METHOD

A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase

C. KAO,¹ M. ZHENG,² and S. RÜDISSE²

¹Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

²Department of Chemistry, University of California at Berkeley, and Physical Bioscience Division, Berkeley Lawrence National Laboratory, Berkeley, California, USA

ABSTRACT

DNA templates modified with C2'-methoxyls at the last two nucleotides of the 5' termini dramatically reduced nontemplated nucleotide addition by the T7 RNA polymerase from both single- and double-stranded DNA templates. This strategy was used to generate several different transcripts. Two of the transcripts were demonstrated by nuclear magnetic resonance spectroscopy to be unaffected in their sequence. Transcripts produced from the modified templates can be purified with greater ease and should be useful in a number of applications.

Keywords: nontemplated nucleotide addition; nucleotide analog; T7 RNA polymerase; transcription

INTRODUCTION

Transcription by T7 RNA polymerase *in vitro* is currently the most widely used method to produce RNAs for a wide variety of applications, including structural and biochemical studies and potential therapeutics (Milligan et al., 1987; Turek & Gold, 1990; Symensma et al., 1996; Siegel et al., 1997). Despite its usefulness, a number of undesired reactions increase the complexity of the T7 polymerase products and necessitate the careful purification of the desired RNAs. These reactions include the synthesis of oligonucleotides aborted during the initiation of transcription (Martin et al., 1988; Moroney & Piccirilli, 1991), polymerase slippage (MacDonald et al., 1993), the use of alternative template initiation sites (Pleiss et al., 1998; Helm et al., 1999), and the addition of one or more nontemplated nucleotides at the 3' terminus of the nascent RNA (henceforth called the N+1 activity). N+1 activity can be a major contributing factor to heterogeneity in transcription products (Milligan et al., 1987; Krupp, 1988). Previously, Moran et al. (1996) demonstrated that DNA templates

containing nucleoside with base analogs that cannot form hydrogen bonds with the substrate nucleotide can induce the termination of transcription prior to the nucleoside analog. However, these analogs are not widely available and a significant amount of N+1 activity remained with some templates. We report that modification of either the penultimate nucleotide or the last two nucleotides at the 5' terminus of the DNA template with methoxy moieties at the ribose C2' position can significantly reduce N+1 activity by the T7 RNA polymerase and, in several cases, increase the abundance of the desired RNA.

RESULTS AND DISCUSSION

Transcription from DNA template containing ribose 2'-methoxy analogs

In studies of RNA synthesis by a viral RNA-dependent RNA polymerase, we noticed that nontemplated nucleotide addition on the nascent RNA was decreased when the 5' terminus of the template was modified with deoxyribonucleotides (Siegel et al., 1999). This observation prompted us to determine whether a DNA template containing nucleotide analogs at the 5' terminus would similarly affect RNA synthesis by the T7 RNA polymerase.

Reprint requests to: C. Cheng Kao, Department of Biology, 1001 East 3rd St., Indiana University, Bloomington, Indiana 47405-3700, USA; e-mail: ckao@bio.indiana.edu.

ase. We tested several nucleotide analogs and found that modification of the ribose at the C2' position with a methoxy moiety was the most effective. Nucleotides with ribose 2'-methoxyls are a good choice for further characterization for two reasons: (1) this analog is commercially available and inexpensive to incorporate during the chemical synthesis of DNA; (2) because the base is unchanged, the identity of the nucleotide incorporated into the nascent RNA should be unaffected. As an initial test, we used a DNA named S5 that should encode a 22-nt RNA within P5 of the *Tetrahymena thermophila* group I intron (Fig. 1A). (For nomenclature, DNA templates with modified 5' terminal nucleotide(s) will be denoted with one or two asterisks following the name shared with the unmodified template.) In a typical transcription reaction as performed by the protocol of Milligan et al. (1987), the DNA template S5 directed production of a 23-nt N+1 RNA at amounts between 50 and 200% of the amount of the desired 22-nt RNA (Fig. 1B, lanes 1 and 2). When S5** was used in the transcription reaction, we noted that the amount of N+1 product made was not only significantly reduced, but the 22-nt transcript was reproducibly more abundant than that from S5 (Fig. 1B, lanes 3 and 4). In more careful analysis, the 22-nt transcript from S5** was reproducibly increased two- to threefold compared to amounts made from S5.

