Reverse Transcriptase Adds Nontemplated Nucleotides to cDNAs During 5'-RACE and Primer Extension

Dayue Chen and John T. Patton
National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

INTRODUCTION

Both rapid amplification of cDNA ends (RACE) and primer extension assays use reverse transcriptase to catalyze the template-directed synthesis of a cDNA. The RACE procedure is frequently used to generate full-length cDNA clones and to determine the precise terminal sequences of cellular and viral RNAs (5–7). In the initial step of RACE, a cDNA is produced by incubating an RNA with reverse transcriptase and one of two types of primers, either oligo-dT (3'-RACE) or a specific deoxyoligonucleotide that anneals closely to the 5' end of the RNA (5'-RACE). In the case of 5'-RACE, a homopolymer tail is then added to the 5' end of the cDNA product with terminal deoxynucleotidyl transferase (TdT). Afterwards, the cDNAs are amplified by PCR using a gene-specific primer and an oligonucleotide complementary to the homopolymer tail. The amplified cDNAs can be sequenced directly or cloned into a plasmid for further characterization. In a typical primer extension assay, a primer anneals to the RNA of interest through specific base pairing and is extended by reverse transcriptase until the growing chain of cDNA reaches the 5' end of the RNA template. The technique has long been used to measure the level of gene expression and to map the 5' end of RNAs (8,9).

In this study, we report that commercially purchased Moloney murine leukemia virus (MMLV) and avian myeloblastosis virus reverse transcriptases contained the TdT-like activity. This work implies that 5'-RACE and primer extension assays must be used carefully in determining the terminal sequences of nucleic acids because, under standard reaction conditions, reverse transcriptase can add nontemplated nucleotides to the 3' ends of cDNAs following template-directed synthesis. Under typical reaction conditions, the termini of 70%–80% of 5'-RACE cDNAs prepared with MMLV reverse transcriptase contained one or more nontemplated nucleotides. The TdT-like activity preferentially adds A residues to cDNAs and uses DNA/RNA or DNA/DNA duplexes as substrates. Because of this activity, strict reliance on 5'-RACE and primer extension to define the 5'-terminal sequence of RNA can yield misleading results.

MATERIALS AND METHODS

Cloning and Sequencing of 5'-RACE Products

dsRNA was purified from rotavirus SA11-4F (2), heated to 100°C for 2 min, and used as a template for making gene 8 cDNAs. The primer for cDNA synthesis, 5'-ATTAGAGACGTGCTTGCGTAC-3', was complementary to nucleotides 322–345 of the gene 8 (+)RNA. Reaction mixtures (50 μL) contained 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM MgCl₂, 10 mM dithiothreitol, 400 μM each of the dNTPs, 0.5 μg dsRNA, 100 nM primer, and 200 U MMLV reverse transcriptase (SUPERSCRIPT II™; Life Technologies, Rockville, MD, USA) and were incubated at 45°C for 50 min. The cDNA products were purified with GlassMax® spin cartridges (Life Technologies) and then subjected to incubation with 10 U TdT (Life Technologies) and 200 μM dGTP, dATP, or dCTP in a volume of 25 μL for 10 min at 37°C. The cDNAs were amplified by 30 cycles (94°C for 60 s, 50°C for 45 s, 72°C for 90 s) of PCR using an anchor primer and the gene 8 primer, 5'-aatgtgactAGAGCTTCATA-3'.
TCAACATAGCC-3', which contained a sequence complementary to nucleotides 285–308 of the gene 8 (+)RNA (upper case) and a BamHI site (underlined). The oligonucleotides 5'-ccggactgaggagggaaagggg-3' (EcoRI site underlined), 5'-ggcaagcgagcatgacgccggg-3' (SalI site underlined, I = inosine), and 5'-cggagcagttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
quence of the gene 8 (+)RNA of ro-
tavirus SA11-4F by 5¢-RACE (4),
cDNAs were prepared from virion-de-
rived gene 8 dsRNA using MMLV re-
verse transcriptase. The cDNAs were
tailed with dATP, dCTP, or dGTP, am-
plified by PCR, and cloned into pUC19.
Of 29 dATP-tailed clones, sequence
analysis showed that 22 (75.9%) had the
5¢-terminal sequence, 5¢-GGCTTTT-3¢
(Figure 2A). Based on earlier work
(1,12), this is the expected sequence for
the 5¢ end of the gene 8 RNA. However,
seven clones (24.1%) contained either
one (G or C) or two (GT, GC, or CT) ex-
tra nucleotides upstream of the expected
5¢-terminal sequence (Figure 2A). Of 34
dCTP-tailed clones, the 5¢-terminal se-
quence of only 10 (24.9%) agreed with
expected gene 8 sequence (Figure 2B).
For these 10 clones, the origin of the
first two nucleotides GG in the 5¢-termi-
nal sequence (5¢-GGCTTTT-3¢) is
masked because of dCTP-tailing of the
first-strand cDNA. Of the remaining 24
dCTP-tailed clones, 18 (52.9%) con-
tained an extra T, and 6 (17.6%) con-
tained an extra C upstream of the ex-
pected 5¢-terminal sequence (Figure
2B). Similar to the dCTP-tailed clones,
the majority (19 of 22 clones, 84.6%) of
the dGTP-tailed clones contained one
or two extra nucleotides at their 5¢ ends
(Figure 2C). Of the 19 clones, 17
(77.3%) contained an extra T, one con-
tained an extra GT, and one contained
an extra TT. These results showed that
the 5¢-terminal sequences of the gene 8
cDNA clones produced by 5¢-RACE were
heterogeneous and that, in total, the
majority of the clones contained one
or more additional nucleotides upstream
of the expected 5¢-terminal sequence.
The extra nucleotides were detected
more often with cDNAs tailed with
dCSP or dGTP than those tailed with
dATP. For cDNAs tailed with dCSP or
dGTP, the extra nucleotides upstream of
the expected 5¢-terminal gene 8 se-
quence were usually T residues.

