

LETTER TO THE EDITOR

More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs

MARK HELM, HERVÉ BRULÉ, RICHARD GIEGÉ and CATHERINE FLORENTZ

Unité Propre de Recherche 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire,
F-67084 Strasbourg Cedex, France

Keywords: human; mitochondria; ribozyme; tRNA

INTRODUCTION

In vitro transcription with T7 RNA polymerase is one of the most powerful tools in RNA research. This is mainly linked to the easy preparation of the enzyme, the quite unlimited range of sizes and sequences of the RNA that can be synthesized, as well as the efficiency and accuracy of synthesis. So far only two major limitations are common knowledge, namely poor transcription efficiency of non-G-rich initial sequences (Dunn & Studier, 1983) and 3'-end heterogeneities of the transcripts by 1 or 2 nt (Milligan et al., 1987; Draper et al., 1988; Kholod et al., 1998). These drawbacks have been overcome in most cases by improving the initiating sequence and/or purifying the transcription products to single-nucleotide resolution when necessary. The Uhlenbeck laboratory has published the first evidence for shifty errors by the T7 RNA polymerase at the very 5' end of the synthesized RNA, an as yet unsuspected event (Pleiss et al., 1998). They demonstrated that in the particular case of tRNA sequences starting with G-rich stretches, the enzyme incorporates additional nontemplated G residues in up to 30% of the transcripts. The authors point to the fact that these errors, in combination with the 3'-end heterogeneity common to T7 transcripts, may be crucial, for instance, for functional studies of tRNAs, because some in vitro-transcribed tRNA molecules of rigorously correct final size may be erroneously considered as functional.

Here we bring additional evidence for novel types of sly errors by the T7 RNA polymerase at the 5' end of transcripts. These errors occur (1) at very high frequency (80 to 100% for a given template), (2) on initial sequences other than homopolymeric G stretches, and

(3) either by incorporation of a single nontemplated nucleotide (predominantly an A) or omission of the first nucleotide. Here also, a combination of these errors of the variety of those possible at the 3'-end leads to incorrect RNAs, despite their rigorous size purification.

RESULTS

Template with a 5'-C¹ACUGU⁶ sequence leads to transcripts with one 5' additional nucleotide

Human mitochondrial tRNA^{Lys} (Anderson et al., 1981) was initially produced by in vitro transcription of the corresponding synthetic gene cloned directly downstream of the T7 RNA promoter (named herein Kwt). Despite a supposedly inefficient 5'-end initial transcription sequence C¹ACUGU⁶, Kwt was obtained in sufficient amount for structural and functional investigations, usually 170 pmol/pmol of plasmid. However, a sequence problem at the 5' end of the transcripts was soon suspected (Helm et al., 1998). Here we investigate the 5'-end quality of Kwt by comparison with RNAs produced by two other approaches. tRNA^{Lys} Kwt-Rz was obtained as a "transzyme" from in vitro transcription of the corresponding gene cloned downstream of a hammerhead ribozyme (Price et al., 1995; Fechter et al., 1998), and Kwt-Lg has been constructed by ligation of a chemically synthesized 9-mer 5'-C¹ACUGUAAA⁹-3' corresponding to the 5' end of the tRNA to an in vitro-transcribed 64-mer 5'-G¹⁰CUAA----CCA⁷⁶-3' forming the 3' part of the tRNA (M. Helm, P. Guerné, R. Giegé, & C. Florentz, in prep.).

The quality of the 5' end of Kwt has been compared to that of the two other tRNA preparations, Kwt-Rz and Kwt-Lg considered as reference molecules, by 5'-end labeling and investigation of their partial RNase T1 hydrolysis products, as well as for the nature of the 5'-end nucleotide.

Sequencing ladders obtained by limited digestions of the three types of RNA with G-specific RNase T1

Reprint requests to: Catherine Florentz, Unité Propre de Recherche 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, F-67084 Strasbourg Cedex, France; e-mail: florentz@ibmc.u-strasbg.fr.

