THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 11, Issue of March 17, pp. 6298–6307, 1995
Printed in U.S.A.

Self-coded 3′-Extension of Run-off Transcripts Produces Aberrant Products during *in Vitro* Transcription with T7 RNA Polymerase*

(Received for publication, September 27, 1994, and in revised form, January 4, 1995)

Francisco J. Triana-Alonso‡§, Marylena Dabrowski‡, Jörg Wadzack‡, and Knud H. Nierhaus‡¶

From the ‡Max-Planck-Institut für Molekulare Genetik, A.G. Ribosomen, Ihnestrasse 73, D-14195 Berlin, Federal Republic of Germany

More than 70% of the RNA synthesized by T7 RNA polymerase during run-off transcription *in vitro* can be incorrect products, up to twice as long as the expected transcripts. Transcriptions with model templates indicate that false transcription is mainly observed when the correct product cannot form stable secondary structures at the 3′-end. Therefore, the following hypothesis is tested: after leaving the DNA template, the polymerase can bind a transcript to the template site and the 3′-end of the transcript to the product site and extend it, if the 3′-end is not part of a stable secondary structure. Indeed, incubation of purified transcripts with the polymerase in transcription conditions triggers a 3′-end prolongation of the RNA. When two RNAs of different lengths are added to the transcription mix, both generate distinct and specific patterns of prolonged RNA products without any interference, demonstrating the self-coding nature of the prolongation process. Furthermore, sequencing of the high molecular weight transcripts demonstrates that their 5′-ends are precisely defined in sequence, whereas the 3′-ends contain size-variable extensions which show complementarity to the correct transcript. Surprisingly, a reduction of the UTP concentration to 0.2–1.0 mM in the presence of 3.5–4.0 mM of the other NTPs leads to faithful transcription and good yields, irrespective of the nucleotide composition of the template.

Run-off transcription *in vitro* using bacteriophage RNA polymerases (Chamberlin and Ryan, 1982) was developed during the last decade as an efficient method for the synthesis of biologically active RNA (Melton *et al.*, 1984; Krieg and Melton, 1984). Among the known variants of this technique the systems based on T7 RNA polymerase are the most widely used. This enzyme can be isolated in large amounts from an overproducing *Escherichia coli* strain (Davanloo *et al.*, 1984), and a number of suitable plasmid vectors containing the strong class III T7 promoters (Dunn and Studier, 1983) or synthetic DNA (Milligan *et al.*, 1987) can be used as templates.

The transition from small scale transcription assays as described in the early reports (Chamberlin *et al.*, 1970; Chamberlin and Ring, 1973) to large scale synthesis of RNA requires an increase in the input of nucleoside triphosphates to the millimolar level (Milligan and Uhlenbeck, 1989; Weitzmann *et al.*, 1990). The optimal concentration of nucleotides can depend on the particular template in use (Milligan and Uhlenbeck, 1989), and it has been shown that an excess of nucleotides (more than 4 mM each) can inhibit the polymerization (Draper *et al.*, 1988). In general, an upper value of 2–4 mM for all four NTPs combined with a magnesium concentration of 14–22 mM gives satisfactory yields.

The fidelity of the *in vitro* transcription process is usually sufficient for low scale synthesis. However, non-coded 3′ prolongation of run-off transcripts has been reported as one or two extra nucleotides in the case of synthetic DNA templates (Milligan *et al.*, 1987), or as very long RNA chains in the case of linearized plasmid templates with 3′ overhanging ends ("snap back" effect; Schenborn and Mierendorf, 1985, Weitzmann *et al.*, 1993).

Using *in vitro* run-off transcription with T7 RNA polymerase, we found that more than 70% of the products could be aberrant transcripts which were longer than the coded RNA (>15 additional nucleotides). We analyzed the reasons for this dramatic mis-synthesis and observed that the processivity of the T7 RNA polymerase is specifically very sensitive to the UTP concentration. Under conditions where the initial concentrations of all four NTPs were the same, a new phenomenon was revealed: after a faithful run-off transcription, the T7 RNA polymerase can accept a former transcript as second template and extend this RNA at the 3′-end in a self-coded fashion, if the 3′-end is not folded into a stable secondary structure. Reduction exclusively of the UTP concentration enables an efficient synthesis of correct products but effectively prevents that of false products.

EXPERIMENTAL PROCEDURES

Materials—Deoxynucleoside triphosphates as well as all the restriction enzymes (except *Ssp*I), calf intestine alkaline phosphatase, T4 polynucleotide kinase, T4 DNA ligase, poly(U)- and oligo(dT)-cellulose were purchased from Boehringer Mannheim. *Ssp*I and the Klenow fragment of *E. coli* DNA polymerase were from New England Biolabs and mung bean nuclease from U. S. Biochemicals. The ribonucleoside triphosphates and the inorganic pyrophosphatase were from Sigma. Plasmid pSP65 and RNasin (40 units/μl) were obtained from Promega Biotec, [^{32}P]UTP, [^{32}P]CTP, [^{32}P]ATP, and [^{32}P]GTP from Amersham. The oligodeoxyribonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer and purified by high performance liquid chromatography on a Hypersil ODS 5-C18 column. Chemical synthesis of RNA was performed on an Applied Biosystems 392 DNA/RNA synthesizer using phosphoramidite chemistry (Scaringe *et al.*, 1990). The plasmid pTZDec-m, a derivative form pTZ18R (Pharmacia) coding for the decacodon O-mRNA (see Table I), was a kind gift from Drs. H.-J. Rheinberger and B. Lewicki, MPI Berlin. The T7 RNA polymerase was isolated from the *E. coli* strain BL21 containing the plasmid pAR1219, according to published procedures (Davanloo *et al.*, 1984).

Construction of Plasmid DNA Templates—Oligodeoxynucleotides were synthesized containing the sequence to be cloned (50–60 nucleotides, see corresponding RNA sequences in Table I). A short complementary oligo(DNA) served as a primer for double-stranded DNA synthesis (Cobianchi and Wilson, 1987). The sequence corresponding to the O-mRNA (Table I) was deleted by successive digestions of pTZDec-m plasmid with *Eco*RI, mung bean nuclease, and *Hpa*I, respectively; and

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Present address: Centro de Investigaciones Biomédicas, Universidad de Carabobo, Núcleo Aragua, La Morita, Maracay, Venezuela.

¶ To whom correspondence should be addressed: MPI Molekulare Genetik, Ihnestr. 73, D-14195 Berlin, Germany.

the 5'-phosphorylated double-stranded sDNAs were inserted into the pTZDec-m plasmid by blunt-end ligation.

The sequence of a $tRNA^{Phe}$ gene from *E. coli* together with a T7 promoter sequence was cloned at the *SmaI* site of the plasmid pSP65. The double-stranded sDNA insert was obtained in the following way. Two oligo(DNA)s were synthesized, the first contained the sequence of the T7 promoter and the sequence corresponding to that of $tRNA^{Phe}$ from nucleotide 1 to 40, the second contained the sequence complementary to that of $tRNA^{Phe}$ from position 22 to 76 (18 nucleotides are complementary to the first oligo) plus a *BstNI* site for linearization. The recombinant plasmids were used for transformations of *E. coli* DH5 according to standard procedures. The positive clones were sequenced in both directions in the region of the insert.

Large Scale Plasmid Preparations—Large scale preparation of plasmid was done by means of a protocol combining alkaline lysis (Birnboim, 1983) and polyethylene glycol precipitation (Lis, 1980), followed by anion-exchange chromatography purification on QIAGEN columns. This combination of methods yields about the same amount and quality of plasmid DNA as the standard protocols (Sambrook *et al.*, 1989) and has the advantage of not requiring the use of RNase A during the preparation, thus avoiding the risk of contamination of the transcription templates. The run-off transcription templates were then prepared by incubation of the purified plasmids with the appropriate restriction enzymes.