When gene 8 (+)RNA made by T7
transcription was used as the template
for 5¢-RACE, instead of virion-derived
RNA, the 5¢ ends of the gene 8 cDNA
clones contained extra nucleotides in
frequencies similar to those observed
above (data not shown). Thus, even
when the RNA used for 5¢-RACE were
made by transcribing a gene 8 cDNA
beginning with the sequence 5¢-
GGCTTTT-3¢, analysis of the clones
indicated that the 5¢ end of the RNA
was heterogeneous in sequence and
contained extra nucleotides upstream
of the expected 5¢-terminal sequence.

Addition of Nontemplated
Nucleotides to DNA/RNA Duplexes
by Reverse Transcriptase

Previous reports have indicated that
the reverse transcriptases of AMV,
MMLV, and HIV have an associated
TdT-like activity (3,10,11,15). To test
whether or not this activity was respon-
sible for the presence of the extra nu-
cleotides, oligonucleotides I (18 nu-
cleotides), II (24 nucleotides), and III
(25 nucleotides) were 5¢-end-labeled
with 32P and then annealed to gene 8
(+)-RNAs, yielding DNA/RNA duplex-
es that contained a 5¢-RNA overhang
(I), a blunt-end (II), and a single-base
3¢-DNA overhang (III), respectively
(Figure 1). The terminal structure of
duplex II mimics the blunt-end generat-
ed by reverse transcriptase during first-
strand synthesis on an RNA template.
The three DNA/RNA duplexes were

Figure 2. Sequence analysis of gene 8 5¢-RACE clones. 5¢-RACE was used to prepare cDNAs of the 5¢ end of rotavirus gene 8 (+)-RNAs. The cDNAs were tailed with dATP (A), dCSP (B), or dGTP (C), am-
plified by PCR, and ligated into pUC19. Dideoxynucleotide sequencing was used to analyze the junction
region between the homopolymer tail and the 5¢-end of the gene 8-specific sequence of these clones. Un-
expected nucleotides found upstream of the gene 8-specific sequence are indicated with asterisks. The
5¢-terminal sequence of the gene 8 (+)RNA is shown.
incubated with MMLV or AMV reverse transcriptase, at concentrations suggested by the supplier, in the presence of all four dNTPs or one individual dNTP. The sizes of the $^{32}$P-labeled oligonucleotides of the DNA/RNA duplexes were evaluated by electrophoresis (Figure 3A). When DNA/RNA duplex II was incubated with all four dNTPs and MMLV or AMV reverse transcriptase, 90% and 30%, respectively, of the oligonucleotides recovered from the reaction mixtures had increased in size by one or more nucleotides (lanes 3 and 11). When the four dNTPs were incubated individually with MMLV and AMV reverse transcriptase, in all cases the products included oligonucleotides that had increased in size by at least one nucleotide (lanes 4–7 and 12–15). However, the efficiency with which the non-templated nucleotides were added to the oligonucleotides varied depending on which dNTP was present (lanes 4–7 and 12–15) and was calculated to follow the order A>G>C>T. The addition of two or more non-templated nucleotides to oligonucleotides occurred much less frequently than did the addition of a single nucleotide, and the addition of two or more non-templated nucleotides was most efficient in reaction mixtures that contained both dATP and MMLV reverse transcriptase. Because oligonucleotides that contained a single non-templated nucleotide (i.e., the 25-nucleotide products) exhibited slight differences in electrophoretic mobility depending on which nucleotide was added (lanes 4–7 and 12–15), it was possible to conclude that in the presence of all four dNTPs, dATP was added to oligonucleotides more efficiently than the other three dNTPs (lanes 3 vs. 5 and lanes 11 vs. 13).

