS primer (lanes 1, 2, 6 and 7), random primer d(N)$_6$ (lane 3) or gene-specific primer [IGFR5 (lane 4) or IGFG6 (lane 5)]. For 5′ RACE, we used primers IGFR1 and AP1 (lanes 2–6), or IGFR1 and IGFR3 (positive control, lane 1). For 3′ RACE, primers were IGFR2 and AP1 (lane 6) or IGFR2 and 3′-IGFR (positive control, lane 7). Lane M: DNA size marker. (B) Size distribution of 5′- and 3′-RACE gene-specific products. PCR fragments (Panel A) were transferred to nylon membrane and hybridized with ^{32}P-labeled IGFR2. (C) Mapping of the longest 5′-RACE cDNA amplification products. PCR fragments (Panel A) were transferred to nylon membrane and hybridized with ^{32}P-labeled IGFR3.

Tecan Ex. 2043, Page 6 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

RNA was used as a template for cDNA synthesis, single, major 5′- and 3′-RACE products were also generated (lanes 7–10), but the level of nonspecific amplification was somewhat higher (lane 6). Thus, total RNA should be used only for the amplification of short (up to 2–3 kb), abundant [more than 0.1% of poly(A)$^+$ RNA] mRNAs.

Subsequent cloning of the 5′-RACE product (2.8 kb) for transferrin receptor and sequence analysis of 7 randomly picked clones (identified by hybridization with a ^{32}P-labeled TFR2 probe) showed that all of the 5′-RACE products extended to within 1–28 nt of the 5′ end of mRNA (Figure 3). D'Alessio and Gerard (8) have shown that combined action of T4 DNA polymerase and RNase H often removes 1–30 nt from the ends of cDNA during creation of blunt ends.

To study amplification of cDNAs belonging to a multigene family, we designed actin-specific primers (ACT1 and ACT2) to a conserved part of the protein coding region of actin cDNA. Then we isolated, cloned and sequenced the most abundant 5′- and 3′-RACE actin cDNA products. Twelve random clones of the 3′-RACE product (Figure 2, lane 3) corresponded in size (1.35 kb) and sequence to the 3′-end fragment of the human cytoskeletal γ-actin cDNA isoform found in human placenta (10). On the other hand, sequence analysis of 25 random clones of the 5′-RACE product (lane 2) revealed the presence in the same band of at least 3 human actin cDNA isoforms: smooth muscle (enteric type) γ-actin, cytoskeletal γ-actin and vascular smooth muscle (aortic type) α-actin (Figure 3). The expected size of 5′-RACE products of these cDNAs are 1.20, 1.22 and 1.19 kb, respectively (10,21,23). Thus, the 5′-RACE products for all three actin isoforms co-migrate on a 1.2% agarose gel. As in the case of the transferrin receptor, mapping the 5′ ends of cloned 5′-RACE fragments for cytoskeletal γ-actin and vascular smooth muscle α-actin revealed that all 18 clones lost less than 30 nt from the 5′-noncoding mRNA sequence. Unexpectedly, two out of four smooth muscle γ-actin 5′-RACE fragments were longer than the previously mapped cap-site for this gene (21), and this extended sequence corresponded to genomic sequence just upstream of the regular cap-site. This finding is probably due to the presence of an alternative transcriptonal start site in the promoter region of vascular smooth muscle γ-actin gene. To confirm these data, we mapped the alternative transcriptional start site at -21G (Figure 3) from the regular cap-site using ligation-anchored PCR technology (9) (data not shown).

These sequencing data show that although it may be difficult to accurately determine transcriptional start sites using cDNA amplification technology described here, in most cases, the size of RACE products are significantly longer than cDNAs isolated from cDNA libraries.

We then performed 5′ and 3′ RACE on the insulin-like growth factor II receptor (IGFRII) gene. This was chosen because it is rather long (9.0 kb) and is expressed at a low level. Figure 4 shows the result of separate amplifications of the 5′ and 3′ ends of the IGFRII cDNA (see Figure 5 for position of the primers). 3′-RACE amplification generated two bands; both of which were found to be gene-specific based on results of Southern blot analysis (Figure 4B, lanes 6 and 7). The larger band of about 6.5 kb corresponds to the expected length of full-length 3′-RACE product. The shorter one appears to be a non-full-sized PCR RACE product. This was confirmed by subsequent cloning and sequencing (data not shown).

5′-RACE produced multiple bands (Figure 4A, lane 2).

