On the use of T7 RNA polymerase transcripts for physical investigation

Alexander A.Szewczak, Susan A.White, Daniel T.Gewirth and Peter B.Moore
Departments of Chemistry and Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

Received April 20, 1990; Revised and Accepted June 18, 1990

ABSTRACT

A few years ago we made some observations which raised questions about the accuracy with which T7 RNA polymerase transcribes templates *in vitro*, and the suitability of its *in vitro* products for biophysical study (1). The experiments described below demonstrate that there is no reason for concern; the products of T7 RNA polymerase transcription *in vitro* are as suitable for biophysical characterization as RNAs synthesized *in vivo*. It is likely that aggregation involving the transcribed portions of the T7 RNA polymerase promoter caused our initial observations.

INTRODUCTION

RNAs synthesized *in vitro* by the RNA polymerase from bacteriophage T7 have increasingly been used for biophysical studies in recent years. The only limitation on the sequences that can be made this way is imposed by the fact that the promoter for T7 RNA polymerase includes the first 4 bases of the transcript (2). The first few nucleotides of the transcript must be those of a high efficiency promoter if large amounts of product are to be made (3). The best sequences are rich in G residues.

A few years ago we made several T7 transcripts related to the 5S RNA from *E. coli*, and were disappointed to find that their imino proton NMR spectra were markedly inferior to those of similar molecules produced by *E. coli* RNA polymerase *in vivo* (1, 4). The resonances of the *in vitro* transcripts were abnormally broad.

Contamination of samples with low molecular weight, abortive initiation products was considered a likely cause of this difficulty (3), but additional purification did not improve our spectra. Samples were also denatured and reannealed to reduce their possible conformational heterogeneity, but to no effect. The hypothesis that intermolecular aggregation might be responsible was also explored. The first transcript we made was a 5S sequence with 3 unpaired G residues added to its 5′ end to ensure efficient transcription. Since G residues are prone to aggregation, the 5′ end of the transcript was re-engineered so that 2 of the 3 G's were base paired (see 5 and references therein). The terminal G was juxtaposed to an A. This version of the molecule behaved no better than the first (4).

Having dealt with the three most likely sources of difficulty, we believed, we were led to a fourth hypothesis, that the poor quality of our spectra might reflect internal sequence heterogeneity in our transcripts. We suggested that the error rate of T7 RNA polymerase might be abnormally high under the conditions used for large scale transcription *in vitro* (1). In our view, the result other groups were obtaining with T7 transcripts at the time were not incompatible with this hypothesis (6, 7, 8, 9, 10), but clear evidence of its validity was lacking.

Below we report the results of experiments we did to determine if the T7 system can be used to make RNA samples appropriate for study by NMR. The data reported below indicate that T7 RNA polymerase, used in the 'standard' manner, transcribes templates with the required accuracy. Aggregation involving promoter sequences appears to have been the cause of the poor spectra we obtained earlier. Provided aggregation is controlled, T7-produced RNAs should be satisfactory for biophysical characterization.

MATERIALS AND METHODS

Strains and Plasmids

E. coli strains HB101, JM101 and CJ236 (dut$^-$, ung$^-$) were used in this work (see 11 and references therein). Strain BL21(DE3), which contains a chromosomal T7 RNA polymerase gene under control of the lacUV5 promotor, and *E. coli* BL21 carrying plasmid pAR1219 were generous gifts from Dr. William Studier of Brookhaven National Laboratory (12,13). M13mp18 was purchased from BRL and pUC19 was supplied by IBI. The construction of plasmids pDG07, pJK1 and pZZ601 has been described elsewhere (1,11,14)

Enzymes and reagents

Restriction endonucleases *Hin*dIII and *Bst*NI and the enzyme T4 DNA ligase were purchased from New England Biolabs. Sodium salts of ribonucleotide triphosphates were from Pharmacia. Stock nucleotide triphosphate solutions were prepared as described previously (1). T7 RNA polymerase was prepared following the protocol of Drs. Dunn and Studier (13). IPTG was purchased from Sigma and DNAse I was obtained from Worthington. Oligonucleotides for site-directed mutagenesis were synthesized on an Applied Biosystems Model 380B DNA synthesizer by Mrs. Grace Sun.