Due to the six nucleotide 5'-RNA overhang, incubation of DNA/RNA duplex I with reverse transcriptase is expected to extend the oligonucleotide from 18 to 24 nucleotides via template-dependent cDNA synthesis (Figure 1). If reverse transcriptase transfers nucleotides to the cDNA in a non-template-dependent manner upon the completion of template-dependent synthesis, then the length of the oligonucleotide will exceed 24 nucleotides. Indeed, such experiments revealed that greater than 90% of the total extended oligonucleotides recovered from reaction mixtures containing MMLV reverse transcriptase were 25 nucleotides or more in size (Figure 3A, lanes 8 and 9). In contrast, when AMV reverse transcriptase was used, approximately 25% of the total extended oligonucleotides were 25 nucleotides in length (lanes 16 and 17). The results obtained with DNA/RNA duplexes I and II indicate that both these commercially purchased reverse transcriptases have TdT-like
activity, which adds nontemplated nucleotides to the 3’-end of cDNAs. Unlike duplexes I and II, MMLV and AMV reverse transcriptase did not add nontemplated nucleotides to oligonucleotides I and II (Figure 3B). Thus, the substrate for the TdT-like activity of reverse transcriptase does not include ssDNA. In related experiments, it was shown that the TdT-like activity of reverse transcriptase could add nontemplated nucleotides to blunt-ended DNA/DNA duplexes but could not add nucleotides to ssRNA (data not shown).

The addition of the first nontemplated nucleotide to the oligonucleotides of duplexes I (5’-RNA overhang) and II (blunt-end) occurred more frequently in reaction mixtures containing reverse transcriptase and all four dNTPs than did the addition of the second or third nontemplated nucleotide (Figure 3A). This suggested that the transfer of one nontemplated nucleotide produced 3’ overhangs on these duplexes, which then made them poorer substrates for the TdT-like activity of reverse transcriptase. To test this, we compared the ability of MMLV and AMV reverse transcriptase to transfer nontemplated nucleotides to DNA/RNA duplexes that contained a 5’-RNA overhang (I), a blunt-end (II), and a single-base 3’-DNA overhang (III). The results showed that MMLV and AMV reverse transcriptase added nontemplated nucleotides to 91% and 25%, respectively, of the extended oligonucleotides of duplex I (Figure 4, lanes 5 and 8), to 87% and 26% of the oligonucleotides of duplex II (lanes 6 and 9), but to only 27% and 0% of the oligonucleotides of duplex III (lanes 7–10). The decreased frequency of the transfer of nontemplated nucleotides to the oligonucleotides of DNA/RNA duplex III relative to those of duplexes I and II showed that the TdT-like activity of reverse transcriptase adds nucleotides to blunt-ended substrates more efficiently than to substrates with 3’ overhangs.

**Effect of Reverse Transcriptase Concentration and Temperature on the Addition of Nontemplated Nucleotides**

To evaluate the effect of reverse transcriptase concentration on the addition of nontemplated nucleotides, DNA/RNA duplex I (5’-RNA overhang) was incubated with serial dilutions of MMLV and AMV reverse transcriptase. Analysis of the extended oligonucleotides (i.e., those ≥ 24 nucleotides) from the assays showed that the TdT-like activity decreased by eightfold as the amount of MMLV reverse transcriptase was reduced from 100 to 12.5 U (Figure 5A, lanes 6–9). Further reduction in the amount of MMLV reverse transcriptase had little effect on TdT-like activity (lanes 10–12). In contrast, dilution of AMV reverse transcriptase had no impact on its TdT-like activity (lanes 13–15). Higher concentrations of AMV reverse transcriptase were not available for testing. These results indicate that reverse transcriptase concentration affects the efficiency of nontemplated nucleotide addition.
transcriptase concentration affects the frequency of the addition of nontemplated nucleotides and that the level of TdT-like activity is lower at lower reverse transcriptase concentrations.