In Vitro Transcription Assays—The assays were performed using a set of conditions optimized according to previous reports (Milligan and Uhlenbeck, 1989; Weitzmann *et al.*, 1990). The standard analytical assays were performed in 25 μl of a transcription mix containing 40 mM Tris-HCl, pH 8.0 (37 °C), 22 mM $MgCl_2$, 1 mM spermidine, 5 mM dithioerythitol, 100 μg/ml of bovine serum albumin (RNase and DNase free), 1 unit/μl ribonuclease inhibitor (RNasin), 5 units/ml of inorganic pyrophosphatase, 0.5 pmol (~1 μg) of linearized plasmid DNA template (or 1 pmol of synthetic DNA templates), 3.75 mM (or the indicated values) each of ATP, GTP, CTP, and UTP, and 40 μg/ml (or the indicated amount) of purified T7 RNA polymerase. The components were mixed at room temperature and incubated at 37 °C for the times indicated (3 h standard). The transcription products were analyzed by cold trichloroacetic acid precipitation in order to estimate the extent of reaction (when a ^{32}P label was included), and by electrophoresis in denaturing polyacrylamide gels to determine the quality of the RNA synthesized. The electrophoretical patterns were visualized by autoradiography and/or via staining with ethidium bromide or toluidine blue. The synthetic DNA templates were assembled by annealing a 17-mer oligodeoxynucleotide with the sequence of the region −17 to −1 of the consensus class III T7 promoter to a template DNA strand containing complementary sequence of the promoter followed by that of the RNA to be synthesized. In assays where purified RNA was used as sole template, it was added in amounts ranging from 0.25 to 20-fold molar excess over polymerase.

Transcript Purification—The transcription products of the plasmids pTZDec-m, pTDMS, and pTDLS (linearized with *BamHI*) were purified after phenol-chloroform extraction of the transcription mix by chromatography on oligo(dT)-cellulose (Sambrook *et al.*, 1989) followed by ion-exchange chromatography on QIAGEN columns. Transcripts obtained from the pTMF template (linearized with *SspI*) or synthetic DNA templates were purified by polyacrylamide gel electrophoresis under denaturing conditions.

RNA Sequencing—The run-off transcript of the plasmid pTZDec-m was sequenced with the chain termination method (Hahn *et al.*, 1989) using avian myeloblastosis virus reverse transcriptase (from Pharmacia-PL Biochemicals) and a synthetic DNA primer. The transcripts derived from the pTMF template were directly labeled at the 5'-end by including [$\gamma^{32}P$]GTP in the *in vitro* transcription reaction. The sequences were determined with the enzymatic sequencing method using the RNA Sequencing Kit from Pharmacia-PL Biochemicals.

The sequencing of the 27-mer product obtained from transcription of MF-mRNA was performed using the phosphorothioate method of Gish and Eckstein (1988) with slight modifications as described in Schatz *et al.* (1991). Five transcription reactions were performed in parallel, in four of which 20% of one nucleotide was replaced by the corresponding [αS]NTP. The fifth reaction was lacking [αS]NTPs, and the corresponding product was used as control for detecting spontaneous chain disruptions. The transcription reactions were performed under standard conditions (see above) in a total volume of 50 μl containing 200 pmol of MF-mRNA as template and 20 pmol of T7 polymerase. The transcription products were separated on a polyacrylamide gel, the bands visualized by UV shadowing, and the transcripts corresponding to the "27-mer" band were extracted. After 5'-labeling, 10 μl (~200,000 disintegrations/min) were treated with 1 μl of a 10 mM J_2 solution in ethanol, incubated for 1 min at room temperature, precipitated with ethanol, and analyzed by polyacrylamide gel electrophoresis.

Secondary Structure Predictions—The secondary structure of minimum free energy for the synthetic RNA molecules was determined with the program FOLD (Zuker and Stiegler, 1981) using the energy parameters defined by Turner and co-workers (Freier *et al.*, 1986), and/or the program MFOLD (Zuker, M., 1989).

RESULTS

Transcription under Standard Conditions—The startling observation was derived from analytical *in vitro* transcription assays performed with templates designed for the synthesis of small model mRNAs devoid of strong secondary structure (see Table I). Under standard conditions (all NTPs at 3.75 mM), the products obtained with the plasmid templates revealed a surprising phenomenon. In the cases of pTZDec-m, pTDMS and pTDLS (decacodon template series) the calculated transcription efficiency (RNA copies synthesized per template) suggested more than quantitative utilization of the ATP added to the assay, if one assumes exclusive synthesis of the expected transcript; the calculated ATP consumption was 12–15% higher than the total input of non-labeled ATP (the label was incorporated as an [^{32}P]NTP different from ATP in these cases).

The gel analysis of these transcripts showed a second anomaly which might be the reason for the mis-estimate. The RNA synthesized ran in the sequencing gel as a smear starting at the position of the full-length transcript and going to higher molecular weights (Fig. 1*A*, *lanes 4–6*). The track corresponding to the MF-mRNA transcript (Fig. 1*A*, *lane 7*) also showed a smear ranging from the expected 46-mer to the region corresponding to about 75 nucleotides. The RNA obtained from the synthetic DNA templates similarly showed a non-homogeneous electrophoretic pattern (Fig. 1*C*, *lane 2*). It is clear that the assumption of exact transcription was not justified. Since the nucleotide composition of the longer transcripts is not known, the transcription efficiency calculation based on one radioactive nucleotide is not reliable. In contrast, the unmodified $tRNA^{Phe}$ transcript consisted of a single product of the right size (Fig. 1*C*, *lane 4*). At first glance it appeared that the partial homopolymeric nature of the model mRNAs could be the cause of the aberrant transcription as indicated in previous reports of "slippage" of T7 RNA polymerase on poly(A) tracks (Milligan and Uhlenbeck, 1989). However, a second series of templates with balanced base composition and lacking homopolymeric tracks gave similar results (data not shown).

Kinetics of the MF-mRNA synthesis were performed by subjecting the samples withdrawn from the transcription assay at various times to gel electrophoresis and analyzing the resulting autoradiogram by densitometry (Fig. 1*B*). At the onset of incubation predominantly correct products were synthesized, as evidenced by the main band product with the expected 46 nucleotides. However, the longer the incubation for transcription lasts the more false, overlong products are produced, reaching a comparable level with the expected product after 180 min of incubation (Fig. 1*B*). The apparent lag period between the incubation start and the efficient synthesis of high molecular weight products indicates that the synthesis of longer RNA is a secondary process which needs an induction time. This induction time appears to be inversely related to the speed of RNA synthesis (*i.e.* with more enzyme the lag period is smaller, data not shown) and seems to be different for every template. These observations suggest that a minimal concentration of transcripts is required for the onset of aberrant transcription.

Secondary Structure at the 3'-Ends of Products and Aberrant Transcription—The observation that some templates reproduc-

TABLE I
Structure of the various transcripts

Decacodon O-mRNA, model mRNA for *E. coli* systems; the main structural elements are Shine-Dalgarno element (in bold face) with a 7-nucleotide long spacer to the initiation codon (underlined), a coding region containing 10 unique codons followed by a stop codon, and a poly(A) tail (57-nt long). The two additional variants of the decacodon were designed to have a medium stability (decacodon MS-mRNA) and low stability (decacodon LS-mRNA) in their predicted secondary structures (according to the program FOLD) when compared with the original mRNA (decacodon O-mRNA). The MF-mRNA contains two unique codons for methionine (M) and phenylalanine (F). The 14-mRNA contains a Shine-Dalgarno element, an 8-nt spacer, and a coding region composed of 14 codons.