Construction of Plasmids

The methods used for constructing and selecting plasmids for *in vivo* and *in vitro* production of RNA are described in detail

elsewhere (11, 15). Plasmid pAF1 was created by inserting the 0.5 kb *Hin*dIII fragment from pDG07 into the 4 kb *Hin*dIII fragment from pZZ601 which contains an ampicillin resistance gene and a phage T7 late promotor. In this way, the coding sequence for the DG07 molecule is positioned for transcription by T7 RNA polymerase. Termination of the primary transcript occurs at the normal *E. coli* sites which are contained within the 0.5 kb fragment. Therefore, the AF1 RNA should be identical to DG07 if transcribed accurately and processed correctly.

For the oligonucleotide-directed mutagenesis necessary to create pYL2, the 0.5 kb *Hin*dIII fragment from pJK1 was placed in M13mp18 (16). This fragment contains a coding sequence closely related to DG07 which has been fused to a T7 RNA polymerase promotor. Mutants were selected using the Kunkel method (17), and sequenced to verify that the desired 8 nucleotides from the 5′ end of the JK1 coding region had been deleted. The mutagenized 0.5 kb fragment was then transferred into the *Hin*dIII site of pUC19 to create pYL2. An existing *Bst*NI site within the JK1 sequence was later used to linearize the template DNA for run-off transcriptions. Thus, the resulting YL2 RNA is 5 base pairs shorter than JK1 RNA and has no 5′ pppGpGpG overhang.

Production of RNA Molecules *in vivo*

A large scale culture of *E. coli* BL21(DE3)/pAF1 was grown in YT broth containing (per liter) 10 g Bacto Tryptone, 5g Bacto Yeast Extract, 10g NaCl and 100 μg/mL ampicillin. When the culture reached an optical density of 1.0 at 550nm, isopropyl-β,D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. This stimulated the production of T7 RNA polymerase, which in turn transcribed the DG07 gene. The IPTG induction was continued for 3 hours and then the cells were collected by centrifugation. Small soluble RNA was released from the cells by grinding with alumina in buffer containing 100 mM NH_4Cl, 10 mM $MgOAc_2$, 0.5 mM EDTA, 3 mM mercapto-ethanol, 20 mM Tris-HCl (pH 7.4) and 40 μg/mL DNAse I. After centrifugation to remove the cell debris and alumina the resulting supernatant was extracted twice with phenol. The nucleic acid was recovered by ethanol precipitation.

AF1 RNA was further purified by high performance liquid chromatography (HPLC). Precipitated RNA was dissolved in a small amount of HPLC buffer containing 20 mM MES (pH 6.0), 2 mM $MgCl_2$, and 20% CH_3CN and loaded on a Nucleogen DEAE 500-7 column. A 300−600 mM NaCl gradient in the same buffer system was used to separate AF1 from other small RNAs.

The production of DG07 RNA (*in vivo*) has been described elsewhere (11).

Production of RNA Molecules *in vitro*

A large scale culture of *E. coli* HB101 containing plasmid pYL2 was grown to saturation in YT broth containing 100 μg/mL ampicillin. Plasmid DNA for *in vitro* transcriptions was prepared using a modified version of standard alkaline lysis and CsCl gradient protocols (15). In order to create DNA fragments suitable for run-off transcription, the plasmid DNA was thoroughly digested with an excess of the required restriction enzyme. The digested DNA was extracted one time with phenol/chloroform (1:1), extracted with chloroform, and then ethanol precipitated before use.

Transcription reactions contained 40 mM Tris (pH 8.1), 2 mM spermidine, 5 mM DTT, 5 mM each NTP, 0.005% Triton X-100, 22.5 mM $MgCl_2$, 0.1 mg/mL T7 RNA polymerase, and 0.1 mg/mL digested plasmid DNA. The reactions were incubated at 37°C for 3 hours, quenched with 1/10 volume of 500 mM EDTA, dialyzed against H_2O overnight and then ethanol precipitated. Final purification was carried out using a Nucleogen DEAE 500−7 column at 60°C in a buffer containing 6 M urea and 20 mM KH_2PO_4 (pH 6.5). YL2 RNA was separated from undesired transcription products using a 450−600 mM KCL gradient. Samples were heated to 80°C for 5 minutes before loading to ensure full denaturation.

The production of JK1 RNA has been described previously (1). For this work an additional step was included in the protocol. The sample was purified using denaturing polyacrylamide gel electrophoresis. RNA was recovered from the gel matrix by electroelution using a Schleicher and Schuell Elutrap apparatus, then ethanol precipitated.