The above observation prompted us to investigate whether either or both modified guanylates in S5** were required to reduce the N+1 activity. DNA templates that contained one guanosine ribose 2'-methoxy moiety at the 5' terminus (S5^{G*}) or at the penultimate position (S5^{G*}) were made, and their transcription products compared with those from S5 and S5**. Template S5** and S5^{G*} significantly reduced the N+1 transcript whereas S5^{G*} did not (Fig. 1B, lanes 7 and 8). Therefore, modification of the penultimate position is the most important in reducing N+1 activity. In addition to the N+1 product, a minute amount of N+2 RNA was observed in overloaded gels. The transcripts from S5** had a reduction of the N+2 RNA. Therefore we routinely used DNA templates modified in the two terminal nucleotides.

To determine whether the modified template would affect the identity of the nucleotide incorporated into the RNA, we purified the 22-nt RNA generated from a 30-mL transcription using template S5 and a 10-mL transcription that used S5** and compared their nuclear magnetic resonance (NMR) spectra. Since the nucleotides of interest are at the 3' terminus of the respective RNAs, and the imino protons of the corresponding base pair cannot be observed because of fast exchange with the solvent, we examined the non-exchangeable ribose H1' and H5 protons, whose chemical shifts are affected by the identity of the base (Fig. 1C). Previous work (S. Rüdisser, unpubl. results) had identified that the terminal cytidylate in the RNA

FIGURE 1. DNAs containing 5' guanosine nucleotides modified with ribose 2'-methoxyls reduced the N+1 activity of the T7 RNA polymerase. **A:** Schematic of the DNA templates S5 (unmodified) and S5** (modified with two guanylates containing ribose 2'-methoxyls) used for transcription. Gm denotes the position of modified guanylates. The sequence in lower case letters is the bottom strand of the T7 promoter, and the sequence in capital letters is the template for transcription. The arrow denotes the position for the initiation of transcription. The dotted line indicates that the sequence of S5** that is not shown is identical to that of S5. The gray lines denote the positions in S5 and S5** most relevant to this analysis. **B:** Transcripts generated from various DNA templates. An asterisk denotes the presence of a methoxy-modified guanylate. The RNAs were visualized after staining with Toluidine Blue and a scan of the gel is shown. Lanes 1–4 contain the products of independent transcription reactions. The 22-nt product and the result of the N+1 activity are indicated to the right of the gel image. **C:** NMR spectroscopy of the RNA produced by DNA template S5 (upper spectrum) and S5** (lower spectrum). The signals from 5 to 6.2 ppm containing the H1'/H5 region of the RNA are shown.

has an H1' peak at 5.75 ppm and a H5 peak at 5.6 ppm. We observed these peaks at the same chemical shifts when using RNA transcribed from either S5** or S5. In addition, all other peaks from the two RNAs corresponded exactly. Finally, the H6 proton of cytidine

22 was found at 7.66 ppm in both samples, further confirming that modifications in S5** did not alter the sequence of the transcription product.

Four additional DNA templates that should direct the synthesis of RNAs from 27 to 64 nt were examined. The modified DNAs were named P1, V Ψ , M5-54, and M5-64, and their 5' terminal nucleotides were, respectively, 5'GA3', 5'AT3', 5'GG3', and 5'GG3'. The transcription products from each pair of unmodified and modified DNAs are shown in Figure 2A. The unmodified versions of all four DNAs generated significant amounts of the respective N+1 products (Fig. 2A, lanes 1, 4, 5, and 7) whereas the modified DNAs did not (Fig. 2, lanes 2, 3, 6, and 8). We have now produced four additional RNAs from modified templates that had significantly reduced N+1 RNAs. Transcription reactions using P1 and V Ψ generated RNAs that were either 1 or 2 nt less than full-length. The modified DNA templates did not significantly affect the abundance of these truncated RNAs (Fig. 2A, lanes 1–4).