Template	Transcript	RNA sequence
Plasmid templates		
		1 10 20 30 40
pTZDec-m	Decacodon	GGGAAUUCGG·ACACGGUGAC·CAGAGCCCUC·**AAAGGAGGUU**·UGAAGU
(*Bam*HI digested)	O-mRNA	AUGA·CUUCAAUCUU·CGCUCAGGUU·AACGACUAA·(A)57·GGGG·AUC
	(143 nt)	50 60 70 140
		1 10 20 30 40
pTDMS	Decacodon	GGGCCC**GGAG·GU**UUGAAGUA·UGGCUUCUAU·CGUUUUCCAG·
(*Bam*HI digested)	MS-mRNA	ACCGAACGUU·AACUUAAGCA·ACGACUAA·(A)57·GGGGA·UC
	(132 nt)	50 60 130
		1 10 20 30 40
pTDLS	Decacodon	GGGCCC**GGAG**·UUUAAAAUAU·GCAAUCUAUC·UUCACCGUUU·
(*Bam*HI digested)	LS-mRNA	ACGAACGCUA·AAGAUCUAAA·CGACUAA·(A)57·GGGGAU·C
	(131 nt)	50 60 130
	Decacodon	1 10 20 30 40
pTDLS	LS-mRNA	GGGCCC**GGAG**·UUUAAAAUAU·GCAAUCUAUC·UUCACCGUUU·
(*Bgl*II digested)	(short-form)	ACGAACGCUA·AAGAUC
	(56 nt)	50
		1 10 20 30 40
pTMF	MF-mRNA	GGGAAAAGAA·AAGAAAAGAA·AAUGUUCAAA·AGAAAAGAAA·AGAAAU
(*Ssp*I digested)	(46 nt)	
		1 10 20 30 40 50
pSTtPhe	$tRNA^{Phe}$	GCCCGGAUAG·CUCAGUCGGU·AGAGCAGGGG·AUUGAAAAUC·CCCGUGUCCU·
(*Bst*NI digested)	(unmodified)	UGGUUCGAUU·CCGAGUCCGG·GCACCA
	(76 nt)	60 70
Synthetic DNA templates		
MF-mRNA template + T7 promoter	MF-mRNA	See above
		1 10 20 30 40
14-mRNA template	14-mRNA	GGGACC**AAGG·AGG**AAAAGAA·AAUGAAAGCA·ACCGCAACCG·CAACU
+ T7 promoter	(65 nt)	AUGAU·AUCAAUAAUA·AUAAU
		50 65

ibly yielded exact products whereas others provided large amounts of overlong products was a surprise. A possible reason could be seen in the fact that the transcripts systematically differed in their capability to form secondary structure at the 3′-end region. For example, unmodified $tRNA^{Phe}$ is obviously able to form the known secondary structures of the wild type tRNA since the transcript was functional in aminoacylation assays and in poly(Phe) synthesis with poly(U) programmed 70 S ribosomes from *E. coli* (data not shown). It follows that the 3′-end of the $tRNA^{Phe}$ transcript should have only four nucleotides which are not involved in secondary structure. In contrast, the other transcripts listed in Table I (except the short form of the decacodon LS-mRNA) show a low overall tendency to form double helices and their 3′-end regions are not involved in stable secondary structures (as predicted with the programs FOLD and MFOLD; data not shown).

The template for the decacodon LS-mRNA offered a convenient system to test the importance of secondary structure at the 3′-end of the correct transcript with regard to the quality of transcription. Upon restriction with *Bam*HI, this plasmid serves as a template for the synthesis of the long form of the decacodon LS-mRNA, which contains a poly(A) tail of 59 adenyl residues and is thus free of secondary structure at the 3′-end (Fig. 2*A*). When restricted with *Bgl*II the plasmid codes a short form of the mRNA which lacks the poly(A) tail and carries a secondary structure motif at the 3′-end (Fig. 2*B*). The results of parallel transcription assays with both forms of the template were striking: products with 3′-end secondary structure show one main product corresponding to the expected length practically free of longer transcripts (Fig. 2*B*), whereas large amounts (~60% of the toluidine blue-stainable material) of prolonged false transcripts were found with the RNA containing the poly(A) tail (Fig. 2*A*).

One possible explanation for the appearance of high molecular weight products is that after running off the T7 RNA polymerase can rebind the previously synthesized RNA and continue the RNA synthesis thus making longer transcripts. A prerequisite for an RNA to function in that way is a loose 3′-end free of secondary structure. It is not yet clear at this point whether one RNA molecule can serve as template for the prolongation of another RNA molecule or the template RNA is prolonged at its own 3′-end (self-coding; Fig. 3*C*).

Self-coded Transcript Extension—The possibility of using RNA as template for the DNA dependent T7 RNA polymerase has in fact already been described by Chamberlin and Ring (1973) and Kornaska and Sharp (1989), among others. We confirmed that poly(U) could indeed be used as template, yielding a rate of polymerization of 0.3–0.5 nucleotides/s and enzyme (at saturation; data not shown), which is about 100-fold slower than in the presence of a DNA template under the conditions used. Furthermore, the fact that for detectable transcription the concentration of RNA template must be approximately 20 times higher than that of DNA templates indicates a higher affinity of the enzyme for DNA templates.

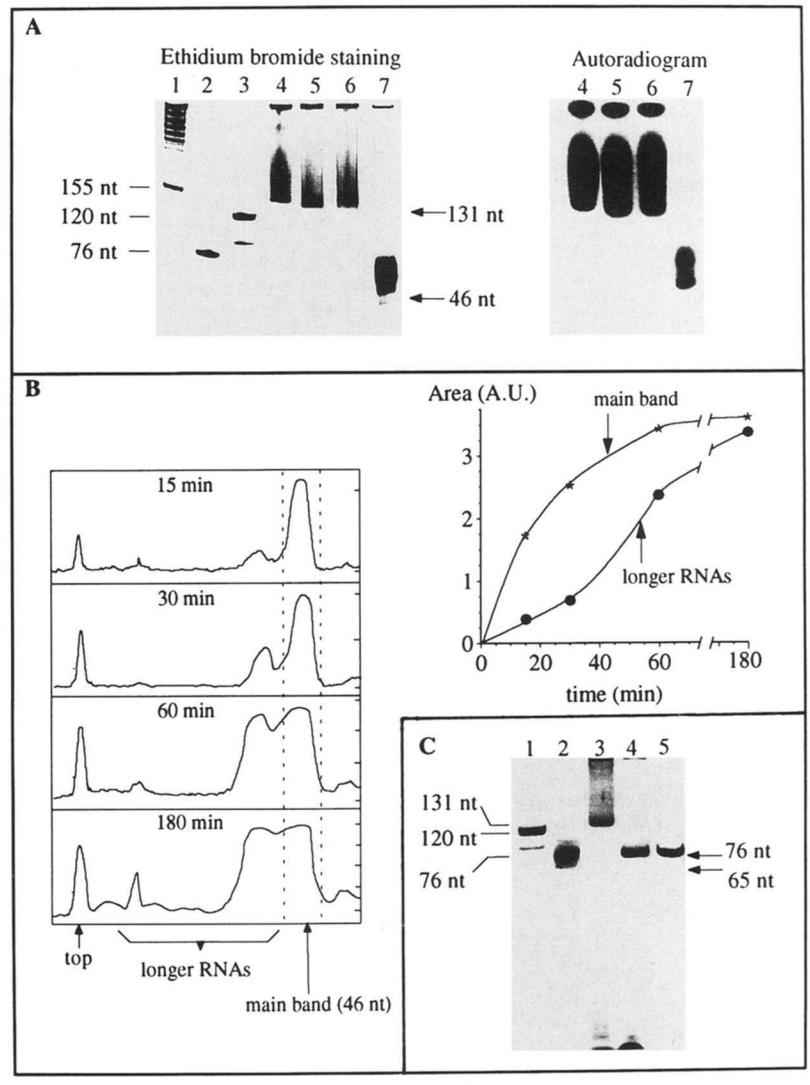

FIG. 1. **Gel electrophoresis analysis of different *in vitro* transcription products.** 4-μl samples from the indicated standard transcription assays were applied to appropriate sequencing gels. After separation the gels were stained with toluidine blue or ethidium bromide and, when indicated, exposed for 12–48 h at −80 °C to an x-ray film. *A*, 10% sequencing gel, ethidium bromide staining: *lane 1*, 0.16–1.77-kilobase RNA molecular weight standards; *lane 2*, $tRNA^{Phe}$ (0.02 A_{260} units); *lane 3*, 5 S rRNA (0.02 A_{260} units); *lane 4*, pTZDec-m transcript; *lane 5*, pTDMS transcript; *lane 6*, pTDLS transcript; *lane 7*, pTMF transcript. Autoradiogram (*lanes 4–7*). *B*, scans of an autoradiogram. 1 pmol (2 μg) of *Ssp*I-restricted pTMF was incubated at 37 °C in 50 μl of standard transcription mix containing [^{32}P]ATP (30 cpm/pmol) with T7 RNA polymerase at a final concentration of 4 μg/ml. At the indicated times (0, 15, 30, 60, and 180 min), 5-μl samples were taken, frozen in liquid nitrogen, and dried under vacuum. The material was then dissolved in 5 μl of RNA sample buffer and run in a 15% sequencing gel. The quality of the RNA synthesized was analyzed by densitometry of the corresponding autoradiograms (*B*, *left panel*). The area in arbitrary units, (*A.U.*) from the peaks corresponding to the main band (*) and the longer RNAs, ●, respectively, are plotted against the time of incubation (*B*, *right*). *C*, 15% polyacrylamide-urea gel. Toluidine blue staining: *lane 1*, 5 S rRNA (0.02 A_{260} units); *lane 2*, 14-mRNA transcript (65 nt), *lane 3*, pTDLS transcript; *lane 4*, pSTtPhe transcript (76 nt); *lane 5*, $tRNA^{Phe}$ from *E. coli* (0.03 A_{260} units).