NMR Methods

All RNA samples were redissolved in a small volume of 1.05× NMR buffer (final concentration = 100 mM KCl, 4 mM $MgCl_2$, 0.2 mM EDTA and 4 mM cacodylate, pH 7.0), and then dialyzed against several changes of the same 1.05× buffer. Sample volumes were reduced to about 0.4 mL using Centricon microconcentrators. Samples were made 5% (v/v) in D_2O to ensure the operation of the lock system, and p-dioxane was added to a final concentration of 1.5 mM. The chemical shift of dioxane was assumed to be 3.741 ppm relative to the methyl resonance of 3-(trimethylsilyl)-1-propanesulfonic acid. The concentrations of AF1, DG07 and YL2 samples examined were all approximately 1 mM, while the JK1 sample was about 0.5 mM. Each sample migrated as one band on both native and denaturing polyacrylamide gels.

Proton nuclear magnetic resonance experiments were performed on the 490-MHz spectrometer at the Yale University Chemical Instrumentation Center. Spectra of imino protons were obtained using the twin-pulse method (18), to avoid exciting the water resonance, and alternate delay accumulation (19) was employed to overcome computer word length limitations. Unless otherwise noted, 1280 scans were collected for each spectrum in 16K blocks with the offset at about 15 ppm and with a sweep width of 14,300 Hz. The spectral resolution after Fourier transformation was therefore 1.7 Hz/pt. Data sets were not resolution enhanced prior to transformation. All spectra were collected at 30°C.

In order to measure the widths of well resolved singlet imino resonances the downfield spectrum for each sample was plotted on an expanded scale of 38 Hz/cm. Line widths were then measured at half maximum height with a ruler. An estimate of the error in each measurement was calculated using the formula:

$$\sigma^2(w) = (w(\delta h/h))^2 + (\delta w)^2$$

where w = peak width at half max. height, δw = uncertainty in width measurement, h = height and δh = uncertainty in determining baseline.

RESULTS

DG07 is a deletion mutant of *E. coli* 5S ribosomal RNA consisting of the native molecule's helices I, IV, and V (see Figure 1). Its protein binding properties and secondary structure are similar to those of the corresponding regions of intact 5S RNA. The imino proton spectra of both DG07 and its close

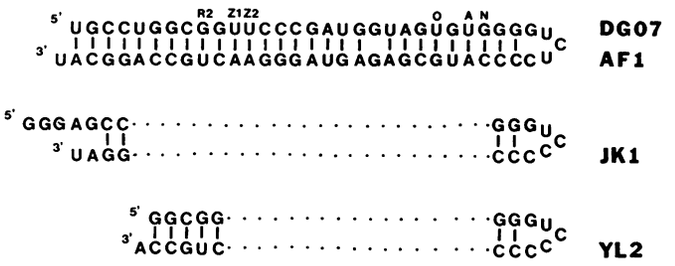

Figure 1: The sequences and secondary structure of the RNAs discussed in this work. Although transcribed by different polymerases *in vivo*, DG07 and AF1 are identical. Molecules JK1 and YL2 were produced *in vitro*. The nucleotides which give rise to the well-resolved imino proton resonances Z1, Z2, A, N, O, and R2 (see Figure 2) are indicated in the DG07/AF1 structure.

relative JK1 (see below) have been discussed elsewhere (1, 11).

Several of the imino protons in DG07 contribute well resolved resonances to its downfield spectrum. The resonances in question and the imino protons to which they correspond are indicated in Figures 1 and 2. The linewidths of these resonances are about 22 Hz (see Table 1). DG07's spectrum is the standard against which the spectra of all the other RNAs discussed below are judged.

JK1 is a derivative of DG07 created so that T7 could be used to produce a DG07-like molecule *in vitro*. In JK1 the 3′ and 5′ sequences of DG07 have been modified to allow run-off transcription (see Figure 1). The most important result of these modifications is that the 5′ end of JK1 has an unpaired pppGpGpG overhang, which DG07 does not.

In general, the imino proton spectrum of JK1 RNA looks like the spectrum of DG07 (Figure 2). Many resonances in the JK1 spectrum correspond to peaks found in the DG07 spectrum, indicating that JK1's secondary structure is normal (see 1, and 4). However, the quality of JK1's spectrum is poor (1). Although JK1 is nearly identical to DG07 in sequence and molecular weight, the resonances in its spectrum are about 20 Hz wider than DG07 resonances (Table 1).

A second JK1 sample was prepared by purification on polyacrylamide gels under denaturing conditions. Its spectrum was no different from the one shown here. Contamination by small RNAs is not the source of the problem, consistent with our earlier conclusions (1).