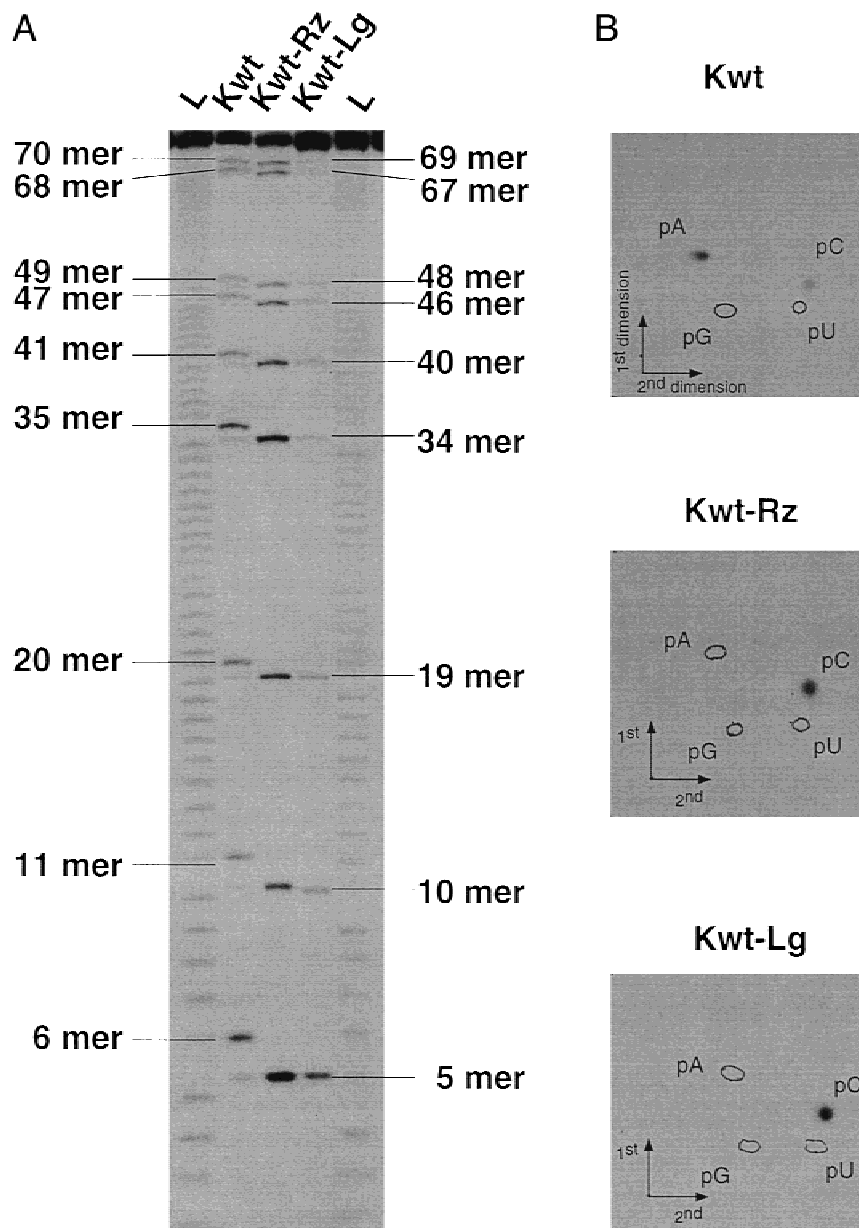


FIGURE 1. Analysis of the 5'-ends of human mitochondrial tRNA^{Lys} preparations. tRNAs were prepared either by *in vitro* transcription of the cloned synthetic gene (Kwt) or by the "transzyme" approach (Kw-Rz; Fechter et al., 1998), or by ligation of a chemically synthesized fragment with an *in vitro* transcript (Kw-Lg) (see text). Transcription conditions for Kwt were for 3 h at 37 °C in 40 mM Tris-HCl, pH 8.1 (at 37 °C), 22 mM MgCl₂, 5 mM dithiothreitol, 0.01% Triton-X100, 1 mM spermidine, 4 mM each nucleoside triphosphate, 0.1 mg/mL linearized plasmid (Helm et al., 1998) and 15 μg/mL T7 RNA polymerase prepared according to Becker et al. (1996). tRNA was first 5'-end labeled by T4 polynucleotide kinase in the presence of a large excess of ATP under otherwise standard conditions (Silberklang et al., 1979). The excess of ATP is aimed at avoiding preferential labeling of some RNA families by the kinase (Nishimura, 1979b). Labeled molecules were purified by short migration on a 12% polyacrylamide/urea gel. **A:** RNase T1 hydrolysis products. Alkaline ladders (L) and denaturing RNase T1 digestion profiles were obtained as described (Helm et al., 1998). The size of the fragments was assigned according to the T1 ladder obtained with the unambiguous reference molecule Kwt-Lg. **B:** 5'-end nucleotide analysis. Labeled tRNA, mixed with carrier yeast total tRNA, was digested to completion by nuclease P1, releasing 5'-phosphate- 3'OH mononucleosides. The digests were spotted onto 10 × 10 cm² Cellulose Avicel thin layer chromatographic plates, and mononucleosides were separated by two-dimensional chromatography. The first dimension was run in isobutyric acid:25% NH₄OH:H₂O (50:1.1:28.9), and the second dimension in 0.1 M sodium phosphate buffer (pH 6.8):ammonium sulfate: *n*-propanol [100:60:2 (v:w:v)] (Nishimura, 1979a). Radioactive spots were detected by exposure of a PhosphorImager screen to the chromatogram (FUJIX Bio-Imaging Analyzer BAS 2000). Assignment of spots was done according to an established table (Keith, 1995). Quantitative analysis was performed with MacBas software.