When a small RNA practically lacking predictable secondary structure was used as template (MF-mRNA, for sequence see Table I) the polymerization was also efficient (up to 1600 nt[1] polymerized/μg of T7 RNA polymerase in standard conditions, data not shown). The products of a transcription incubation including [^{32}P]UTP as label were analyzed by gel electrophoresis. The corresponding autoradiogram (Fig. 3*A*, *lanes 1–3*) showed no products with the length of the starting template (46 nt) but rather a group of short aberrant products 5–26 nucleotides long (about 20% of the synthesized RNA); a strong band corresponding to a 27–28-mer (~40% of the total RNA) is also present, suggesting that the RNA is mainly used as template at about the middle of the sequence, and relatively large amounts of long transcripts (~40%) also appear which are longer than the starting template. Similar results with respect to nucleotide incorporation and size distribution of the transcripts were obtained using chemically synthesized RNAs having totally unrelated sequences but preserving the size and the lack of predictable secondary structure as common features (data not shown).

In order to determine the sequence of the strong band with about 27 nt, the transcription was repeated, but now allowing for random incorporation of a few phosphorothioated nucleotides at either A, G, C, or U positions. The oligonucleotides corresponding to the 27-mer were isolated, labeled at their 3′-ends with ^{32}P, and then cleaved with iodine at the phosphorothioated positions and separated on a sequence gel (Gish and Eckstein, 1988). Fig. 4 demonstrates that the oligonucleotides were complementary to the 5′-end of the MF-mRNA. The prolongation of the transcripts exceeded the 23 nt identified on the sequence gel, since the last one (*3′-C*, see Fig. 4) and the first 2–3 nucleotides (*AAG-5′*) could not be detected for technical reasons. For convenience we call the respective band the "27-nt" product. The fact that a clean sequence pattern is obtained indicates that the 27 nt band contains one defined transcript. It is therefore clear that the 27-nt oligomer is the product of a defined *de novo* synthesis starting at a position which is about 19 nucleotides from the 3′-end of the template MF-mRNA.

The fact that the addition of the non-labeled MF-mRNA yields two more or less defined product families (Fig. 3*A*), one shorter (about 27 nt) and one longer (50–80 nt) than the added MF-mRNA with 46 nt, could be explained as follows. The MF-mRNA binds around its 27th nucleotide to the template site of the T7 RNA polymerase. Then either *de novo* synthesis (Fig. 3*C*, *left*) occurs yielding the 27-nt long products; in other cases the polymerase prolongs added MF-mRNAs for about 27 nucleotides yielding the 75-nt products. It is possible that a template mRNA serves to prolong its own 3′-end at the later process.

In the next experiment we wanted to analyze the fate of the input MF-mRNA, which therefore carries a ^{32}P label at the 5′-end. The transcription was performed with non-labeled NTPs (Fig. 3*A*, *lanes 5–9*). The radioactivity was distributed

[1] The abbreviations used are: nt, nucleotide(s); cpm, counts/min.

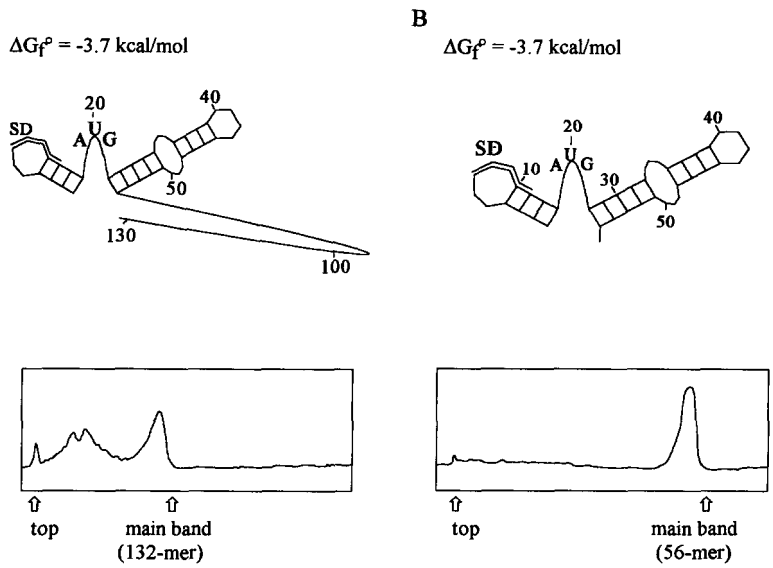

FIG. 2. **Predicted secondary structures and *in vitro* transcription with two forms of the pTDLS template.** The optimal predicted secondary structures and their ΔG_f^o were determined for the long (*A*) and the short (*B*) forms of the decacodon LS-mRNA (see Table I) using the program MFOLD (*S.D.* = Shine-Dalgarno sequence). The scans of autoradiograms correspond to analytical assays where 0.5 pmol of pTDLS digested with *Bam*HI (*A*) or *Bgl*II (*B*) were incubated in standard transcription conditions during 3 h at 37 °C. 5-μl samples of the transcription mix were applied to a 10% sequencing gel, and after separation the nucleic acids were stained with toluidine blue.

between the original mRNA band (see *lane 5*) and a smear of overlong products of up to 80 nt comprising about 60% of the original input (determined by densitometry of the autoradiogram and by direct counting of gel slices) and showing a pattern of the main products (Fig. 3*A*, *lane 6*) very similar to that observed when a pTMF DNA template is used. These results again indicate that the T7 RNA polymerase uses the RNA as transcription template and is also able to prolong this RNA at the 3'-end. Interestingly, the input of RNA template, which varied from 10- to 0.5-fold molar excess over polymerase, had practically no effect on the pattern of 3'-end prolonged products (Fig. 3*A*, *lanes 6–9*). The fact that the same pattern is seen even at the lowest input of MF-mRNA is in favor of the view that the prolongation process is a self-coding mechanism (Fig. 3*C*, *right*).

A rigorous test of this view is made in the next experiment. [^{32}P]MF-mRNA (46 nt) is mixed with various amounts of nonlabeled decacodon O-mRNA (143 nt). Both mRNAs can trigger different patterns of overlong products in a transcription assay. A self-coding mechanism (Fig. 3*C*, *right*) will show in the mixture experiment the two distinct patterns without any interference, whereas the possibility that "one RNA molecule serves as template for the prolongation of a second RNA molecule" (intermolecular mechanism; Fig. 3*C*, *center*) implies a significant change in the pattern of radioactive prolonged RNAs. Fig. 3*B*, *lanes 1–4*, shows that increasing amounts of decacodon O-mRNA do not at all interfere or disturb the distribution of the overlong products derived from MF-mRNA. Ethidium bromide staining of a second sequencing gel reveals that the decacodon O-mRNA also triggered its distinct pattern of prolonged RNA (Fig. 3*B*, *lanes 7–10*) independently of the presence of MF-mRNA. It follows that the 3'-end prolongation is most probably an intramolecularly coded, *i.e.* a self-coded process.

Sequencing the Products—According to the conclusions of the preceding section the extended transcripts must all have a defined 5'-end but various 3' extensions, which should be complementary to the RNA template. This expectation was checked by sequencing the transcripts produced in a preparative assay (3.75 mM NTPs were used in these cases) and purified as described under "Experimental Procedures." The DNA template for the MF-mRNA was selected for this purpose.