Tecan Ex. 2043, Page 7 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

Based on results of Southern blot analysis (Figure 4B, lane 2), these include both gene-specific and nonspecific bands. In accordance with the data of Frohman (11), we believe that many of the gene-specific 5′-RACE products are probably due to premature termination of reverse transcription in the course of first-strand cDNA synthesis, which results in non-full-sized RACE templates. The longest gene-specific 5′-RACE products (about 3 kb, Figure 4, B and C, lane 2) are slightly shorter than the expected full-length 5′-RACE product (3.2 kb). Priming cDNA synthesis with a random primer (Figure 4, A–C, lane 3) or different gene-specific primers (Figure 4, A–C, lanes 4 and 5) instead of oligo(dT) did not significantly improve the yield or size of the longest 5′-RACE product (Figure 4C, lanes 2–5). Moreover, inclusion of 10% dimethyl sulfoxide (DMSO) in the PCR buffer, replacement of dGTP for 7-methyl-dGTP, or using other combinations of thermophilic enzymes and buffers developed for LA PCR (*Taq* DNA Polymerase + *Pwo* DNA Polymerase, ExTaq DNA Polymerase or r*Tth* DNA Polymerase, XL) also did not improve the yield or size of the longest gene-specific 5′-RACE product.

Sequence analysis of 15 random clones of the longest RACE product revealed that all of the amplified 5′-RACE fragments lacked 250–270 nt from the 5′ end of the mRNA (Figure 5). Sequence analysis of the 5′ end of IGFRII cDNA (Figure 5) showed that there is a 170-nt G/C-rich sequence (90% G/C), which can form a strong secondary structure (ΔG = -227 kcal/mol). Such a structure could cause termination of reverse transcription or directly prevent amplification of full-length 5′-RACE IGFRII fragments. To distinguish between these two possibilities, we analyzed cDNA generated by extension of the IGFR4 primer, which is 0.55 kb from the 5′ end of IGFRII mRNA (Figure 5). We used M-MLV (RNase H^-) reverse transcriptase with placenta poly$(A)^+$ RNA as a template. We found that reverse transcriptase could read through the G/C-rich sequence and generate a full-length cDNA extension product with the expected size (550 nt) with at least 50% yield (data not shown). However, attempts to amplify this cDNA using primers IGFR4 and 5′ IGRF generated only a shorter DNA fragment (about 250 bp), which was similar in size to the PCR fragment generated by primers IGFR4 and IGFR3 (Figure 5). Subsequent sequence analysis of this shorter PCR fragment revealed that it corresponded to a 5′ fragment of IGFRII cDNA with a deletion of the G/C-rich sequence and flanked by primers IGFR4 and 5′ IGFR from both ends (data not shown). These results agree with the data of Bertling et al. (4) and indicate that extended G/C-rich sequences can, in some cases, block amplification or result in amplification and cloning of incorrect, although similar, cDNAs. In these "difficult" cases, the sequence of the ampli-

