METHOD

T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates

JEFFREY A. PLEISS, MARIA L. DERRICK, and OLKE C. UHLENBECK

Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA

ABSTRACT

The use of T7 RNA polymerase to prepare large quantities of RNA of a particular sequence has greatly facilitated the study of both the structure and function of RNA. Generally, it has been believed that the products of this technique are highly homogeneous in sequence, with only a few noted exceptions. We have carefully examined the transcriptional products of several tRNAs that vary in their 5' end sequence and found that, for those molecules that begin with multiple, consecutive guanosines, the transcriptional products are far from homogenous. Although a template beginning with GCG showed no detectable 5' end heterogeneity, two tRNA templates designed to have either four or five consecutive guanosines at their 5' ends had more than 30% of their total transcriptional products extended by at least one untemplated nucleotide at their 5' end. By simply reducing the number of consecutive guanosines, the heterogeneity was reduced significantly. The presence of this 5' end heterogeneity in combination with the 3' end heterogeneity common to T7 transcriptions results in a mixture of RNA molecules even after rigorous size purification.

Keywords: aminoacyl-tRNA synthetase; initiation; phosphorothioate; slippage; tRNA

INTRODUCTION

In vitro transcription by T7 RNA polymerase is used widely to produce RNA for biophysical, biochemical, and molecular biological experiments (Milligan et al., 1987; Sampson & Uhlenbeck, 1988; Tuerk & Gold, 1990; Xiong & Lommel, 1991; Batey et al., 1992; Nikonowicz et al., 1992; Doudna et al., 1993). Most transcription reactions show a significant amount of 3'-terminal heterogeneity where the expected runoff product is contaminated by shorter products arising from premature terminations and longer products arising from nontemplated additions of residues (Schenborn & Mierendorf, 1985; Milligan et al., 1987). In addition, certain templates yield unexpectedly large products caused by internal slipping at homopolymeric runs (Groebe & Uhlenbeck, 1988; Macdonald et al., 1993) or self-priming (Konarska & Sharp, 1989; Krupp, 1989; Cazenave & Uhlenbeck, 1994). However, in almost all cases, it is possible to obtain homogeneous product by rigorous purification of RNA

of the correct size, usually by PAGE. Here we report evidence that transcripts beginning with consecutive guanosine residues contain significant 5'-terminal heterogeneity that can make it impossible to recover pure RNA by simple size purification.

RESULTS AND DISCUSSION

Inconsistencies in the kinetic properties of several tRNAs prepared by in vitro transcription with T7 RNA polymerase prompted a closer examination of the products of their transcription reactions. Figure 1 shows the desired products from in vitro transcription of five different tRNAs, chosen for this study primarily because they have varying numbers of consecutive guanosines at their 5' ends. Using the usual methods, each of these sequences was cloned into pUC19 plasmid DNA between the wild-type 17-nt T7 promoter sequence and a *Bst*N 1 restriction site (Sampson & Uhlenbeck, 1988). Plasmids were linearized with *Bst*N 1 and used for in vitro transcriptions under reaction conditions typically used to produce high yields. Each reaction contained [γ - 32 P] GTP to specifically label the 5' end of each transcription product, and ATP α S to insert a phosphorothioate linkage 5' of every adenosine in the sequence. At the desired times, the

Reprint requests to: Olke C. Uhlenbeck, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA; e-mail: olke.uhlenbeck@colorado.edu.

FIGURE 1. Desired products from in vitro transcription of CQ1 (an A1G mutant of *E. coli* tRNA^{Gln}), CQ10 (a GC 4–69 AU mutant of CQ1), YF0 (yeast tRNA^{Phe}), CA0 (*E. coli* tRNA^{Ala}), and CA2 (a GC 4–69 CG mutant of CA0).

tRNA length products were purified on a Sephadex G-25 column to remove both unincorporated NTPs and short abortive products. The purified RNAs were treated with iodine to cleave the phosphate backbone at positions of phosphorothioate incorporation (Schatz et al., 1991). The products of these cleavage reactions were separated on a sequencing gel and visualized using a phosphorimager.

Figure 2A shows the results of iodine cleavage for each of the five RNAs. As expected, the major cleavage products for each RNA are found at positions corresponding to the proper length from the 5' end of the molecule to each adenosine in the sequence. In the case of YF0, these are the only visible products. However, additional minor bands are observed for the other four sequences, indicating the presence of adenosines at other positions. The number and relative intensity of these bands is different for each of the RNAs. A careful quantitation of the products near one expected adenosine for each RNA is shown in Figure 2B. In the case of YF0, no minor bands were detected within the limit of the experiment (<0.5%). In contrast, CA0 has at least four distinct bands in an area expected to have only one adenosine. The most intense band is present at the expected position, whereas the other bands appear at positions +1, +2, and +3, comprising 20, 9, and 5%

of the total radioactivity, respectively. The three remaining sequences also show such longer products, but with varying intensities. Although the data shown for each RNA are derived from just one set of bands near a single adenosine residue, the same distribution of bands is seen for each adenosine throughout each RNA sequence. This consistent distribution of minor bands leads to the conclusion that the additional products do not result from slipping of the polymerase during the elongation phase of the transcription, but rather are molecules that have additional nucleotides at their 5' end, and otherwise are identical.