(Fig. 1A) highlight a constant and predominant 1-nt upward shift all along the G-ladder of Kwt relative to that of Kwt-Rz and Kwt-Lg. Only a small proportion of RNase T1-generated fragments of Kwt has the same size as the fragments originating from Kwt-Rz or Kwt-Lg. Thus, whereas Kwt-Rz and Kwt-Lg consist of one tRNA population that is homogenous with regard to the 5'-end, Kwt contains two populations of molecules differing in size by 1 nt at the 5' end. Quantitation of the relative distribution of the radioactivity within the two fragments generated by RNase T1 hydrolysis at each position within Kwt reveals that ~80% of the molecules have a 1-nt extension at their 5' end and ~20% have the expected correct size. The erroneous addition of an extra nucleotide at the 5' end has also been found on a series of variants of Kwt (not shown) having individual internal point mutations at nucleotides located at least 9 nt from the 5' end and sharing the same 5'-end sequence with Kwt (Helm et al., 1998).

Analysis of the 5' nucleotide was done by complete digestion to monophosphate nucleosides by nuclease P1, and by two-dimensional thin layer chromatography (TLC). Nuclease P1 liberates 5'-phosphate-3'OH mononucleosides, and thus releases the [³²P]-labeled 5'-end nucleotide from the RNA. As shown in Figure 1B, Kwt contains substantial amounts of adenosine (75%) and guanosine (5%) in addition to the expected cytidine (20%). On the other hand, both Kwt-Rz and Kwt-Lg start with the expected cytidine in the complete population of molecules.

In summary, these data demonstrate that T7 RNA polymerase makes errors at the 5' end of the RNA during transcription of the Kwt gene. In ~80% of the transcription products, an additional nucleotide is incorporated. This nucleotide is always a purine with a large preference for adenines.

Template with a 5'-C¹AGAGA⁶ sequence leads to transcripts with 5' addition or omission of one nucleotide

Aware of the problems encountered with tRNA^{Lys}, human mitochondrial tRNA^{Pro} produced by *in vitro* transcription with T7 RNA polymerase has also been analyzed in detail with respect to its 5' extremity. The corresponding gene starts with a likely unfavorable 5'-C¹AGAGA⁶ sequence, but is rather efficiently transcribed (220 pmol/pmol plasmid). After transcription in the presence of an excess of CMP and 5'-end labeling with T4 polynucleotide kinase, analysis of the RNA population on a denaturing high resolution polyacrylamide gel leads to four well-separated bands. The RNAs contained in each individual band were analyzed for the quality of their 5' extremity by size analysis of RNase T1 digests and by analysis of the 5'-end nucleotide. As seen in Figure 2, only 30–35% of one of these RNA populations (band 2) corre-