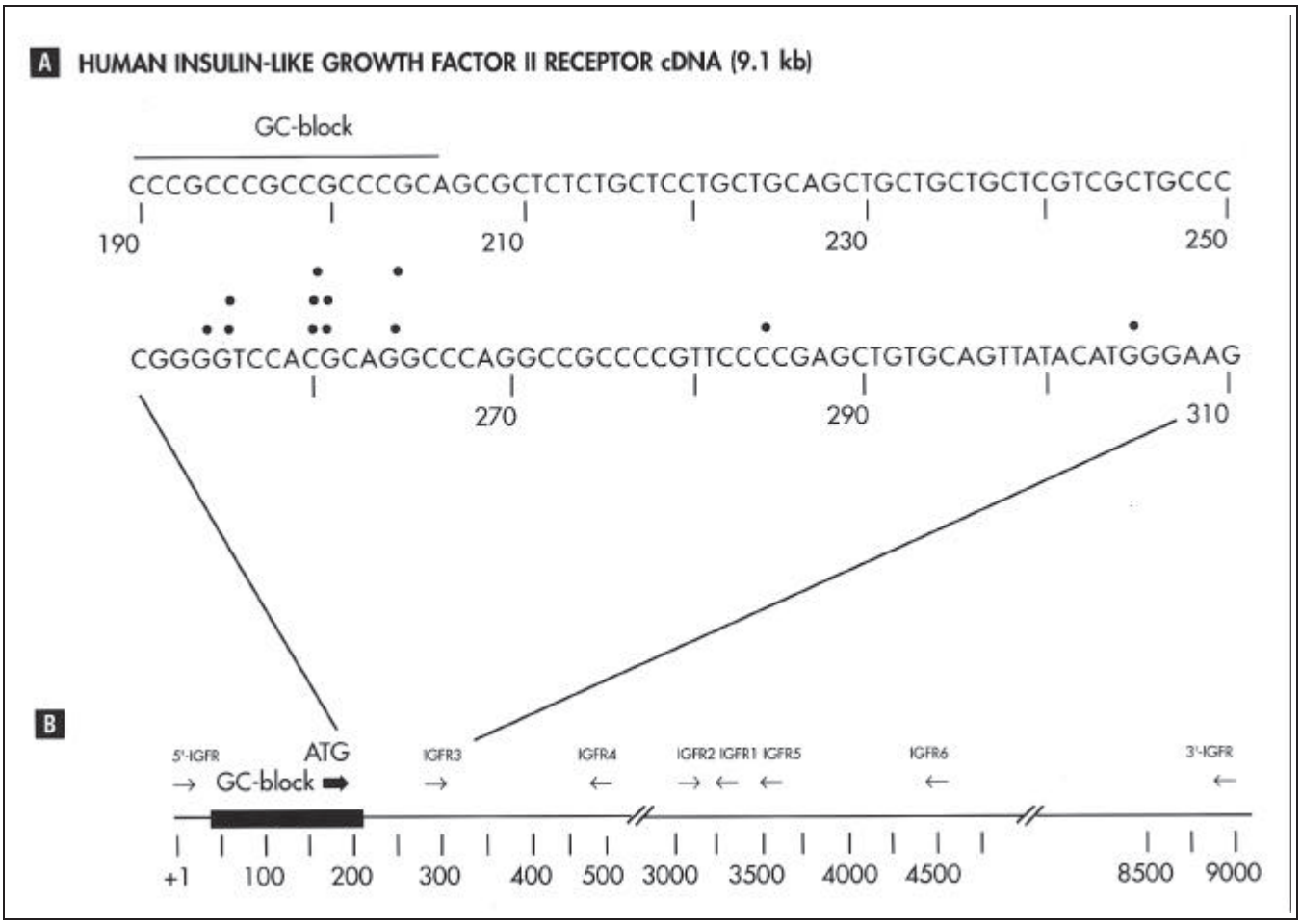

Figure 5. Mapping the 5′-termini of 5′-RACE products of IGFRII receptor cDNA. (A). The IGFRII mRNA sequence at the 5′ end (22) shows the position of the 5′ ends of randomly cloned 5′-RACE fragments (dots) and the 3′ end of the G/C-rich sequence. Numbers reflect nucleotide positions relative to the 5′ end of the longest cDNA fragment cloned from cDNA library. (B). The IGFRII mRNA (9.1-kb-long) map with positions of gene-specific primers (arrows), the G/C-rich sequence near the 5′ end and the translation start site (ATG).

Tecan Ex. 2043, Page 8 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

fied RACE product should be confirmed by comparison to the structure of cDNA fragments cloned from conventional cDNA libraries or genomic sequence.

Generation of Full-Length cDNA by PCR

The full-length cDNA can be directly amplified from the pool of adaptor-ligated cDNA by LA PCR using two gene-specific primers from the extreme 5′ and 3′ ends of the cDNA sequence (5′ GSP and 3′ GSP, Figure 1). When characterizing novel genes, this approach requires partial sequencing of the 5′ and 3′ end of the RACE products in order to design PCR primers to the 5′ and 3′ end of the target gene. We designed 5′- and 3′-GSP primers that began within 30 nt of the 5′ and 3′ ends of the mRNA for TFR, IGFRI and IGFRII. Using these primers, we successfully amplified full-length 5.0-kb TFR cDNA (Figure 6, lane 3), full-length 5.0-kb IGFRI cDNA (lane 5) and nearly full-length 8.8-kb cDNA for IGFRII cDNA (lane 4). In all cases, the PCR products corresponded to the expected size, and we confirmed the identity of PCR products by subsequent cloning and partial sequencing.