To determine if T7 RNA polymerase is inherently capable of producing high quality RNAs we synthesized a sample of DG07 with T7 RNA polymerase *in vivo* using pAF1 as the template. In BL21(DE3) cells T7 RNA polymerase transcribes pAF1 efficiently; our yield was 4.3 mg of pure RNA per gram of cells. To our relief, the imino proton spectrum of the resulting sample was identical to that of DG07 (Figure 2), and its linewidths were the same within error (Table 1). Clearly there is no difference between RNAs synthesized by T7 RNA polymerase and *E. coli* RNA polymerase *in vivo* as far as their spectroscopic quality is concerned.

This finding led us to reconsider the possibility that aggregation might be the cause of the abnormally broad resonances we were obtaining from *in vitro* transcripts. A template was constructed that would direct the synthesis of a DG07-like molecule by T7 RNA polymerase *in vitro*, whose terminal sequences should base pair perfectly, pYL2. It proved to be an excellent template; about 9 mg of pure RNA was produced from a 10 mL reaction mixture,

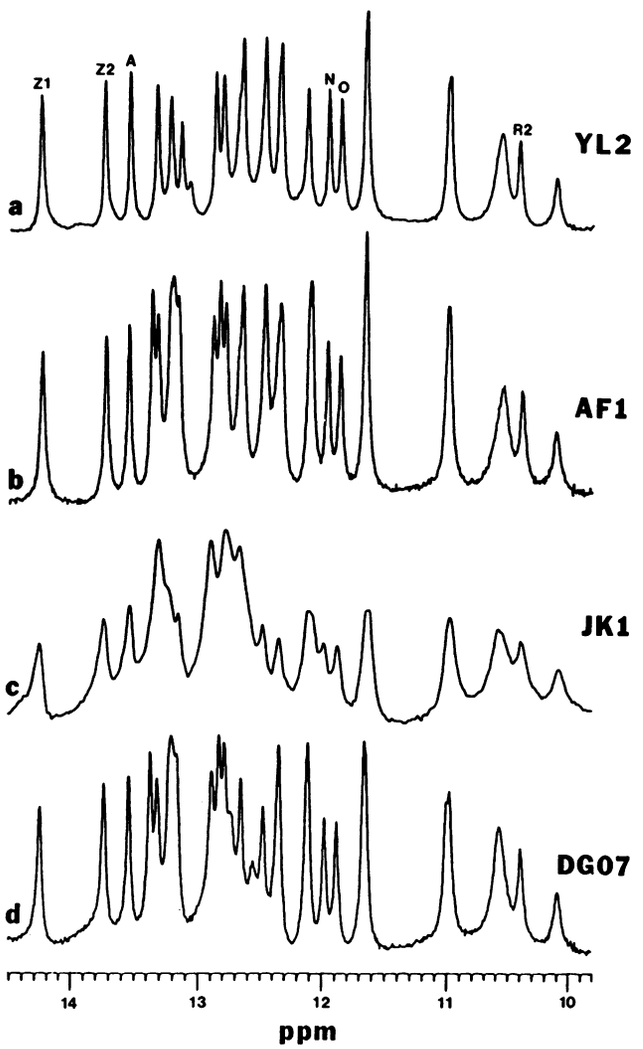

Figure 2: Imino proton spectra of DG07, AF1, JK1 and YL2. RNA samples were dialyzed into 100 mM KCl, 4 mM $MgCl_2$, 0.2 mM EDTA, and 4 mM cacodylate, pH 7.0. The concentration of RNA in each sample was about 1 mM, except for JK1 which was 0.5 mM. Spectrum a) is that of YL2 RNA, while spectrum b) is that of AF1. Spectra c) and d) belong to JK1 and DG07, respectively. All spectra were collected at 30°C and are shown unenhanced.

Table 1. Comparison of NMR linewidths from *in vivo* and in *in vitro* T7 transcripts (linewidths measured in Hertz)

	Resonance					
Transcript	Z1	Z2	A	N	O	R2
DGO7	23±2	23±2	20±2	20±2	21±2	25±2
AF1	21±2	20±2	17±2	20±2	22±2	26±2
JK1	45±3	48±8	41±8	37±8	41±8	48±8
YL2	17±2	16±2	15±2	16±2	18±2	22±2

Linewidth measurements for imino resonances Z1, Z2, A, N, O, and R2. [See Figure 1 for the locations of the protons which give rise to these resonances.] Peak widths were measured at half maximum height with a ruler, and multiplied by 38.34 Hz/cm to produce the values shown. Error estimates were calculated using the formula given in Materials and Methods.

which corresponds to 1200-fold molar amplification of the template sequence.