After transcription under standard conditions, using [^{32}P]GTP as label and pTMF (*Ssp*I restricted) as template, the full-length transcript and one sample from the area containing products with a higher molecular weight were isolated after polyacrylamide gel electrophoresis (Fig. 5*A*) and sequenced using enzymatic techniques. The sequencing patterns (Fig. 5, *B* and *C*) confirmed the expected sequence of the full-length transcript and revealed that the secondary bands (in a range of 50–70 nucleotides) contain a 5'-end with a sequence identical to the correct transcript (Fig. 5*B*) and 3' prolongations with apparent complementarity to the correct transcript (Fig. 5, *C* and *D*). It follows that the aberrant transcripts are not the result of false transcription initiation or errors in the elongation process but rather of an extended synthesis (Fig. 5*D*). Sequencing of the heterogeneous transcripts from the template for the O-mRNA (Table I and Fig. 1*A*, *lane 4*) also revealed defined 5'-ends expected for the correct transcript (data not shown).

What Avoids Self-coded 3'-End Prolongation of Transcripts?—We have identified three parameters which influence the extent of false transcription. One parameter already mentioned is the time of transcription incubation. At the beginning predominantly correct products are produced, whereas false transcripts accumulate at the end of long incubations (Fig. 1*B*).

A second parameter is the concentration of T7 RNA polymerase. With the template for the decacodon LS-mRNA used in the experiment shown in Fig. 6, one should perform the reaction with an enzyme concentration of less than 10 μg/ml to avoid most of the smear of higher molecular weight products. However, adjusting the enzyme concentration for every individual template in order to achieve precise transcription is not always advisable, since for a number of templates the enzyme concentration had to be reduced below 4 μg/ml, which is too low for preparative purposes.

The third parameter was a surprise. A strong and specific effect of the UTP concentration on the accuracy of transcription was found. Fig. 7*A* shows the transcription of the MF-mRNA (46 nucleotides) under conditions where the concentration of the nucleotides was systematically changed (one by one) in a range covering preparative and analytical conditions (3.75 mM to 188 μM). The autoradiogram reveals that a reduction of ATP or GTP lowers the efficiency of transcription with apparently no effect on the proportion of correct/incorrect products. The reduction of CTP seems to have no effect in the range of concentrations used, probably due to the fact that the transcript has only 1 cytosine. In sharp contrast, reduction of UTP predominantly reduces the synthesis of incorrect products. When we evaluated the gels quantitatively (Fig. 7*B*), the following results were obtained. At the lowest UTP concentration applied

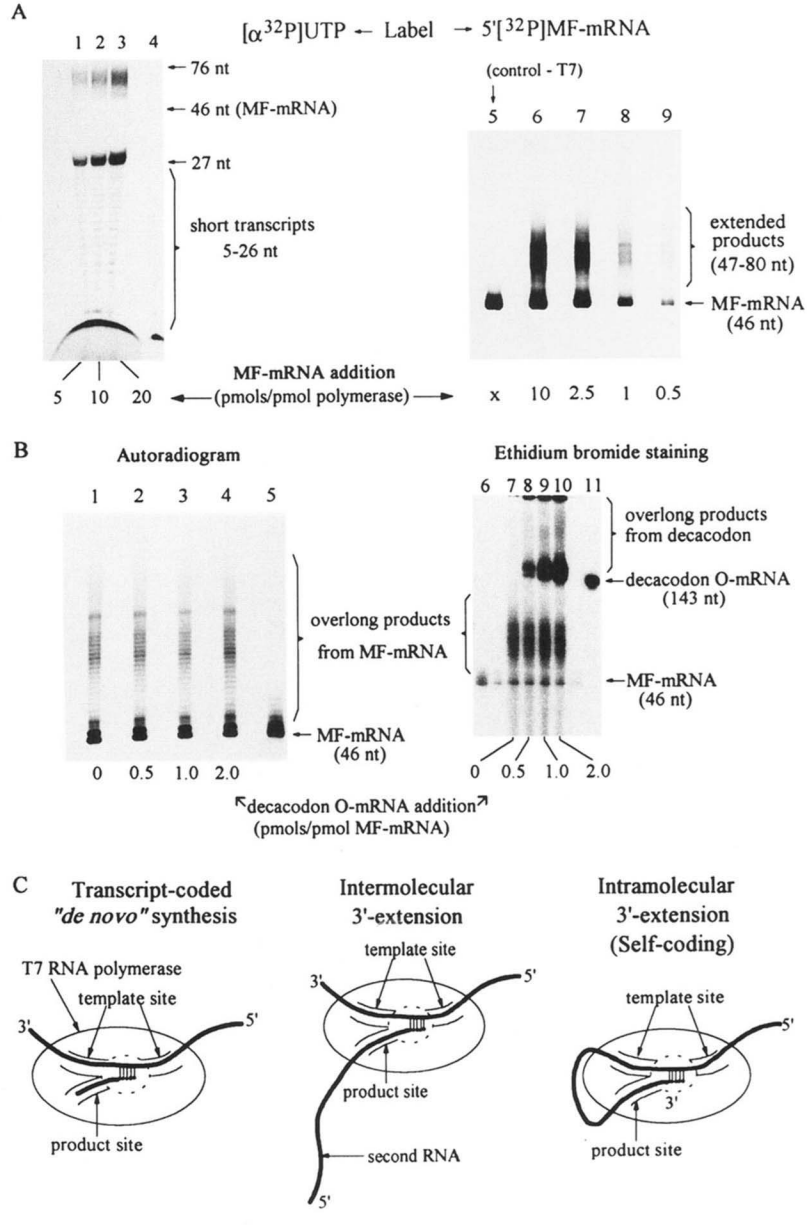

FIG. 3. **RNA dependent *in vitro* transcription with T7 RNA polymerase, MF-mRNA as template.** *A*, autoradiograms from 15% sequencing gels. *Left panel*, *lanes 1–3*: 125, 250, and 500 pmol of non-labeled MF-mRNA, respectively, were incubated in 25 μl of standard transcription mix containing 3.75 mM of each ATP, GTP, CTP, and [^{32}P]UTP (20 cpm/pmol) and 25 pmol of T7 RNA polymerase during 4 h at 37 °C. *Lane 4* corresponds to a control incubation omitting the MF-mRNA. *Right panel*, *lanes 6–9*: 250, 62, 25, and 12.5 pmol of 5'-^{32}P-labeled MF-mRNA (390 cpm/pmol), respectively, were incubated as indicated above in the presence of non-labeled NTPs. *Lane 5* corresponds to a control incubation containing 25 pmol of [^{32}P]MF-mRNA and omitting the T7 RNA polymerase. *B*, *left panel*, autoradiogram from 10% sequencing gels. *Lanes 1–4*, 100 pmol of 5'-^{32}P-labeled MF-mRNA (260 cpm/pmol) were incubated in standard transcription conditions with non-labeled NTPs, 100 pmol of T7 RNA polymerase, and 0 (no addition), 50, 100, and 200 pmol of gel-purified non-labeled decacodon O-mRNA, respectively. *Lane 5* corresponds to a control incubation omitting NTPs and decacodon O-mRNA. *Right panel*, ethidium bromide staining. *Lane 6* corresponds to *lane 5*, and *lanes 7–10* correspond to *lanes 1–4* in the autoradiogram (*left panel*); *lane 11*, untreated decacodon O-mRNA. *C*, three models for RNA-coded RNA synthesis.

(188 μM) the most accurate transcription was observed and the correct transcripts amounted to about 80% of the total transcribed products, whereas the maximal amount of correct products (100%) was found at 0.7 mM UTP. Variation of the CTP concentration did not affect the accuracy of transcription; only 25% of the total transcripts were correct, whereas the yield of correct transcripts showed a shallow maximum at 0.7 mM CTP reaching 86% of the best yields found in the UTP series. Reduction of the ATP or GTP had a disastrous effect. The yield of correct transcripts went down to 3%, whereas the synthesis of incorrect transcripts increased to over 90%.

Qualitatively, the same results were observed with various plasmid templates and synthetic DNA templates coding for transcripts with weak secondary structure at the 3'-end and with different nucleotide compositions For example, the oligo(DNA) containing the fragment of a tRNA gene corresponding to positions 1–40 of $tRNA^{Phe}$ plus the T7 promoter was transcribed. The transcription yielded mainly overlong products of about 60 nt. The prolongation complementary to the first 20 nucleotides of $tRNA^{Phe}$ contained only 3 U residues. Nevertheless, a reduction of the UTP concentration during transcription grossly prevented the synthesis of overlong transcripts (data not shown). It follows that the strong and specific effect observed with decreasing the UTP concentration is not related to the uridine content or nucleotide composition of the transcripts.