Our results show that the RACE cDNA amplification protocol described here can be successfully used for cDNA cloning of high-, medium- and low-abundant cDNAs of up to 9 kb. Unlike other RACE protocols, a second, nested PCR step is rarely needed. On the other hand, our results with IGFRII cDNA show that background can still be high for RACE of certain rare cDNAs [0.01%–0.001% of poly(A)$^+$ RNA]. In these cases, Southern blot analysis, screening of gene-specific clones after cloning of RACE products or

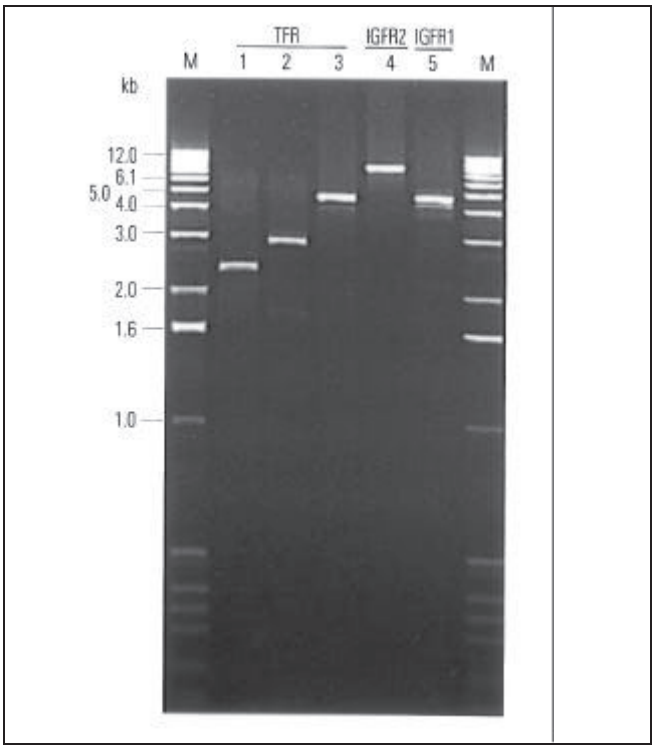

Figure 6. Generation of full-length cDNA by end-to-end amplification from adaptor-ligated ds cDNA. Lane 1: 5′ RACE for TFR cDNA; lane 2: 3′ RACE for TFR cDNA; lane 3: full-length TFR cDNA (5.0 kb) amplified with primers 5′ TFR and 3′ TFR; lane 4: nearly full-length IGFRII cDNA (8.8 kb) amplified with primers IGFR3 and 3′ IGFR; lane 5: full-length IGFR1 cDNA (5.0 kb) amplified with primers 5′ iGFR and 3′ iGFR; lane M: DNA size marker.

Tecan Ex. 2043, Page 9 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016

nested PCR is required to identify the correct RACE products. In spite of limitations in current PCR technology itself (such as, low accuracy and difficulty to amplify G/C-rich sequences), our cDNA amplification procedure is a fast, simple and reliable alternative to conventional library screening for obtaining full-length cDNAs.

ACKNOWLEDGMENTS

We wish to thank professor Eugeney Sverdlov for many helpful discussions, David Gunn and Paul Kitts for critical review of the manuscript, and Theresa Provost and David Gunn for preparation of the figures.