When these data are compared to the sequences of the different RNAs, a clear relationship is seen between the presence of minor bands and the number of consecutive 5'-terminal guanosines. Indeed, the derivatives of CQ1 and CA0 designed to reduce the number of consecutive 5'-terminal guanosines, CQ10 and CA2, respectively, both show significant reductions in the levels of the minor bands relative to the parent molecules. However, the presence of the bands cannot be explained simply by the number of consecutive guanosines because CA0 (with four guanosines) has more minor products than does CQ1 (with five guanosines).

A number of experiments was performed to verify that these additional bands were not a result of the

FIGURE 2. Results of iodine cleavage of phosphorothioate-containing RNAs. **A:** Sequencing gel showing products 10–76-nt long with positions 30–48 enlarged. **B:** Line integration of the radioactivity present near one expected adenosine for each RNA. The major band is present at the left end of the trace, with positions +1, +2, and +3 appearing to the right. The percentage of radioactivity present in each peak is indicated.

method used for detecting these aberrant RNA products. Transcripts prepared with 10% ATP α S and 90% ATP did not change the relative amount or position of the additional bands, indicating that the presence of the phosphorothioate containing NTP did not intrinsically produce the extra bands. Transcriptions with either CTP α S or UTP α S instead of ATP α S also gave very similar patterns for the extra products, suggesting that the substituted ATP α S did not cause the extra products. Finally, RNAs transcribed with only the four normal NTPs and [γ - 32 P] GTP were purified and subjected to partial digestion with RNase T1 to give specific cleavage after guanosines. The minor bands observed by this method were identical in amount and distribution to those seen using phosphorothioate cleavage.

Two different mechanisms may explain the 5'-terminal heterogeneity observed in the transcription products from T7 promoter sequences containing consecutive guanosine residues. One possibility is that slippage occurs during the initiation processes such that, after the second nucleotide is added to form pppGpG, the dinucleotide can slip back by one residue in the enzyme active site prior to further elongation. Such a mechanism has been proposed to explain 5'-terminal heterogeneity produced by *Escherichia coli* RNA polymerase when initiating transcription with consecutive adenosines or uridines (Guo & Roberts, 1990; Xiong & Reznikoff, 1993). A related possibility is based on the observation that the

large amount of abortive initiation products produced by T7 RNA polymerase can serve as primers for future initiation events (Moroney & Piccirilli, 1991). If one assumes that abortive products can reinitiate in an incorrect frame, 5' end heterogeneity would result. More work is needed to clearly define the exact mechanism or mechanisms by which 5'-terminal heterogeneity is produced by this polymerase.

Regardless of the mechanism, the presence of 5'-terminal heterogeneity greatly complicates the isolation of RNA molecules with both length and sequence homogeneity because of the inevitable 3'-terminal heterogeneity that results from premature termination and nontemplated additions of nucleotides by T7 RNA polymerase. To illustrate this point, one of these tRNAs was subjected to a functional assay. Using [α - 32 P] CTP, CA0 was transcribed to a low specific activity, allowing the transcription products to be purified to single-nucleotide resolution on a polyacrylamide gel (Fig. 3A). Each of the three major transcription products was assayed subsequently for its ability to be aminoacylated by *E. coli* alanyl-tRNA synthetase. Because aminoacyl-tRNA synthetases require a 3'-terminal CCA sequence in order to covalently attach an amino acid, the fraction of each band with a properly terminated 3' end can be estimated by measuring the amount of tritiated amino acid incorporated into the RNA. As shown in Figure 3B, approximately 80% of the molecules in the proper sized

FIGURE 3. A: Purification of internally labeled CA0 transcript. Lane 1 is the unpurified transcription reaction, and lanes 2–4 are the purified transcripts. Contamination seen in the 77-nt sample represents less than 10% of the total. **B:** Fraction of each purified transcript able to be aminoacylated by *E. coli* alanyl-tRNA synthetase.