DISCUSSION

The run-off *in vitro* transcription system with T7 RNA polymerase has been used for the synthesis of short (less than 100 nucleotides; Groebe and Uhlenbeck, 1988; Sampson and Uhlenbeck, 1988; Milligan and Uhlenbeck, 1989) and long RNAs (*e.g.* 16 S and 23 S rRNA; Nègre *et al.*, 1989; Weitzmann *et al.*, 1990). The transition from small scale to large scale assays required some changes of the transcription systems to improve the yield of synthetic RNA. Major changes are the inclusions of ribonuclease inhibitor and inorganic pyrophosphatase, which significantly improve the quality and the yield of products (Cunningham and Ofengand, 1990); minor improvements are the additions of polyethyleneglycol and Triton X-100 (Milligan and Uhlenbeck, 1987). Furthermore, the concentrations of nucleotides have to be increased, and it is normally assumed that an increase in the nucleotide concentration has no major effect on the quality of the RNA synthesized.

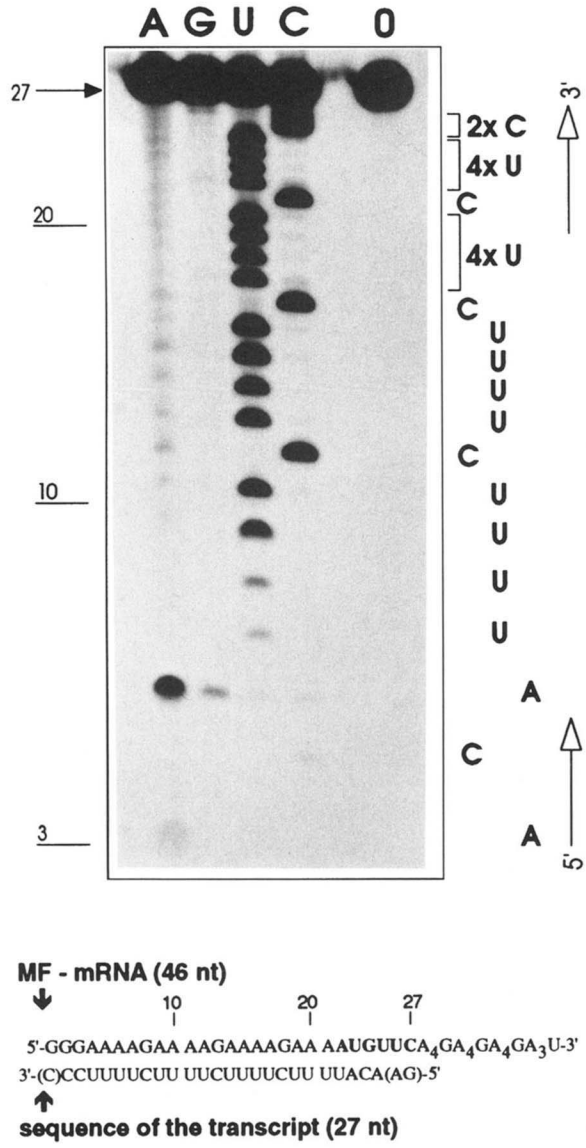

FIG. 4. **The sequence of the short 27-nt transcript of the MF-mRNA (46 nt).** The sequence was assessed with the phosphorothioate method (Gish and Eckstein, 1988). *A*, *G*, *U*, and *C*, transcripts containing the respective phosphorothioated nucleotides. *O*, transcript without phosphorothioated nucleotides but otherwise treated as the thioated transcripts. *Bottom*, the sequence of the transcript is compared to that of the MF-mRNA used as a template. For details see "Experimental Procedures."

However, it is frequently seen that the synthesis of full-length transcript is accompanied by production of higher and lower molecular weight products during large scale transcription assays using T7 and also other bacteriophage RNA polymerases (Milligan *et al.*, 1987, Ling *et al.*, 1989, Paddock, 1989).

The transcription system previously optimized by us for the synthesis of a model mRNA coding for a decapeptide (see "Experimental Procedures") was used to test the templates for different heteropolymeric mRNAs. The system works efficiently with respect to the incorporation of ribonucleotides in RNA, but the fidelity of the synthesis is far from optimal. In all cases presented in Fig. 1, the synthesis of the expected transcripts was accompanied by a significant production of RNA with higher molecular weights than expected, the only exception being the transcription with the $tRNA^{Phe}$ template. Since all the plasmid templates used in this experiment had 5′-protruding or blunt ends, the observed phenomenon is different from the snap back effect observed with DNA templates having 3′-protruding ends. This term describes the fact that an RNA polymerase engaged in transcription can, upon reaching the 5′-end of the template strand, turn to use the 3′-protruding end as a second template (a DNA to DNA jump) continuing the synthesis toward plasmid-sized transcripts (Schenborn and Mierendorf, 1985). Furthermore, the RNA extensions described in this paper were also produced with single-stranded DNA templates, which cannot trigger a snap back event. The template-independent addition of a few nucleotides at the 3′-end of small transcripts observed by Uhlenbeck and colleagues (Milligan *et al.*, 1987) can also not explain the results, since the additions found in the cases reported here are much longer.

Several observations reported in the literature as well as the results presented here led to a possible explanation for the aberrant transcription. First, it is known that the T7 RNA polymerase (as well as many other RNA polymerases) can accept RNA as template (Chamberlin and Ring, 1973; Kornaska and Sharp, 1989, 1990). Second, as shown in Fig. 1*B*, the production of high molecular weight RNA appears to be a late event during the transcription incubation. Third, in all cases where the phenomenon is visible the expected transcript has a 3′-end region which appears to be free from stable secondary structure (Fig. 2). Fourth, the presence of two different RNAs both with 3′-ends lacking stable secondary structures triggers a pattern of overlong products independently from each other. And finally, sequencing of the high molecular weight RNA revealed that it consists of a 5′ portion identical to the expected transcript plus a variable 3′ prolongation, which shows apparent complementarity with the transcript itself (Fig. 3*D*). It follows that the synthesis of aberrant transcripts is not related to abnormal transcription initiation. The problem starts when the "run-off" process is completed.

The data suggest the following. During the run-off transcription and after the production of a certain amount of transcript, the RNA can serve as substrate for a self-coded 3′-end prolongation. A certain amount of transcript is required because the RNA has to compete with the DNA for the template site, which has a higher affinity for DNA. This explains the lag of appearance of false transcripts in the kinetics shown in Fig. 1*B*. In order to be fully efficient in this process, the RNA must have a 3′-end free of secondary structures. This means that at sufficient concentrations of RNA the T7 RNA polymerase swap from a DNA to an RNA template. A stable secondary structure at the 3′-end of a transcript obviously prevents the binding to the *product* site and possibly also to the *template* site of the polymerase (Figs. 2 and 3*C*). Since a secondary structure involving the 3′-end will form only after the release of the respective transcript, the RNA has probably to be released before it can occupy the *template* site and with its 3′-end the *product* site of the T7 RNA polymerase. After the self-coding process, the overlong products have now 3′-ends which are self complementary, are therefore involved in stable secondary structures, and thus cannot serve as template again preventing a further prolongation. This explains why overlong products have a length of less than twice that of the correct transcript. Our observations concerning the importance of a secondary structure at the 3′-end of a transcript agree well to what is known about the requirements for termination. T7 RNA polymerase terminates efficiently *in vitro* at a G:C rich hairpin/U stretch structure (Dunn and Studier, 1983). The hairpin formation is important for disrupting DNA (template)-RNA (transcript) interactions as well as interactions between RNA and polymerase (Sousa *et al.*, 1992, and references therein).