REFERENCES

1. **Adams, M.D., M.B. Soares, A.R. Kerlavage, C. Field and J.C. Venter.** 1993. Rapid cDNA sequencing (expressed sequencing tags) from a directionally cloned human infant brain cDNA library. Nature Genet. *4*:373-380.
2. **Arnold, C. and I.J. Hogson.** 1991. Vectorette PCR: a novel approach to genome walking. PCR Methods Appl. *1*:39-42.
3. **Barnes, W.M.** 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage template. Proc. Natl. Acad. Sci. USA *91*:2216–2220.
4. **Bertling, W.M., F. Beier and E. Reichenberger.** 1993. Determination of 5′ ends of specific mRNAs by DNA ligase-dependent amplification. PCR Methods Appl. *3*:95-99.
5. **Bertrand, E., M. Fromont-Racine, R. Pictet and T. Grande.** 1993. Visualization of the interaction of a regulatory protein with RNA *in vivo*. Proc. Natl. Acad. Sci. USA *90*:3496-3500.
6. **Borson, N.D., W.L. Sato and L.R. Drewes.** 1992. A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. PCR Methods Appl. *2*:144–148.
7. **Brenman, C.A., A.E. Manthey and R.I. Gumport.** 1983. Using T4 RNA ligase with DNA substrates, p. 38-52. *In* R. Wu (Ed.), Methods in Enzymology, Vol. 100. Academic Press, San Diego.
8. **D'Alessio, J.M. and G.F. Gerard.** 1988. Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H: the fate of information at the 5′ terminus and the effect of *E. coli* DNA ligase. Nucleic Acids Res. *16*:1999-2014.
9. **Dumas, J.B., M. Edwards, J. Delort and J. Mallet.** 1991. Oligonucleotide ligation to ss cDNA: a new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. Nucleic Acids Res. *19*:5227-5232.
10. **Erba, H.P., P. Gunning and L. Kedes.** 1986. Nucleotide sequence of the human γ cytoskeletal actin mRNA: anomalous evolution of vertebrate nonmuscle actin genes. Nucleic Acids Res. *14*:5275-5294.
11. **Frohman, M.A.** 1994. On beyond RACE. PCR Methods Appl. *4*:S40-S58.
12. **Frohman, M.A., M.K. Dush and G.R. Martin.** 1988. Rapid production of full-length cDNA from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA *85*:8998-9002.
13. **Frohman, M.A. and G.R. Martin.** 1989. Rapid amplification of cDNA ends using nested primers. Technique *1*:165-170.
14. **Gubler, U. and B.J. Hoffman.** 1983. A simple and very efficient method for generating complimentary DNA libraries. Gene *25*:263-269.
15. **Hofmann, M.A. and D.A. Brian.** 1991. A PCR enhanced method for determining the 5′ end sequence of the mRNAs. PCR Methods Appl. *1*:43-45.
16. **Kellogg, D.E., I. Rybalkin, S. Chen, N. Mukhamedova, T. Vlasik, P. Siebert and A. Chenchik.** 1994. TaqStart Antibody™: "Hotstart" PCR facilitated by a neutralizing monoclonal antibody directed against *Taq* DNA polymerase. BioTechniques *16*:1134-1137.
17. **Liang, P. and A. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science *257*:967-970.
18. **Lukianov, S., N.K. Gurskaya, K.A.. Lukianov, V.S. Tarabukin and E.D. Sverdlov.** 1994. Highly efficient subtractive hybridization of cDNA. Bioorg. Chem. (Russia) *20*:701-704.
19. **Maruyama, K. and S. Sugano.** 1994. Oligo-capping: a simple method to replace the cap structure of eukariotic mRNAs with oligoribonucleotides. Gene *138*:171-174.
20. **Maunders, M.** 1993. DNA and RNA ligases, p. 213-230. *In* M.M. Burrell (Ed.), Methods of Molecular Biology. Enzymes of Molecular Biology, Vol. 16. Humana Press, Totowa, NJ.
21. **Miwa, T., Y. Manabe, K. Kurokawa, S. Kamada, N. Kanda, G. Bruns, H. Ueyama and T. Kakunaga.** 1991. Structure, chromosome location, and expression of the human smooth muscle (enteric type) γ-actin gene: evolution of six human actin genes. Mol. Cell. Biol. *11*:3296-3306.
22. **Morgam, D.O., J.C. Edman, D.N. Standring, V.A. Fried, M.C. Smith, R.A. Roth and W.J. Rutter.** 1987. Insulin-like growth factor II receptor as a multifunctional binding protein. Nature *329*:301-307.
23. **Nakano, Y., T. Nishihara, S. Sasayama, T. Miwa, S. Kamada and T. Kakunaga.** 1991. Transcriptional regulatory elements in the 5′ upstream and first intron regions of the human smooth muscle (aortic type) α-actin encoding gene. Gene *99*:285-289.
24. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. **Siebert, P.D., A. Chenchik, D.E. Kellogg, K.A. Lukyanov and S.A. Lukyanov.** 1995. An improved method for walking in uncloned genomic DNA. Nucleic Acids Res. *23*:1087-1088.
26. **Schneider, M.J., M.J. Owen, D. Banville and J.G. Williams.** 1984. Primary structure of human transferrin receptor deduced from the mRNA sequence. Nature *311*:675-678.
27. **Volloch, V., B. Schweitzer and S. Rits.** 1994. Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of β-globin mRNA with an extended 5′-untranslated region. Nucleic Acids Res. *22*:2507-2511.
28. **Welsh, J. and M. McClelland.** 1991. Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. Nucleic Acids Res. *19*:5275-5279.
29. **Welsh, J. and A. McClelland.** 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res. *20*:4965-4970.

Address correspondence to Alex Chenchik, CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA. Internet: aachenchik@clontech.com

Tecan Ex. 2043, Page 10 of 10
IDT, Inc. v. Tecan Genomics - IPR2025-00016