transcript (76-nt long) are able to be aminoacylated. The transcript one nucleotide shorter was only aminoacylatable to a very low level, whereas the transcript one nucleotide longer reached a level of approximately 55%. These results correlate well with the RNA sequencing data. Because none of the detectable 5' end heterogeneity seen for this molecule corresponded to shorter products, the entire 75-nt transcript must result from premature 3'-end termination events. The level of aminoacylation of the 76-nt sample confirms that at least 80% of the molecules have properly terminated 5' and 3' ends. Although the inability of the remaining 20% to be aminoacylated may be explained partly by misfolding of the RNA (or other such problems), it is also undoubtedly due to contamination of the sample with molecules that are one or more nucleotides longer at the 5' end and an equal number of nucleotides shorter at the 3' end. The 77-nt transcript actually contains a majority of molecules (55%) that have a properly formed 3'-terminal CCA, but are longer by one nucleotide at their 5' end. The molecules with additional 5'-end nucleotides remain active in this assay, whereas the other 45% of the molecules do not.

All products of in vitro transcription using T7 RNA polymerase will potentially contain some amount of 5'-end heterogeneity. The amount of heterogeneity produced for a given RNA will vary from extremely high to undetectable levels, depending on several variables, including the sequence of the RNA, the length of the RNA, the NTP concentrations, and the duration of the transcription. The amount of heterogeneity acceptable to the experimentalist will depend upon the intended use of the RNA molecules. When the RNA is to be used for such experiments as in vitro translation assays, even the high levels of heterogeneity produced from the CA0 template (~35% of the total transcription) may be tolerable. Conversely, RNA intended for structural studies either by NMR or X-ray crystallography is often required to be of very high homogeneity. Although undetectable in our experiments, the levels of heterogeneity produced from the YF0 template may be sufficient to prevent crystallization. For typical biochemical experiments using in vitro-transcribed RNA, 90% sample purity is often considered acceptable. However, even a 10% contamination of molecules with 5'-end heterogeneity may greatly complicate the analysis of certain experiments. For example, in a multiple turnover aminoacylation assay of the 76-nt CA0 substrate presented above, the 20% of the molecules that were inactive for aminoacylation may be able to bind the enzyme and influence the experimentally determined kinetic constants of the active population. Low levels of contamination may also lead to false assignment of minor cleavage sites in ribozyme assays, or incorrect location of branch points in splicing assays.

Given the observations presented here, it is recommended that any RNA made by T7 transcription should be examined carefully. When the amount of heterogeneity is determined to be too high, one of two steps can be taken. In situations where the 5' end sequence can be changed without altering the function of the RNA, it can be changed to a sequence that produces very low levels of heterogeneity, such as that of YF0. When this is not possible, the only way to ensure proper 5' and 3' end formation is to employ some form of post- or co-transcriptional processing event (Dzianott & Bujarski, 1988; Grosshans & Cech, 1991; Price et al., 1995; Ferre & Doudna, 1996; Lapham & Crothers, 1996; Lapham et al., 1997; Santoro & Joyce, 1997).

MATERIALS AND METHODS

Transcription and cleavage of 5' end-labeled RNA

In vitro transcriptions were performed in 50 mM Tris-HCl, pH 8.1, 50 mM NaCl, 1 mM each NTP, 20 mM MgCl₂, 1 mM spermidine, 0.1 mg/mL *Bst*N 1-linearized plasmid DNA, and 0.03 mg/mL T7 RNA polymerase. Included in the reactions was [γ -³²P] GTP (NEN) at a specific activity of 1,000 Ci/mol.

When desired, phosphorothioate incorporation was accomplished by substituting the corresponding NTP with either 1 mM NTP α S (for 100% incorporation), or a mixture of 0.9 mM NTP and 0.1 mM NTP α S (for 10% incorporation). After 4 h at 37 °C, the reactions were quenched by addition of 10 volumes of 300 mM sodium acetate, 0.5 mM EDTA, and subsequently purified on a Sephadex G-25 column pre-equilibrated in the same buffer. The samples were precipitated with three volumes of 100% ethanol, followed by two washes with 70% ethanol. Cleavage of the phosphorothioate-containing RNAs was accomplished in a reaction containing 10 mM Tris-HCl, pH 8.0, 1 mM iodine, and 10% ethanol. After 5 min at room temperature, the reactions were quenched by addition of sodium bisulfite to 10 mM. Partial cleavage of the RNAs with RNase T1 was accomplished as described in Ehresmann et al. (1987). The products of cleavage reactions were analyzed on a 45-cm 12% polyacrylamide sequencing gel run until the bromophenyl blue migrated 45 cm.