Three parameters of the transcription assay were shown to affect the relative amount of aberrant transcription: the time of transcription incubation (Fig. 1*B*), the stoichiometry of T7 RNA

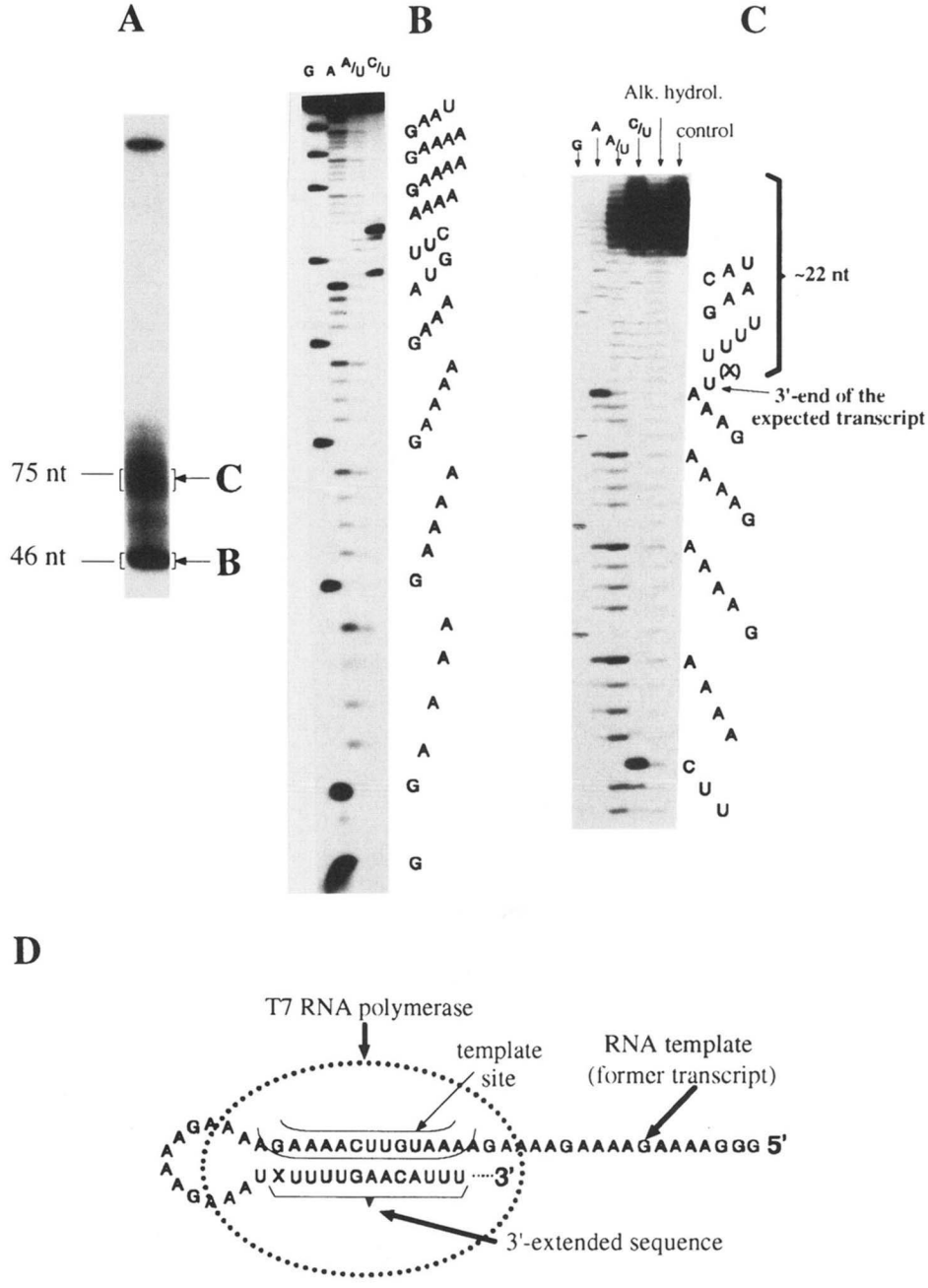

FIG. 5. **The sequence of the T7 transcripts is prolonged at the 3′-end.** 5 pmol of the MF-mRNA DNA template (pTMF) were incubated in standard conditions (all NTPs at 3.75 mM) in the presence of [^{32}P]GTP (40 cpm/pmol). The transcription products were then separated in a 17% sequencing gel (autoradiogram in *A*) from which two sections were cut and the RNA extracted. One section corresponded to the position of the expected transcript (46 nt; *B*) and the other corresponded to an RNA about 20 nt longer (*C*). The sequences of these two RNAs were determined enzymatically showing that the 46-nt long RNA corresponds to the expected transcript (*B*), and the longer fragment contains an identical 5′ region followed by a 3′ prolongation (*C*) which shows complementarity to the expected transcript. *D*, self-coding model showing the complementarity between the MF-mRNA sequence and the 3′ extension.

polymerase used relative to the DNA template (Fig. 6), and the concentrations of nucleotides. Fig. 6 demonstrates that, when the amount of enzyme surpasses a critical level, the synthesis of "smeared" products with a high molecular weight is exacerbated. It follows that an adjustment of the enzyme amount could be sufficient for a synthesis with tolerable amounts of secondary products. However, the enzyme reduction is not a practical solution in general, since the enzyme concentration must be reduced to inefficient low levels for many templates. Since higher concentration of polymerase leads to enhanced synthesis resulting in high concentration of RNA in the transcription assay, a transcript-coded transcription is favored under these conditions. Thus, lowering the polymerase concentration improves the transcription quality not by influencing the mode of action of the polymerase, but rather by lowering the synthesis of a possible RNA template and therefore does not provide a real improvement. According to the results in Fig. 7, manipulation of the nucleotide concentrations dramatically affects the production of high molecular weight RNA. The reduction of purine nucleotides and CTP caused a decrease in the yield of transcripts without improving the fidelity of transcription. In contrast, when the UTP concentration was decreased, a strong reduction in the synthesis of high molecular weight RNA was observed without a severe effect on the yield of the expected transcript. The same phenomenon was observed with different templates coding for RNAs with unrelated sequences and nucleotide composition (with or without homopolymeric tracks) but all preserving the lack of secondary structure at the 3′-end.

Since there is no clear correlation between the composition of the transcripts and the effects of UTP, the data suggest that at high concentrations the UTP is an allosteric effector of the T7 RNA polymerase inducing a conformation which facilitates the acceptance of RNA as a template. This hypothesis is still speculative but has some support from the following observations. In analytical assays using micromolar concentrations of all four nucleotides, it has been shown that a reduction of the pyrimidines, especially the UTP, triggers an increased production of

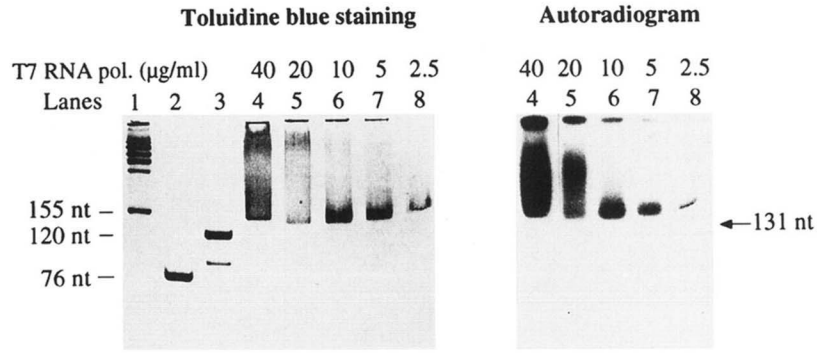

FIG. 6. **The concentration of T7 RNA polymerase affects the quality of the transcripts.** 1 μg (0.5 pmol) samples of the decacodon LS-mRNA DNA template (pTDLS) were incubated in 25 μl of standard transcription mix containing [^{32}P]ATP (2 cpm/pmol) with the indicated amounts of T7 RNA polymerase (2.5–40 μg/ml) during 3.5 h at 37 °C. Samples of 6 μl from every incubation were applied to a 10% sequencing gel. *Lane 1*, 0.16–1.77-kilobase RNA molecular weight standards; *lane 2*, tRNA^{Phe} (0.02 A_{260} units); *lane 3*, 5 S rRNA (0.02 A_{260} units); *lanes 4–8*, incubations with 40, 20, 10, 5, and 2.4 μg of T7 RNA polymerase/ml of incubation mix, respectively. The autoradiogram shows *lanes 4–8*.

