Full-Length cDNA Synthesis for Long-Distance RT-PCR of Large mRNA Transcripts

BioTechniques 34:768-773 (April 2003)

ABSTRACT

Reverse transcription of mRNA often leads to the synthesis of partial, non-full-length cDNAs. Methods to facilitate reverse transcription across RNA regions of secondary structure, as well as enzyme modifications to eliminate RNase H activities inherent to reverse transcriptase enzymes, have been previously reported. However, because all reverse transcriptases have high error rates of polymerization, the misincorporation of nucleotides can also cause the reverse transcriptase to stumble. Hence, even in the absence of RNA secondary structure and RNase H activity, the synthesis of full-length cDNA from long mRNA transcripts still remains a challenge. We describe here the coupling of a 3'→5' exonuclease function during reverse transcription. The incorporation of a proofreading activity, when used in conjunction with denaturant buffers and RNase H-deficient reverse transcriptases, can successfully generate full-length cDNAs of up to 15 kb.

INTRODUCTION

The availability of the draft sequences of the human genome has greatly accelerated the identification of novel protein encoding genes (1,2). Sequencing of genomes of other organisms and improvements in gene prediction algorithms have further contributed to the rapid identification of computationally derived protein-encoding transcripts. However, functional characterization of these predicted genes requires the expression of encoded proteins from physical full-length clones. Unfortunately, many of these full-length clone reagents are not available from public or commercial clone repositories. Most current large-scale efforts for obtaining full-length clone reagents rely on clone selection from cDNA libraries on the

Table 1. First-Strand Primer and Reverse Transcriptase Enzymes Used in the Generation of cDNA

Tube	Primer	Reverse Transcriptase
1	Random Hexamer (20 ng)	MMLV RNase H-minus
2	Random Hexamer (20 ng)	AMV
3	Random Hexamer (20 ng)	MMLV RNase H-minus
4	Random Hexamer (20 ng)	MMLV native
5	Oligo d(T) ₂₀ (500 ng)	MMLV RNase H-minus
6	Oligo d(T) ₂₀ (500 ng)	AMV
7	Oligo d(T) ₂₀ (500 ng)	MMLV RNase H-minus
8	Oligo d(T) ₂₀ (500 ng)	MMLV native

The AMV, the wild-type MMLV, or the RNase-deficient (RNase H-minus) MMLV reverse transcriptase enzyme was used. All reactions were performed following manufacturer-supplied protocols, except tubes 3 and 7, which included DMSO as an RNA denaturant and *Pfu* as a proofreading polymerase. Promega supplied all reverse transcriptases.

basis of 5' and 3' expressed sequence tag clone end sequences (3,4). Such a process is only successful for obtaining full-length clones of short transcripts with an average clone insert size of 1.2 kb and a maximum insert size of 6.3 kb (5). Furthermore, as these clones have been prepared from oligo-d(T) priming, the presence of long 3' untranslated regions significantly limits the actual protein-coding length of the clones isolated. Our laboratory has been generating full-length clone reagents for genes not found among our large collection of approximately 10 million sequenced human cDNA clones. We have primarily used RT-PCR to amplify these genes from mRNA using electronic full-length DNA sequences derived from gene prediction software, assembly algorithms, and hand editing. While we have been successful with this technology in isolating even the lowest-abundance transcripts, we found that the amplification of longer transcripts of more than 10 kb required various optimization strategies during the reverse transcription step. Often, multiple rounds of cDNA synthesis using varying conditions or different enzymes were required before a particular gene could be successfully amplified. As the choice of cDNA synthesis condition could not be predetermined for each gene, testing multiple conditions simultaneously or sequentially significantly increased processing time and cost. With the widespread use of robotics and miniature systems, more and more re-

search laboratories now have the capability to perform high-throughput experiments with large clone sets. The ability to use PCR to generate protein-encoding

Figure 1. Long-distance PCR products amplified using cDNA pooled from the eight separate conditions (Table 1) were separated on a 0.8% agarose gel. Lane 1 is a 1-kb DNA extension ladder from Invitrogen. Molecular weights are indicated as marked. Ten microliters of the nested PCR product (see Materials and Methods section) from GenBank accession nos. 1743884, 34811, 13570016, and 34338 were separated on lanes 2–5, respectively. Bands were excised, and purified DNA was cloned and sequenced.