Transcripts longer than 50 nt are difficult to purify away from their associated N+1 RNAs even after extensive preparative gel electrophoresis. This has led to

contamination of the desired transcripts that can complicate spectroscopic and biochemical analyses of RNAs. For example, Wu and Tinoco (1998) observed in a 56-nt RNA three additional imino signals associated with the last three base pairs of the N+1 RNA. In one-dimensional imino spectrum of the 56-nt RNA (Fig. 2B, bottom spectrum), the peak at 12.58 ppm (labeled G191⁺) and the shoulder at 13.2 ppm (labeled G131⁺) were tentatively assigned to the N+1 RNA because it was not possible to separate the N+1 RNA from the 56-nt RNA. To examine whether transcripts purified from the modified template could remove the confounding signals, we purified the transcripts from a modified template encoding the 56-nt sequence used by Wu and Tinoco and performed one-dimensional imino spectra analysis (Fig. 2B, top spectrum). The transcript from the modified template, M5-56**, lacked the G191⁺ signal, was significantly enhanced in the signal for G131 and G191, and did not have the extra peak shoulder associated with G131⁺. All other peaks were unaffected, demonstrating that the nucleotide sequence produced from modified template M5-56 was identical to the expected sequence.

FIGURE 2. Effect of ribose 2'-methoxy modifications at the last two nucleotides at the 5' termini of four DNA templates on N+1 activity of T7 RNA polymerase. **A:** RNAs in denaturing polyacrylamide gels stained with Toluidine Blue. Both gels contain 20% polyacrylamide, but the samples in lanes 1–4 and lanes 5–6 were electrophoresed for 16 and 24 h, respectively. The DNA templates used to generate the transcripts are listed on top of the gel images, with two asterisks denoting modification of the 5' terminal two nucleotides with ribose 2'-methoxyl moieties. The lengths of the transcripts expected for P1 and V Ψ are indicated on the left of the gel image whereas transcripts from M5-56 and M5-64 are shown on the right side. The positions of the RNA 1 nt longer than the expected size are indicated with "N+1". **B:** NMR spectra of the imino protons of the RNAs transcribed from M5-56 and M5-56**. The positions of G131 and G191 are identified within the spectra as assigned by Wu and Tinoco (1998). G131⁺ and G191⁺ are the peaks associated with an RNA that contains the N+1 nucleotide.

Transcription from double-stranded templates

RNAs of less than a 100 nt can be transcribed from chemically synthesized DNA templates where only the promoter sequence is double-stranded (Milligan et al., 1987). However, chemical DNA synthesis is impractical for transcription of longer RNAs, and the double-stranded templates generated by polymerase chain reaction (PCR) are more suitable (Ausubel et al., 1994). To determine whether the ribose 2'-methoxy modifications could potentially affect transcription from double-stranded DNA templates, we used Taq DNA polymerase and dNTPs to extend the oligonucleotide containing the T7 promoter sequence hybridized to DNA S5 and V Ψ , or to S5** and V Ψ **. Half of the extension products were analyzed by gel electrophoresis to confirm that duplex DNAs were formed. Transcription from the remaining half of the extension reaction yielded the products shown in Figure 3. Double-stranded S5** and V Ψ ** again directed products that lacked the N+1 RNAs whereas unmodified double-stranded S5 and V Ψ each produced significant amounts of N+1 products. In PCR products, a DNA primer generates the 5' terminus of the template strand for transcription. The above result with double-stranded templates suggests that DNAs amplified by PCR using modified primers will also reduce N+1 activity.

The heterogeneity at the 3' ends of transcripts generated by polymerases in vitro can have detrimental effects in a number of applications. In the production of infectious viral transcripts, 3' heterogeneity may reduce the synthesis by viral RNA replicase (Miller et al.,

1986; Sun et al., 1996), or reduce the infectivity of the RNA (Murphy & Park, 1997). In biochemical and structural analyses of RNA structures, 3' heterogeneity may result in additional signals that complicate the interpretation of results and may prevent the effective crystallization of the RNA needed for X-ray crystallography (Price et al., 1995). Previously, one commonly used strategy to generate a homogeneous RNA sample of less than 50 nt is to perform denaturing gel electrophoresis and to excise and elute the desired band (Wyatt et al., 1991). For RNAs longer than 50 nt, where it is difficult to separate the N+1 RNA from the desired RNA by gel electrophoresis, the transcript can be expressed with a *cis*-acting ribozyme sequence that will cleave at the desired positions (Dianzott & Bujarski, 1988; Price et al., 1995), or by use of a mutant T7 RNA polymerase that is defective in its initiation rate (Gardner et al., 1997). Both procedures require extensive effort that may be partially alleviated by the use of chemically synthesized DNA templates or PCR-generated double-stranded templates containing ribose 2'-methoxy moieties. The quality and, often, the yield of the desired transcripts will be improved by this method. The ability to generate less complex transcription products will permit more rapid RNA purification such as by high-pressure liquid chromatography.