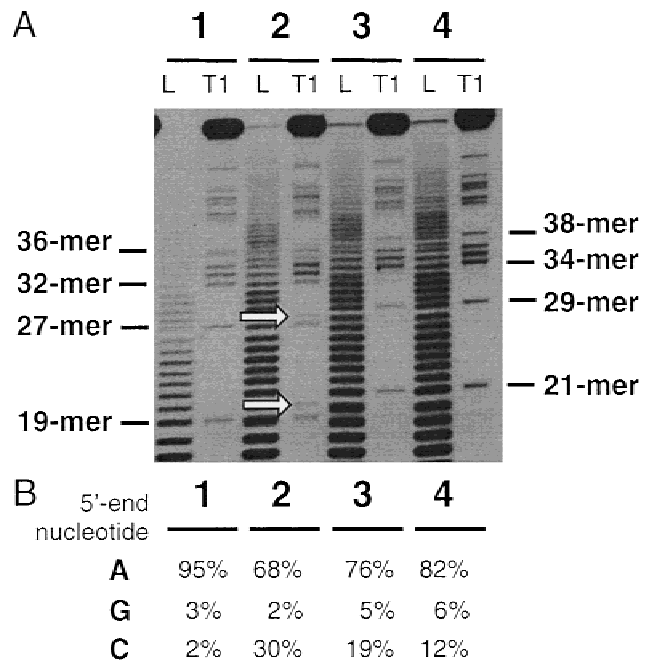


FIGURE 2. Analysis of the 5' ends of the transcription products of human mitochondrial tRNA^{Pro} gene. The synthetic gene of tRNA^{Pro} has been cloned and transcribed in the presence of T7 RNA polymerase and CMP as described (Brulé et al., 1998). Transcription products were dephosphorylated, 5'-end labeled with T4 polynucleotide kinase, and separated to single-nucleotide resolution by long run on a 12% denaturing polyacrylamide/8 M urea gel. Four bands were detected (numbered 1 through 4 starting with the smallest), individually excised, and eluted. **A:** Autoradiography of typical alkali (L) and RNase T1 (T1) digests of the RNAs. The size of the fragments was assigned by comigration of hydrolysis products from tRNA^{Lys} derivatives (not shown). The white arrows point to T1 fragments of expected size according to the cloned gene sequence. **B:** Analysis of 5'-end nucleotides. Experimental conditions were as described in Figure 1 except that labeling has been performed under standard conditions, that is, in the presence of an excess of RNA.

sponds to the expected tRNA^{Pro}. RNase T1 fragments are of expected size (Fig. 2, white arrows) and the 5'-end nucleotide is likely a C. All other RNA populations lead to unexpected T1 fragments (too long or too short by 1 nt) and have predominantly an A as the terminal nucleotide. A complete quantitative analysis of all different RNAs produced during the transcription of the tRNA^{Pro} gene, combining the relative amount of each of the four individual original bands as quantified by densitometry (band 1: 12%, band 2: 18%, band 3: 34%, band 4: 36%) and the proportion of correct-size T1 fragments in population 2, leads to a final value of at most 5% of tRNA^{Pro} with a correct 5' end. Thus, ~95% of the transcription products are erroneous. Interestingly, when transcription is performed in the absence of CMP, none of the transcripts has the correct 5' end (results not shown).

In summary, the T7 RNA polymerase makes two types of errors. Either it starts transcription at nt +2, which corresponds to an A (RNA from band 1 and 65–70% of RNA from band 2) or it adds one extra nucleotide, ei-

ther A or C (RNAs from bands 3 and 4) upstream of nt +1. Since four families of RNAs are produced, heterogeneities at the 3' end are obviously also present.

DISCUSSION

Towards a catalog of error-prone sequences for transcription by T7 RNA polymerase