Table 2. Gene Name, GenBank Accession No., mRNA Transcript Length, and Primer Oligonucleotide Sequences Used for PCR

Gene	GenBank Accession No.	mRNA Length	Forward Primer (5'→3')	Nested Forward (5'→3')	Nested Reverse (5'→3')	Reverse Primer (5'→3')
Reelin	1743884	11580	AAAACTGAATGAGCGAGCG GCGCGTGA	ATGGAGCGCAGTGGCTG GGGCCC	TGATTCTTCATGGGTATCG CCTAAGTGA	CACTCGGTCTTGAGAAGG GCTTTCAGGT
Utrrophin	34811	10302	GTCAAGCTGAACCATCGTAG GAGGTTGA	ATGGCCAAAGTATGGAGAA CATGAAGCCA	TACGTCAGGAAACATTGG TTGGCCGGAT	AGCCATGGAGCTTCTGAT TTAATGGAAC
DNA-dependent protein kinase	13570016	13509	GCATTTCCGGGTCCGGGCC GAGC	ATGGCGGGCTCCGGAGC CGGTGT	CCACAGACCTCACATCCAG GGCTCCCAT	CCAGAAGGGCATTCCCAAG GCTTCCCCAC
LDL-receptor related protein	34338	14896	TGAAGGAGAGAAGTAGCAG GACCAGA	ACCAGAGGGAAGGGGC TGCTGCTTG	TACTGTGCTCGCTTGCCG GCTTGCTC	GGAGAAATAATACTGTGC CTCGCTTGC

clones efficiently for large numbers of genes from a single cDNA source would contribute greatly to these research efforts. We describe here improvements for reverse transcription of mRNA to create a single cDNA pool that is optimized for the isolation of cDNAs for all gene templates and demonstrate the successful isolation of full-length transcripts of up to 15 kb.

MATERIALS AND METHODS

cDNA Synthesis

Eight separate conditions were used for reverse transcription (Table 1). In each, 5 µg mRNA template were used per tube in a final volume of 20 µL. All reactions followed standard reverse transcription procedures using manufacturer-supplied protocols and buffers, except for tubes 3 and 7, which included the following modifications: mRNA was denatured at 70°C in the presence of 40% DMSO for 10 min. cDNA synthesis was carried out with a final concentration of 10% DMSO using 200 U RNase H-deficient Moloney murine leukemia virus (MMLV) reverse transcriptase and 5 U *Pfu* DNA polymerase. After an initial incubation at 42°C for 20 min, the tubes were incubated at 58°C for 10 min to activate the *Pfu* DNA polymerase and to cause partial heat denaturation of RNA secondary structures (6). Two hundred units more of reverse transcriptase were then added, and the reactions were allowed to proceed at 42°C for an additional 20 min. To improve PCR yield, the reactions from all tubes were pooled and treated with 10 U RNase H (Promega, Madison, WI, USA) following reverse transcription (7).

PCR Amplification of cDNA

Primers were selected using Oligo 6.6 software (Molecular Biology Insights, Cascade, CO, USA) (Table 2). PCR was performed with a mixture of *Tth* and *Pfu*. All PCR reagents were purchased from

Short Technical Reports

Perkin-Elmer (Foster City, CA, USA), with the exception of *Pfu* DNA polymerase, which was purchased from Stratagene (La Jolla, CA, USA). The initial primary PCR mixture was constructed into two separate mixtures, a master mixture and an enzyme mixture. The master mixture contained 0.56 μ L dNTP (25 mM each), 2.94 μ L 25 mM magnesium acetate, 18.18 μ L 3.3 $\times$ *Tth* buffer, and 24.32 μ L nuclease-free water to a final volume of 46 μ L. The enzyme mixture contained 3 μ L 3.3 $\times$ *Tth* buffer, 1.75 μ L *Tth* DNA polymerase (2 U/ μ L), 0.28 μ L *Pfu* DNA polymerase (2.5 U/ μ L), and 4.97 μ L nuclease-free water to a final volume of 10 μ L. Oligonucleotides were purchased from Operon Technologies (Alameda, CA, USA), diluted to 5 μ M with TE, and 2 μ L each primer were used in PCR. Pooled cDNA (13 ng) in 10 μ L nuclease-free water was added into a 96-well PCR plate (MJ Research, Reno, NV, USA) with 46 μ L master mixture and 4 μ L primer. PCR was performed using a PTC-200™ peltier thermal cycler (MJ Research). Conditions were 82°C for 2 min (10 μ L enzyme mixture were added during this step) and 94°C for 3 min, followed by 11 cycles of 94°C for 50 s, 65°C for 1 min, 68°C 14 min, and 32 cycles of 94°C for 50 s, 65°C for 14 min, plus 15 s/cycle. This was concluded with an extension cycle of 68°C for 10 min. One microliter of the primary PCR product was used as template for secondary PCR using nested primers following the same conditions. Ten microliters of the nested PCR product were separated on a 0.8% SeaKem® GTG® agarose gel and stained with Gelstar (BMA, Rockland, ME, USA) (Figure 1). DNA in visualized bands was isolated with QIAquick® spin columns (Qiagen, Valencia, CA, USA) and eluted into 50 μ L 10 mM Tris (pH 8.0). Purified PCR products were cloned into pCR2.1-TOPO® (Invitrogen, Carlsbad, CA, USA), and insert ends were sequenced to confirm their identity.