Reverse transcription is routinely performed at elevated temperature (45°C–50°C) to reduce the impact of RNA secondary structure on cDNA synthesis. To examine the effect of temperature on the TdT-like activity of reverse transcriptase, DNA/RNA duplex I (5'-RNA overhang) was incubated with MMLV and AMV reverse transcriptases at 25°C–50°C. The results showed that temperature was not a significant factor in the template-dependent elongation (i.e., cDNA synthesis) of the 18-nucleotide oligonucleotide primer of the duplex (Figure 5B, lanes 5–14). However, the addition of non-templated nucleotides was affected by temperature. At 25°C, approximately 36% and 48% of the extended oligonucleotides generated by MMLV and AMV reverse transcriptase, respectively, contained nontemplated nucleotides (lanes 5 and 10). At 32°C–45°C, the level of extended oligonucleotides with nontemplated nucleotides increased to 90% in assays containing MMLV reverse transcriptase (lanes 6–8). Unlike the results with MMLV reverse transcriptase, the level of extended oligonucleotides with nontemplated nucleotides decreased to 20% in assays containing AMV reverse transcriptase at 32°C–45°C (lanes 11–13). At the highest temperature (50°C), nontemplated nucleotides were present on only approximately 27% and 19% of the extended products recovered from assays containing MMLV and AMV reverse transcriptase, respectively (lanes 9 and 14). Therefore, although functioning over a wide range of temperatures, the TdT-like activity of reverse transcriptase is influenced by temperature and is reduced at 50°C.

**DISCUSSION**

This study was prompted by our initial observation that the majority of clones generated by the 5'-RACE procedure contained extra nucleotides upstream of the 5' end of the rotavirus gene 8-specific sequence. By analyzing DNA/RNA duplexes with defined terminal sequences, we have demonstrated that reverse transcriptase can add one, and less commonly two or three, nucleotides in a template-independent manner to the 3' end of cDNAs upon the completion of template-dependent synthesis. The TdT-like activity of reverse transcriptase is substrate specific, as it catalyzed the addition of non-templated nucleotides to DNA/RNA and DNA/DNA duplexes, but not to ssDNA or ssRNA. In comparison to DNA/RNA duplexes with blunt-ends, the TdT-like activity...
activity added nucleotides to duplexes with 3'-DNA overhangs much less efficiently. While any of the four dNTPs were used as substrates, dATP was the preferred substrate for TdT-like activity. Reverse transcriptase added nontemplated nucleotides to cDNAs at all enzyme concentrations examined, but only at high concentrations of reverse transcriptase (≥ 100 U) was the addition of more than one nontemplated nucleotide to cDNAs readily detected. The TdT-like activity of reverse transcriptase functioned over a wide range of temperatures (25°C–50°C) but exhibited peak activity at temperatures that are typically used in generating cDNAs by RACE and primer extension (32°C–45°C).

5'-RACE and primer extension have been widely used to determine the precise terminal sequences of cellular and viral RNAs. Under typical reaction conditions, MMLV reverse transcriptase can add nontemplated nucleotides to 70%–80% of 5'-RACE clones (Figure 1) and to greater than 90% of primer extension products (Figure 4), while AMV reverse transcriptase adds nontemplated nucleotides to 40% of 5'-RACE clones (data not shown) and to 25% of primer extension products (Figure 4). These findings raise concerns about the strict reliance of using these methods to determine the terminal sequences of RNAs. For cellular RNA, the presence of nontemplated nucleotides may be readily identified by comparing the terminal sequence of the cDNA to the genomic DNA sequence from which the RNA was derived. However, this approach is not applicable for cDNAs of RNA viruses because of the lack of a genomic DNA sequence to make a sequence comparison. Indeed, when using 5'-RACE or primer extension to determine the terminal sequences of RNAs of RNA viruses, it may be necessary to confirm its accuracy by an alternative method such as direct RNA sequencing.

Despite significant effort, we could not find assay conditions under which the TdT-like activity of reverse transcriptase was abolished. However, our analysis indicates that it is possible to reduce the incidence of the addition of nontemplated nucleotides in 5'-RACE and primer extension by using less reverse transcriptase, by raising the reaction temperature, or by a combination of both. Since the addition of nontemplated nucleotides cannot be completely eliminated, 5'-RACE clones can be expected to exhibit some degree of sequence heterogeneity at the junction of cDNAs and homopolymer tail. To avoid misinterpreting the sequence information gained from 5'-RACE clones, we recommend that at least two different dNTPs should be used in homopolymer tailing and that several cDNA clones should be independently prepared and their sequences comparatively analyzed.

REFERENCES


Received 19 June 2000; accepted 31 August 2000.

Address correspondence to:
Dr. John T. Patton
Laboratory of Infectious Diseases
National Institutes of Allergy and Infectious Diseases
National Institutes of Health
7 Center Drive, MSC 0720, Room 117
Bethesda, MD 20892, USA.
e-mail: jpatton@niaid.nih.gov