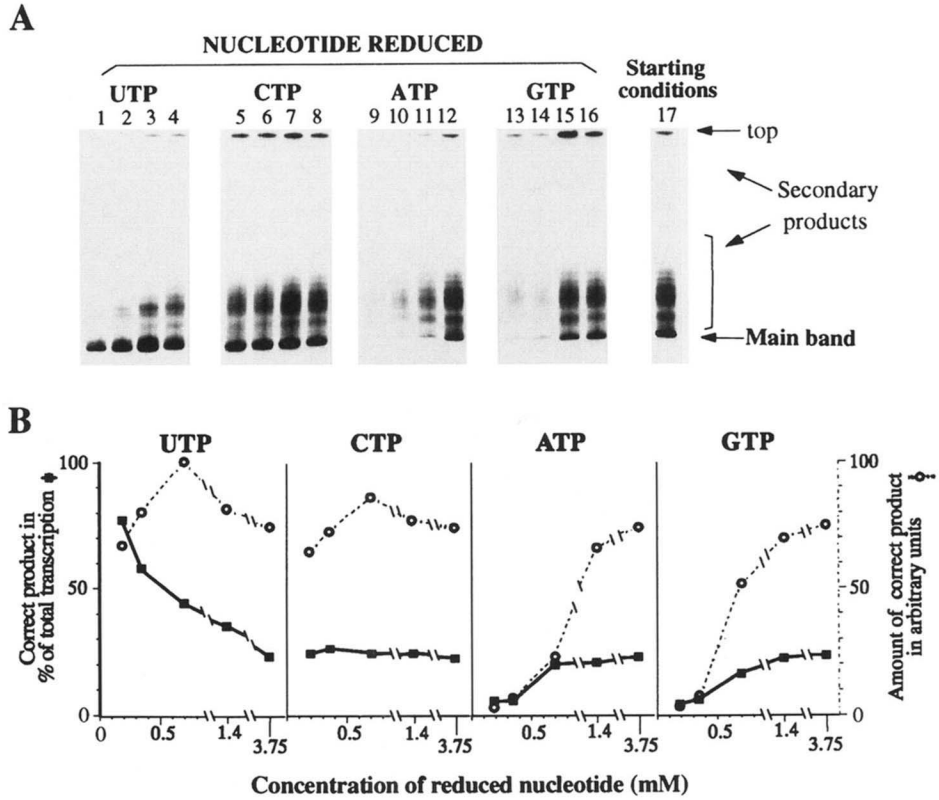

FIG. 7. **Influence of the concentrations of nucleotides on the accuracy of the transcription.** 0.5 pmol of the MF-mRNA template (pTMF) were incubated in 25-μl aliquots under standard conditions adjusting the concentration of one nucleotide at a time to the following values: 3.75 mM (standard conditions); 1.4 mM; 0.7 mM; 0.35 mM and 0.188 mM. 17 parallel incubations changing the concentration of ATP, GTP, CTP, and [^{32}P]UTP (2 cpm/pmol) as well as a control omitting the plasmid were performed. After 3.5-h incubation at 37 °C the reaction was stopped, and samples from every incubation were applied to a 10% sequencing gel. The corresponding autoradiograms were analyzed by densitometry. *A*, autoradiograms: *lanes 1–4*, UTP increasing from 0.188 mM to 1.4 mM; *lanes 5–8*, corresponding variation of CTP; *lanes 9–12*, variation of ATP; *lanes 13–16*, variation of GTP; *lane 17*, starting conditions (all nucleotides at 3.75 mM). *B*, results of densitometry: ■, fraction of correct transcripts in % of the total RNA synthesized; ○, amounts of correct transcript in arbitrary units.

abortive transcripts (Ling *et al.*, 1989) indicating loose enzyme-RNA complexes, *i.e.* the processivity of the enzyme is reduced when the UTP concentration is lower than that of the other nucleotides. A reduced processivity at decreased UTP conditions would certainly counteract the tendency to self-coded transcript prolongation.

It is not clear whether RNA-coded *de novo* synthesis or self-coded 3′-end extension events occur *in vivo*. However, the intracellular molar ratio of ATP/UTP in *E. coli* is similar to that which we found optimal for an accurate run-off transcription (the ATP concentration *in vivo* is about 3 mM, while that of UTP is about 0.8 mM; Neuhard and Nygaard, 1987), suggesting that these types of activities of the T7 RNA polymerase are not common events (or constitute part of well controlled processes) inside the living cell. *In vitro*, a reduction of the UTP concentration to 0.2–1.0 mM while keeping the other nucleotides in the range of 3.5–4 mM (here the base composition of the expected transcript should be taken into account) is generally applicable and improves the fidelity of transcription, as well as giving a high yield of correct RNA synthesized with T7 RNA polymerase.

Acknowledgments—We thank Dr. Richard Brimacombe and Ralf Jünemann for discussions and Cordula Stiege and Uwe Fehner for expert technical assistance.

REFERENCES

Arnott, S., Chandrasekaran, R., and Marttila, C. M. (1974) *Biochem. J.* **141,** 537–543
Birnboim, H. C. (1983) *Methods Enzymol.* **100,** 243–255
Chamberlin, M., and Ring, J. (1973) *J. Biol. Chem.* **248,** 2235–2244
Chamberlin, M., McGrath, J., and Waskell, L. (1970) *Nature* **228,** 227–231
Chamberlin, M. J., and Ryan, T. (1982) in *The Enzymes* (Boyer, P. D., ed) Vol. 15B, pp. 87–109, Academic Press, New York
Cobianchi, F., and Wilson, S. (1987) *Methods Enzymol.* **152,** 94–110
Davanloo, P., Rosemberg, A. H., Dunn, J. J., and Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 2035–2039
Draper, D. E., White, S. A., and Kean, J. M. (1988) *Methods Enzymol.* **164,** 221–237
Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 9373–9377
Gish, G., and Eckstein, F. (1988) *Science* **240,** 1520–1522

Groebe, D. R., and Uhlenbeck, O. C. (1988) *Nucleic Acids Res.* **16,** 11725–11735
Hahn, Ch.-S., Strauss, E. G., and Strauss, J. H. (1989) *Methods Enzymol.* **180,** 121–130
Konarska, M. M., and Sharp, P. A. (1989) *Cell* **57,** 423–431
Konarska, M. M., and Sharp, P. A. (1990) *Cell* **63,** 609–618
Krieg, P. A., and Melton, P. A. (1984) *Nucleic Acids Res.* **12,** 7057–7070
Ling, M.-L., Risman, S. S., Klement, J. F., McGraw, N., and McAllister, W. T. (1989) *Nucleic Acids Res.* **17,** 1605–1617
Lis, J. T. (1980) *Methods Enzymol.* **65,** 347–353
Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056
Milligan, F., and Uhlenbeck, O. C. (1989) *Methods Enzymol.* **180,** 51–62
Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 8783–8798
Nègre, D., Weitzmann, C., and Ofengand, J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 4902–4906
Neuhard, J., and Nygaard, P. (1987) in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology* (Neidhart, F., Ingraham, J., Low, K., Magasanik, B., Schaester, M., and Umbarger, H. E., eds) pp. 445–473, American Society for Microbiology, Washington D. C.
Paddock, G. V. (1989) *BioTechniques* **8,** 856–865
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Sampson, J. R., and Uhlenbeck, O. C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 1033–1037
Scaringe, S. A., Franklyn, C., and Usman, N. (1990) *Nucleic Acids Res.* **18,** 5433–5441
Schatz, D., Leberman, R., and Eckstein, F. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 6132–6136
Schenborn, E. T., and Mierendorf, R. C. (1985) *Nucleic Acids Res.* **13,** 6223–6236
Sousa, R., Debabrata, P., and Lafer, E. M. (1992) *J. Mol. Biol.* **224,** 319–334
Weitzmann, C. J., Cunningham, P. R., and Ofengand, J. (1990) *Nucleic Acids Res.* **18,** 3515–3520
Weitzmann, C. J., Cunningham, P. R., Nurse, K., and Ofengand, J. (1993) *FASEB J.* **7,** 177–180
Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975) *J. Mol. Biol.* **92,** 181–192
Zuker, M. (1989) *Science* **244,** 48–52
Zuker, M., and Stiegler, P. (1981) *Nucleic Acids Res.* **9,** 133–148