The data presented here confirm and enlarge the concept reported by the Uhlenbeck laboratory in demonstrating that T7 RNA polymerase can make a variety of errors at the 5' end of in vitro transcripts. The enzyme adds or omits a nucleotide in a very large proportion of the synthesized RNAs. This error is not necessarily related to a homopolymeric nucleotide stretch in the template, but can hit a number of other sequences. Preliminary data on in vitro transcripts of several other human mitochondrial tRNAs show that the link with the initial sequence to be transcribed (in other words the part of the T7 promoter that is internal to the gene) is not yet obvious. Indeed, whereas 5'-ACUCUU... (tRNA^{Gly}), 5'-AGAAAU (tRNA^{Ile}), and 5'-GUUAAG (tRNA^{LeuUUR}) sequences, which rather resemble those from tRNA^{Lys} and tRNA^{Pro}, are synthesized without errors, the A-rich 5'-GAAAAA (tRNA^{SerAGY}) is produced partially with an additional 5'-end nucleotide. Thus, as long as the sequence rules for 5'-end error-prone transcriptions are not understood, the possibility of 5' errors likely should be considered for any tRNA or RNA transcribed by T7 RNA polymerase. The list of unfavorable sequences may well become longer. In agreement with Pleiss et al. (1998), we believe that, because of the well-known errors unavoidable at the 3'-end of the transcripts, even tRNAs rigorously purified for a correct size may bear overlooked errors at both ends. Thus, analysis of the 5' extremity of the transcript is recommended.

The variety of internal promoter sequences leading to errors at the initial stage of transcription suggests the existence of a variety of mechanism(s) by which these mistakes occur. Whereas a "slippage" mechanism likely applies for errors on homooligomeric G stretches as discussed by Pleiss et al. (1998), alternative mechanisms have to be considered for nonhomooligomeric sequences such as those existing in the mitochondrial tRNAs investigated here. More work is needed to elucidate the sequence contexts that lead to this new type of transcription error.

ACKNOWLEDGMENTS

We thank Anne Théobald-Dietrich for advise in the ribozyme construction. This work was supported by grants from the

Centre National de la Recherche Scientifique (CNRS), Université Louis Pasteur (Strasbourg) and Association Française contre les Myopathies (AFM). M.H. was supported by a Marie-Curie fellowship from the EC. H.B. was supported by an AFM grant.

Received December 30, 1998; returned for revision February 16, 1999; revised manuscript received March 1, 1999

REFERENCES

- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon JC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Becker HD, Giegé R, Kern D. 1996. Identity of prokaryotic and eukaryotic tRNA^{Asp} for aminoacylation by aspartyl-tRNA synthetase from *Thermus thermophilus*. *Biochemistry* 35:7447–7458.
- Brulé H, Holmes WM, Keith G, Giegé R, Florentz C. 1998. Effect of a mutation in the anticodon of human mitochondrial tRNA^{Pro} on its post-transcriptional modification pattern. *Nucleic Acids Res* 26:537–543.
- Cech TR. 1992. Ribozyme engineering. *Curr Biol* 2:605–609.
- Draper DE, White SA, Kean JM. 1988. Preparation of specific ribosomal RNA fragments. *Methods Enzymol* 164:221–237.
- Dunn JJ, Studier FW. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the location of T7 genetic elements. *J Mol Biol* 166:477–535.
- Fechter P, Rudinger J, Giegé R, Théobald-Dietrich A. 1998. Ribozyme-processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: Production and activity. *FEBS Lett* 436:99–103.
- Helm M, Brulé H, Degoul F, Cepanec C, Leroux J-P, Giegé R, Florentz C. 1998. The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res* 26:1636–1643.
- Keith G. 1995. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* 77:142–144.
- Kholod N, Vassilenko K, Shlyapnikov M, Ksenzenko V, Kisselev L. 1998. Preparation of active tRNA gene transcripts devoid of 3'-extended products and dimers. *Nucleic Acids Res* 26:2500–2501.
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA template. *Nucleic Acids Res* 15:8783–8798.
- Nishimura S. 1979a. Chromatographic mobilities of modified nucleotides. In: Schimmel P, Söll D, Abelson JN, eds. *Transfer RNA: Structure, properties, and recognition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. Appendix III, pp 551–552.
- Nishimura S. 1979b. Modified nucleosides in tRNA. In: Schimmel P, Söll D, Abelson JN, eds. *Transfer RNA: Structure, properties, and recognition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 59–79.
- 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, Oubridge C, Avis M, Nagai K. 1995. Crystallization of RNA-protein complexes. I. Methods for large-scale preparations of RNA suitable for crystallographic studies. *J Mol Biol* 249:398–408.
- Silberklang F, Gillum AM, RajBhandary UL. 1979. Use of in vitro ³²P labeling in the sequence analysis of nonradioactive tRNAs. *Methods Enzymol* 59:58–109.