Template Strand Gap Bypass Is a General Property of Prokaryotic RNA Polymerases: Implications for Elongation Mechanisms[†]

Jiang Liu[‡] and Paul W. Doetsch*[,‡,§]

Department of Biochemistry and Division of Cancer Biology, Department of Radiation Oncology, Emory University School of Medicine, Atlanta, Georgia 30322

Received June 18, 1996; Revised Manuscript Received September 18, 1996[⊗]

ABSTRACT: It has previously been shown that T7 RNA polymerase is capable of bypassing gaps on the template strand ranging in size from 1 to 24 nucleotides. This as well as other observations suggested a role for the nontemplate strand during elongation. To establish the generality of this gap bypassing event, we have extended these studies to SP6 and *Escherichia coli* RNA polymerases. SP6 RNA polymerase bypasses template gaps from 1 to 19 nucleotides in size with various degrees of efficiency and produces runoff transcripts of decreasing length corresponding to increasing gap size. RNA sequence analysis of the resulting runoff transcripts revealed that SP6 RNA polymerase faithfully transcribed both parts of the template strand flanking the gapped region. Similar experiments were carried out with *E. coli* RNA polymerase (a multiple subunit enzyme) and indicate that it is also capable of gap bypass albeit with reduced efficiency compared to T7 and SP6 RNA polymerases. It appears that the ability to bypass gaps present on the DNA template strand is a general property of prokaryotic RNA polymerases. These results have implications with respect to the mechanism of elongation and the role of the nontemplate strand in transcription.

To carry out DNA template-dependent transcription, RNA polymerase has to go through four major phases: promoter binding and activation, promoter clearance and RNA chain initiation, chain elongation, and chain termination (Chamberlin, 1995). During elongation, RNA polymerase unwinds duplex DNA so that the template strand is available for RNA polymerase to carry out RNA synthesis via complementary base pairing with the template strand. A melted duplex DNA region plus a nascent RNA−DNA hybrid duplex forms the transcription bubble, in which the function of the nontemplate strand is not well understood (Yager et al., 1987). Recently, a role for the nontemplate strand during transcription elongation was suggested for bacteriophage λ (Ring & Roberts, 1994). A sequence specific structure in the nontemplate strand caused the *Escherichia coli* RNA polymerase elongation complex to pause at positions +16 and +17 on the λ late gene operon. Point mutations at positions +2 and +6 on the nontemplate strand within the transcription bubble abolished the pausing both *in vitro* and *in vivo*. In addition, these mutations on the nontemplate strand also abolished the ability of the late gene regulatory antiterminator Q protein to modify RNA polymerase. These experiments suggest that specific nontemplate strand−RNA polymerase interactions may have a regulatory role in the elongation process.

Recent studies from our group (Zhou & Doetsch, 1994; Zhou et al., 1995) have shown that T7 RNA polymerase transcribes through gaps from 1 to 24 nucleotides in length when present on the template strand and generates runoff transcripts containing a correctly templated sequence but deleted for the gapped region (missing sequence) on the template strand. It was also shown that the downstream portion of the broken template can be threaded into T7 RNA polymerase during the bypass of gaps via a "gap closing" event (Zhou et al., 1995). In addition, it was shown that the nontemplate strand is required for the ternary complex to resume transcription elongation from a stalled state. These findings suggest that T7 RNA polymerase interacts with the nontemplate strand to mediate efficient elongation.

We wished to determine whether the ability to bypass template strand gaps was a general property of other RNA polymerases during elongation. SP6 RNA polymerase is a single polypeptide enzyme of approximately 100 kDa and is similar in many respects to T7 RNA polymerase (Kassavetis et al., 1981). For the present study, SP6 RNA polymerase templates were constructed containing a range of small to large gaps from 1 to 19 nucleotides in length. This collection of templates was utilized under conditions in which only a single round of elongation was allowed to occur, and the efficiency of gap bypass could be determined directly as well as the sequence of the resulting runoff transcripts.

E. coli RNA polymerase is a multiple subunit RNA polymerase. Unlike simple phage RNA polymerases, *E. coli* RNA polymerase transcription is carried out through the coordination of the σ initiation factor and the core enzyme (Chamberlin, 1982). *E. coli* RNA polymerase shares sequence homology with eukaryotic RNA polymerase II core enzyme (Allison et al., 1985; Sweetser et al., 1987). In addition, both of these RNA polymerases share similarities with respect to processes of promoter recognition, transcription initiation, and transcription elongation (Eick et al., 1994). We investigated whether gap bypass is also a property of a multiple subunit RNA polymerase. The results from these studies indicate that both SP6 and *E. coli* RNA polymerase

[†] This work is supported by Research Grant CA 55896 from the National Cancer Institute.

* Corresponding author.
[‡] Department of Biochemistry.
[§] Division of Cancer Biology.

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0006-2960(96)01455-9 CCC: $12.00

are capable of bypassing gaps on the template strand with various degrees of efficiency. Runoff transcripts generated from gap bypass are also internally deleted for the gapped region. These findings suggest that the ability to bypass gaps on the template strand is a general property of prokaryotic RNA polymerases and underscores the importance of the nontemplate strand in the elongation process.