Unlike JK1, YL2 produces an imino proton spectrum of

excellent quality that is similar to the DG07 spectrum in most respects (Figure 2). It is missing a number of resonances found in the DG07 spectrum because it lacks DG07's 5 terminal base pairs (Figure 1). In addition, some of the remaining resonances have altered chemical shifts because of changes in ring current effects brought about by the sequence difference (see 11). The result of these changes is that the imino proton spectrum of YL2 has fewer overlaps than that of DG07. It may be an even better molecule in which to study the structure of the helix V-loop E region of 5S RNA than DG07.

The linewidths of the resolved resonances in the YL2 spectrum are compared with those of the other molecules we have discussed in Table 1. They are somewhat narrower than those in the spectra of DG07 and AF1, as expected since YL2 is a somewhat smaller molecule.

DISCUSSION

It is clear that T7 RNA polymerase is capable of transcribing templates accurately both *in vivo* and *in vitro*, contrary to our earlier suspicions. The cause of the poor spectra given by our earlier *in vitro* transcripts is not known in every case. However, our experience suggests that the 5′ overhanging guanylic acid residues many of our transcripts include may have caused aggregation, which broadens NMR spectra.

The construct we originally made to test the hypothesis that overhanging G's were causing aggregation had all its promoter residues base paired except for its 5′ most G residue, which was paired with an A (4). Apparently even that was not good enough. The only product that is fully satisfactory (YL2) is one in which all promoter residues are perfectly paired. In the case of YL2 we know that its 5′ terminal G is paired; we can detect its imino proton resonance in YL2's downfield spectrum.

We have recently done some experiments with an RNA stem-loop of completely different sequence that support the conjecture that unpaired G residues should be avoided. The version of the molecule we first examined was a T7 transcript with an overhanging 5′ pppGpGpG. It aggregated, as we could demonstrate on gels, and its resonances were unreasonably broad. We have now examined a second version of the molecule whose terminal bases are fully base paired. It does not aggregate, and its resonances are narrow (Szewczak, A., Chan, Y.-L., & Wool, I., unpublished data).

ACKNOWLEDGEMENTS

We would like to thank Dr. M. John Rogers and Joyce Sherman for their help with mutagenesis and sequencing, and Yale Liang for assisting in the construction of pYL2. S.A.W. was supported by an American Cancer Society postdoctoral fellowship and A.A.S. was the recipient of a NSF graduate fellowship. This work was supported by grants from the National Institutes of Health (AI09167, GM41651).

REFERENCES

1. Gewirth, D.T. & Moore, P.B. (1988) Nucl. Acids Res. 16, 10717–10732.
2. Dunn, J.J. & Studier, F.W. (1983) J. Mol. Biol. 166, 477–535.
3. Milligan, J.F., Groebe, D.R., Wetherell, G.W. & Uhlenbeck, O.C. (1987) Nucleic Acids Res. 15, 8783–8798.
4. Gewirth, D.T. (1988) Ph.D. thesis, Yale University.
5. Sen, D. & Gilbert, W. (1990) Nature 344, 410–414.
6. Hall, K.B., Sampson, J.R., Uhlenbeck, O.C. & Redfield, A.G. (1989) Biochemistry 28, 5794–5801.
7. Sampson, J.R. & Uhlenbeck, O.C. (1988) Proc. Natl. Acad. Sci. USA 85, 1033–1037
8. Puglisi, J.D., Wyatt, J.R. & Tinoco, I., Jr. (1988) Nature 331, 283–286.
9. Puglesi, J.D. (1989) Ph.D. thesis, University of California at Berkeley.
10. Varani, G., Wimberly, B. & Tinoco, I. (1989) Biochemistry 28, 7760–7772.
11. Gewirth, D.T. & Moore, P.B. (1987) Biochemistry 26, 5657–5665.
12. Studier, F.W. & Moffat, B.A. (1986) J. Mol. Biol. 189, 113–130.
13. Davanloo, P., Rosenberg, A.H., Dunn, J.J. & Studier, F.W. (1984) Proc. Nat. Acad. Sci. USA 81, 2035–2039.
14. Zhang, P. (1989) Ph.D. thesis, Yale University.
15. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y..
16. Zoller, M.J. & Smith, M. (1983) Methods in Enzymol. 100, 468–500.
17. Kunkel, T.A. (1985) Proc. Nat. Acad. Sci. USA 82, 499–492.
18. Kime, M.J. & Moore, P.B. (1983) FEBS Letters 153, 199–203.
19. Roth, K., Kimber, B.J. & Feeney, J. (1980) J. Mag. Res. 41, 302–309.