RESULTS AND DISCUSSION

The development of long PCR technology using DNA polymerase mixtures has enabled the robust amplification of ultra-long targets from cloned

inserts and genomic DNA (8,9). However, RT-PCR on long cDNA templates has been difficult to achieve because of the inability to synthesize full-length cDNA efficiently. Multiple reverse transcription conditions often need to be tested before each long cDNA template can be successfully amplified. Presumably, the truncation of cDNA is the result of the reverse transcriptase's inherent RNA-degrading RNase H activity and the enzyme's lack of processivity in regions of RNA secondary structure. A reverse transcriptase lacking RNase H activity has been developed as well as enzymes capable of thermophilic reverse transcription (10–16). Reagents such as DMSO, betaine, formamide, glycerol, and trehalose, which facilitate the disruption of RNA secondary structure, have also been used to achieve full-length cDNA synthesis (17–19). Although these improvements have increased our ability to amplify some large transcripts, we found that neither the additives alone nor the use of the RNase H-minus reverse transcriptase can guarantee the synthesis of full-length cDNA for any given gene. Furthermore, for some genes the avian myeloblastosis virus (AMV) or the RNase H-plus wild-type MMLV outperforms RNase H-minus MMLV in full-length cDNA RT-PCR.

We have developed a method to synthesize full-length cDNA to enable the large-scale amplification of multiple long cDNA templates from a single optimized cDNA source. We test our cDNA by amplifying four long full-length mRNA transcripts ranging from 10.3 to 14.9 kb (Table 2) (20–23). We demonstrate successful RT-PCR of these long transcripts (Figure 1) without needing to optimize the cDNA synthesis reaction because, with our cDNA pool, at least one of the eight distinct conditions used to generate the cDNA pool is capable of full-length cDNA synthesis. For the four long mRNA transcripts tested, none could be amplified using cDNA prepared following standard conditions that did not include the modified reaction conditions used in tubes 3 and 7 (data not shown). However, even though cDNA from either tube 3 or 7 alone was sufficient for the full-length PCR amplification of these four selected mRNAs, the amplification

of other mRNAs may have required standard reverse transcription conditions. As our goal is to devise a strategy to generate a single cDNA source for the large-scale PCR of multiple mRNAs in high-throughput setting, reverse transcription products from the other standard conditions were also included in our final single cDNA pool for PCR. Sequence analysis of the first and last 600 bp of the clone insert ends confirmed that the PCR products were amplified from the intended mRNA transcripts. Although the entire clone insert was not sequenced, analysis of the insert ends indicated that the combined error rates for RT and PCR were low at 1/600 bp, with all changes being single-base substitutions. It is not known whether these changes were created during RT-PCR or if they were naturally occurring single nucleotide polymorphisms. Interestingly, with only a single exception, all changes detected led to silent mutations that do not alter the encoded amino acids. We believe that our addition of a proofreading function during the reverse transcription event significantly increased the full-length content of our cDNA. Incorporation of a wrong nucleotide by the reverse transcriptase at the 3' end of the growing cDNA would inhibit the synthesis of full-length cDNAs. Because of the lack of a proofreading 3'→5' exonuclease activity, error rates for all reverse transcriptases are high. An average error rate of 1 in 1000 bases has been shown (24,25). MMLV will only resume faithful incorporation 50% of the time after a mis-insertion (24). Analogous to the use of proofreading activities in DNA-dependent polymerase for long PCR, the proofreading activity can correct for errors made by the RNA-dependent polymerase during cDNA synthesis. This editing function is particularly useful for long cDNA synthesis, as base mis-incorporation frequency increases with length. While other reports have described methods to overcome premature termination of reverse transcription due to RNase H activity and mRNA secondary structure, none so far has described the use of a proofreading function for reverse transcription. We believe the method presented here will greatly improve a technique that is widely used by molecular biologists.