MATERIALS AND METHODS

Molecular manipulations

The majority of the oligonucleotides used in this study were synthesized using an Applied Biosystems DNA synthesizer (model 381A) with reagents purchased from Glen Research. Some oligonucleotides were purchased from Operon Technologies (Alameda, California). Transcriptions used the conditions described in Wyatt et al. (1991). Briefly, the template DNA strand and the T7 promoter DNA strand were purified via denaturing polyacrylamide gel electrophoresis and then adjusted to 8 μ M. One microliter of each DNA was used in a 20- μ L transcription reaction containing a final concentration of 40 mM Tris, pH 8.1, 1 mM spermidine, 0.01% Triton X-100, 80 mg/reaction of polyethylene glycol 8000, and 4 mM of each NTP. The T7 RNA polymerase used was purified using the protocol of Grodberg and Dunn (1988). The sequence of the RNA transcribed from S5 is: 5'-GGCGAAGUCGAAAGAUGGCGCC-3'. The sequence of the RNA transcribed from P1 is: 5'-GAUACCUUUGGAGGGCUUCGGCUCUCU-3'. The sequence of the RNA transcribed from V Ψ is: 5'-GGCGCAGUGGGCUAGCGCCACUCAAAAGGCCCUA-3'. The sequence of the RNA transcribed from M5-56 is: 5'-GGCAGUACCAAGUCGCGAAAGCGAUGGCUUUGCAAAGGGGUAUGGUAUAAGCUGCC-3'. The sequence of the RNA transcribed from P5-64 is: 5'-GGCAGUACCAAGUCGCGAAAGCGAUGGCUUUGCAAAGGGGUAUGGUAUAAGCUGCC-3'.

Double-stranded DNA templates were synthesized using Taq DNA polymerase in the buffer described in Ausubel et al. (1994). The reaction was performed in two cycles as follows: the reaction was heated to 90 °C and cooled to 45 °C to allow

FIGURE 3. Ribose 2'-methoxy modification present in double-stranded templates can decrease N+1 activity by the T7 RNA polymerase. The double-stranded templates used for transcription are shown at the top of the Toluidine Blue-stained gel. The lengths of the expected transcript and the N+1 products are shown to the right of the gel image.

annealing of the oligonucleotides. Taq polymerase was then added as the reaction was incubated for 30 min at 70°C. After the extension, the reaction was extracted with a 1:1 mixture of phenol:chloroform and then precipitated with 3 vol of ethanol, washed with 70% ethanol, and dried prior to transcription. The RNAs were visualized following staining with Toluidine Blue as described by the manufacturer (Boehringer Mannheim Inc.).

Nuclear magnetic resonance spectroscopy

RNA transcripts used for NMR samples were prepared as described by Wyatt et al. (1991). NMR spectroscopy was performed on a Bruker AMX-600 or a Bruker DRX-500 (Billerica, Massachusetts) at field strength of 600.14 MHz and 500.13 MHz, respectively. All samples were dissolved in a phosphate buffer containing 90% H₂O and 10% D₂O. Spectra for the RNAs generated from S5 (2.9 mM sample) and S5** (2.7 mM sample) were taken at 30°C with 16,384 complex points in 64 scans. The proton carrier frequency was set at 4.72 ppm and the maximum for excitation was set at 7.7 ppm. For the spectra of RNAs transcribed from M5-56 and M5-56** (each at 1 mM), the spectra were recorded at 10°C using a 1-1 pulse sequence for water suppression (Plateau & Gueron, 1982; Hore, 1983). The carrier frequency was centered on the water resonance frequency and the excitation profile was maximized at 12 ppm.

ACKNOWLEDGMENTS

This work would not have been possible without the guidance and generous support of Professor Ignacio Tinoco Jr., in whose laboratory these experiments were performed. We thank David Koh for synthesizing oligonucleotides and Laura Kao for editing of the manuscript. Funding was provided in part by a National Institutes of Health grant GM10840 and Department of Energy grant DE-FG03-86ER60406 to I.T. and by the National Science Foundation grant MCB9507344 and a U.S. Department of Agriculture grant 9702126 to C.K. S.R. is grateful for postdoctoral fellowship J1499-PHY from the Fonds zur Förderung der Wissenschaftlichen Forschung of Austria, and M.Z. gratefully acknowledges a postdoctoral fellowship from the Burroughs Wellcome Fund.