Transcription and aminoacylation of internally labeled RNAs

The RNAs for aminoacylation reactions were made by transcription with the four NTPs under the same conditions, but with [α - 32 P] CTP (50 Ci/mol) (NEN). After 4 h at 37 °C, 10 units of calf intestinal phosphatase (New England Biolabs) were added to each of the reactions and incubated for an additional hour at 37 °C. Reactions were purified with Sephadex G-25 and precipitated as before. Samples were purified on a 45-cm 12% polyacrylamide gel that was run until the xylene cyanol had migrated a total distance of 65 cm. The gel was exposed to film in order to isolate the individual bands. Aminoacylation reactions were performed in 30 mM HEPES, pH 7.5, 30 mM KCl, 15 mM MgCl₂, 4 mM ATP, 7 mM β -ME, 20 μ M 3 H-alanine (40 Ci/mmol) (Amersham), 0.2 μ M RNA, and either 10 nM or 100 nM *E. coli* alanyl-tRNA synthetase. RNA concentration was determined based on the known specific activity of the sample. Reactions were performed at 37 °C, with aliquots removed at various time points and quenched by spotting onto 5% trichloroacetic acid soaked Whatman 3MM paper. After washing in 5% TCA, the individual filters were counted by liquid scintillation counting while separating the counts resulting from the 3 H and the 32 P.

ACKNOWLEDGMENTS

This research was supported by a National Institutes of Health grant GM37552 to O.C.U.

Manuscript accepted without revision June 23, 1998

REFERENCES

- Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR. 1992. Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res* 20:4515–4523.
- Cazenave C, Uhlenbeck OC. 1994. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci USA* 91:6972–6976.
- Doudna JA, Grosshans C, Gooding A, Kundrot CE. 1993. Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc Natl Acad Sci USA* 90:7829–7833.
- Dzianott 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.
- Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res* 15:9109–9128.
- Ferre DAAR, Doudna JA. 1996. Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res* 24:977–978.
- Groebe DR, Uhlenbeck OC. 1988. Characterization of RNA hairpin loop stability. *Nucleic Acids Res* 16:11725–11735.
- Grosshans CA, Cech TR. 1991. A hammerhead ribozyme allows synthesis of a new form of the *Tetrahymena* ribozyme homogeneous in length with a 3' end blocked for transesterification. *Nucleic Acids Res* 19:3875–3880.
- Guo HC, Roberts JW. 1990. Heterogeneous initiation due to slippage at the bacteriophage 82 late gene promoter in vitro. *Biochemistry* 29:10702–10709.
- Konarska MM, Sharp PA. 1989. Replication of RNA by the DNA-dependent RNA polymerase of phage T7. *Cell* 57:423–431.
- Krupp G. 1989. Unusual promoter-independent transcription reactions with bacteriophage RNA polymerases. *Nucleic Acids Res* 17:3023–3036.
- Lapham J, Crothers DM. 1996. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA* 2:289–296.
- Lapham J, Yu YT, Shu MD, Steitz JA, Crothers DM. 1997. The position of site-directed cleavage of RNA using RNase H and 2'-O-methyl oligonucleotides is dependent on the enzyme source [Letter to the Editor]. *RNA* 3:950–951.
- Macdonald LE, Zhou Y, McAllister WT. 1993. Termination and slippage by bacteriophage T7 RNA polymerase. *J Mol Biol* 232:1030–1047.
- 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.
- Moroney SE, Piccirilli JA. 1991. Abortive products as initiating nucleotides during transcription by T7 RNA polymerase. *Biochemistry* 30:10343–10349.
- Nikonowicz EP, Sirt A, Legault P, Jucker FM, Baer LM, Pardi A. 1992. Preparation of 13 C and 15 N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res* 20:4507–4513.
- Price SR, Ito N, Oubridge C, Avis JM, Nagai K. 1995. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 249:398–408.
- Sampson JR, Uhlenbeck OC. 1988. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc Natl Acad Sci USA* 85:1033–1037.
- Santoro SW, Joyce GF. 1997. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci USA* 94:4262–4266.
- Schatz D, Leberman R, Eckstein F. 1991. Interaction of *Escherichia coli* tRNA(Ser) with its cognate aminoacyl-tRNA synthetase as determined by footprinting with phosphorothioate-containing tRNA transcripts. *Proc Natl Acad Sci USA* 88:6132–6136.
- Schenborn ET, Mierendorf RC Jr. 1985. A novel transcription property of SP6 and T7 RNA polymerases: Dependence on template structure. *Nucleic Acids Res* 13:6223–6236.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.
- Xiong XF, Reznikoff WS. 1993. Transcriptional slippage during the transcription initiation process at a mutant lac promoter in vivo. *J Mol Biol* 231:569–580.
- Xiong ZG, Lommel SA. 1991. Red clover necrotic mosaic virus infectious transcripts synthesized in vitro. *Virology* 182:388–392.