REFERENCES

1. Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
2. Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, et al. 2001. The sequence of the human genome. *Science* 291:1304-1351.
3. Strausberg, R.L., E.A. Feingold, R.D. Klausner, and F.S. Collins. 1999. The mammalian gene collection. *Science* 286:455-457.
4. Osato, N., M. Itoh, H. Konno, S. Kondo, K. Shibata, P. Carninci, T. Shiraki, A. Shinagawa, et al. 2002. A computer-based method of selecting clones for a full-length cDNA project: simultaneous collection of negligibly redundant and variant cDNAs. *Genome Res.* 12:1127-1134.
5. Kawai, J., A. Shinagawa, K. Shibata, M. Yoshino, M. Itoh, Y. Ishii, T. Arakawa, A. Hara, et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* 409:685-690.
6. Huttemann, M. 2002. Partial heat denaturation step during reverse transcription and PCR screening yields full-length 5'-cDNAs. *BioTechniques* 32:730-736.
7. Polumuri, S.K., A. Ruknudin, and D.H. Schulze. 2002. RNase H and its effects on PCR. *BioTechniques* 32:1224-1225.
8. Cheng, S., C. Fockler, W.M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* 91:5695-5699.
9. Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91:2216-2220.
10. Gerard, G.F., D.K. Fox, M. Nathan, and J.M. D'Alessio. 1997. Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. *Mol. Biotechnol.* 8:61-77.
11. Kotewicz, M.L., C.M. Sampson, J.M. D'Alessio, and G.F. Gerard. 1988. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res.* 16:265-277.
12. Myers, T.W. and D.H. Gelfand. 1991. Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry* 30:7661-7666.
13. Kotewicz, M.L. and G.F. Gerald. 2000. Cloned genes encoding reverse transcriptase lacking RNase H activity. *In* US patent and trademark office. Life Technologies, Inc., United States.
14. Gelfand, D.H., T.W. Myers, and C.L. Sigua. 1997. Reagents and methods for coupled high temperature reverse transcription and polymerase chain reactions. *In* US patent and trademark office. Roche Molecular Systems, Inc., United States.
15. Jendrisak, J.J. 2000. Reverse transcription method. *In* US patent and trademark office. Epicentre Technologies Corporation, United States.
16. Warthoe, P. 2001. Method for improved reverse transcription at high temperatures. *In* US patent and trademark office. Display Systems Biotech, United States.
17. Carninci, P., Y. Nishiyama, A. Westover, M. Itoh, S. Nagaoka, N. Sasaki, Y. Okazaki, M. Muramatsu, et al. 1998. Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full-length cDNA. *Proc. Natl. Acad. Sci. USA* 95:520-524.
18. Chester, N. and D.R. Marshak. 1993. Dimethyl sulfoxide-mediated primer T_m reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal. Biochem.* 209:284-290.
19. Spiess, A.N. and R. Ivell. 2002. A highly efficient method for long-chain cDNA synthesis using trehalose and betaine. *Anal. Biochem.* 301:168-174.
20. DeSilva, U., G. D'Arcangelo, V.V. Braden, J. Chen, G.G. Miao, T. Curran, and E.D. Green. 1997. The human reelin gene: isolation, sequencing, and mapping on chromosome 7. *Genome Res.* 7:157-164.
21. Herz, J., U. Hamann, S. Rogné, O. Myklebost, H. Gausepohl, and K.K. Stanley. 1988. Surface location and high affinity for calcium of a 500-kd liver membrane protein closely related to the LDL-receptor suggest a physiological role as lipoprotein receptor. *EMBO J.* 7:4119-4127.
22. Tinsley, J.M., D.J. Blake, A. Roche, U. Fairbrother, J. Riss, B.C. Byth, A.E. Knight, J. Kendrick-Jones, et al. 1992. Primary structure of dystrophin-related protein. *Nature* 360:591-593.
23. Hartley, K.O., D. Gell, G.C. Smith, H. Zhang, N. Divecha, M.A. Connelly, A. Admon, S.P. Lees-Miller, et al. 1995. DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* 82:849-856.
24. Krug, M.S. and S.L. Berger. 1987. First-strand cDNA synthesis primed with oligo(dT). *Methods Enzymol.* 152:316-325.
25. Schaefer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* 227:255-273.

Address correspondence to Dr. Glenn K. Fu, Director, Molecular Biology, Incyte Genomics, 3160 Porter Drive, Palo Alto, CA 94304, USA. e-mail: gfu@incyte.com

Received 11 October 2002; accepted 30 January 2003.

**Phillip R. Hawkins, Pei Jin,
and Glenn K. Fu**
*Incyte Genomics
Palo Alto, CA, USA*