Received May 17, 1999; returned for revision June 7, 1999; revised manuscript received June 10, 1999

REFERENCES

- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. 1994. *Current protocols in molecular biology*. New York: Wiley.
- Dianzott AM, Bujarski JJ. 1988. An in vitro transcription vector which generates nearly correctly ended RNAs by self-cleavage of longer transcripts. *Nucleic Acids Res* 16:10940.
- Gardner L, Mookhtiar KA, Coleman JE. 1997. Initiation, elongation, and processivity of carboxyl-terminal mutants of T7 RNA polymerase. *Biochemistry* 36:2908–2918.
- Groberg J, Dunn JJ. 1988. Purification of T7 RNA polymerase. *J Bacteriol* 170:1244–1253.
- Helm M, Brule H, Giege R, Florentz C. 1999. More mistakes by T7 RNA polymerase at the 5' ends of in vitro transcribed RNAs. *RNA* 5:618–621.
- Hore PJ. 1983. Solvent suppression in Fourier transform nuclear magnetic resonance. *J Mag Reson* 55:283–300.
- Krupp G. 1988. RNA synthesis: Strategies for the use of bacteriophage RNA polymerases. *Gene* 72:74–89.
- MacDonald LE, Zhou Y, McAllister WT. 1993. Termination and slippage by bacteriophage T7 RNA polymerase. *J Mol Biol* 232:1030–1047.
- Martin CT, Muller DK, Coleman JE. 1988. Processivity in early stages of transcription by the T7 RNA polymerase. *Biochemistry* 27:3966–3974.
- Miller WA, Bujarski JJ, Dreher TW, Hall TC. 1986. Minus-strand initiation by bromo mosaic virus replicase within the 3' tRNA-like structure of native and modified RNA templates. *J Mol Biol* 187:537–546.
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15:8783–8798.
- Moran S, Ren RXF, Sheils CJ, Rumney S, Kool ET. 1996. Non-hydrogen bonding terminator nucleosides increase the 3'-end homogeneity of enzymatic RNA and DNA synthesis. *Nucleic Acids Res* 24:2044–2052.
- Moroney SE, Piccirilli JA. 1991. Abortive products as initiation nucleotides during transcription by T7 RNA polymerase. *Biochemistry* 30:10343–10349.
- Murphy SK, Park GD. 1997. Genome nucleotide lengths that are divisible by six are not essential but enhance replication of defective interfering RNAs of the paramyxovirus simian virus 5. *Virology* 232:144–157.
- Plateau P, Gueron M. 1982. Exchangeable proton NMR without baseline distortion, using new strong-pulse sequences. *J Am Chem Soc* 104:7311–7312.
- Pleiss JA, Derrick ML, Uhlenbeck OC. 1998. T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* 4:1313–1317.
- Price SR, Ito N, Outbridge C, Avis JM, Nagai K. 1995. Crystallization of RNA-protein complexes. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 249:398–408.
- Siegel R, Bellon L, Beigelman L, Kao CC. 1999. Use of RNA, DNA, and chimeric templates by a viral RNA-dependent RNA polymerase (RdRp) suggests that RdRp can mediate the transition from the RNA to the DNA world. *J Virol* 73:6424–6429.
- Siegel RW, Adkins S, Kao CC. 1997. Sequence-specific recognition of a subgenomic promoter by a viral RNA polymerase. *Proc Natl Acad Sci USA* 94:11238–11243.
- Sun JH, Adkins S, Faure G, Kao CC. 1996. Initiation of (–)-strand RNA synthesis catalyzed by the BMV RNA-dependent RNA polymerase: Synthesis of oligonucleotides. *Virology* 226:1–12.
- Symensma TL, Giver L, Zapp M, Takle GB, Ellington AD. 1996. RNA aptamers selected to bind human immunodeficiency virus type 1 Rev in vitro are responsive in vivo. *J Virol* 70:179–187.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:504–510.
- Wu M, Tinoco I. 1998. RNA folding causes secondary structure rearrangements. *Proc Natl Acad Sci USA* 95:11554–11560.
- Wyatt JR, Chastin M, Puglisi JD. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *BioTechniques* 11:764–769.