MATERIALS AND METHODS

Materials. Dihydrouridine was purchased from Sigma. The dihydrouridine dimethoxytrityl-blocked phosphoramidite building block was synthesized by Glenn Research (Sterling, VA). Oligonucleotides were synthesized by the Emory University Microchemical Facility or GIBCO and purified by polyacrylamide gel electrophoresis (Maniatis et al., 1982). The sequence of each oligonucleotide was verified by base-specific chemical DNA sequencing (Maxam & Gilbert, 1980). *E. coli* and SP6 RNA polymerases were purchased from Boehringer Mannheim and Stratagene, respectively. Heparin and ribonuclease inhibitor were purchased from Sigma. T4 polynucleotide kinase was purchased from New England Biolabs. T4 DNA ligase was from Promega. Calf intestinal phosphatase was from Boehringer Mannheim. The RNA Sequencing Kit (nuclease method) and the Sequenase Version 2.0 DNA Sequencing Kit were from United States Biochemicals. [α-^{32}P]CTP (sp act. 3000 Ci/mmol), [γ-^{32}P]-ATP (3000 Ci/mmol), and [α-^{35}S]dATP (45 Ci/mmol) were from Amersham. HPLC[1] -purified nucleoside triphosphates were purchased from Pharmacia. First Strand cDNA Synthesis Kit was from Novagen.

Construction of DNA Templates. To construct the control (unbroken) DNA templates SP6-GAP0 and SP6-GAP0-DHU, oligonucleotides 56mer and 56mer-DHU (Figure 1) were 5′-end labeled as previously described (Liu et al., 1995). 5′-End labeled oligo 56mer (100 μM) or 56mer-DHU (containing dihydrouracil at nucleotide position 19 from the 5′-end) was mixed with oligo 56mer-NT (100 μM) in 10 mM $MgCl_2$ (Figure 1). The mixtures were heated to 70 °C for 10 min and slowly cooled to room temperature (4 h), and the annealed DNA templates were isolated as previously described (Zhou & Doetsch, 1994). To construct the one nucleotide gap-containing template (SP6-GAP1), 3000 pmol of oligo 18mer (5′-CACCCGTCTCCAACTCCA-3′) was 5′-end labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (*12*). 5′-End labeled 18mer (300 μM) was added to 56mer-NT (100 μM) and 37mer (300 μM) (5′-CCCAACCAAC-CGTGTATTCTATAGTGTCACCTAAATC-3′) in 10 mM $MgCl_2$ and annealed under the conditions described above, and the DNA template SP6-GAP1 was isolated as previously described (Zhou & Doetsch, 1994). DNA templates SP6-GAP2 through SP6-GAP19 (Figure 1A) were constructed utilizing a similar approach. To construct templates EC-GAP0, oligos 69mer-NT, 46mer, and 53mer (Figure 1B) were 5′-phosphorylated by T4 polynucleotide kinase in 25 μL of kinase buffer containing 4.8 mM ATP at 37 °C for 18 h. Equal amounts (4000 pmol) of phosphorylated 69mer-NT, 46mer, and 53mer plus 68mer-NT and 5′-^{32}P-labeled 38mer were annealed by heating at 70 °C for 10 min and then cooled to room temperature for 4 h. Ligation was carried out with 3 units of T4 DNA ligase at 16 °C for 18 h. After phenol−chloroform extraction, EC-GAP0 was purified from a 20% nondenaturing polyacrylamide gel. EC-GAP1 was constructed using a similar approach with oligonucleotides 68mer-NT, 53mer, 5′-end labeled 37mer, 5′-phosphorylated 69mer-NT, and 5′-phosphorylated 46mer (Figure 1B).

Single Round Transcription and Multiple Round Transcription Experiments. Single round transcription experiments were carried out by preincubation of 2 pmol of DNA template with 10 mM DTT, 3 mM $MgCl_2$, 5 mM NaCl, 20 mM Tris-HCl, pH 7.9, 0.5 mM of ATP, GTP, and UTP, 3 units of ribonuclease inhibitor, and 20 units (1.2 pmol) of SP6 or 1 unit (8 pmol) of *E. coli* RNA polymerase for 8 min at room temperature. After that, heparin (250 μg/mL), CTP (10 μM), and 20 μCi of [α-^{32}P]CTP (3000 Ci/mmol) were added to the preincubation mixture. Two-microliter aliquots were removed at various times following the start of elongation, and reactions were terminated by addition of stop-loading buffer (9.8 M urea, 50 mM EDTA, 0.1% xylene cyanol) at different time points. The ^{32}P-labeled transcripts were analyzed on denaturing 15% polyacrylamide 7 M urea gels and subjected to autoradiography and phosphorimager analysis (Molecular Dynamics Model 445 SI).

Multiple round transcription experiments were carried out as previously described (Liv et al., 1995) with 1 pmol of DNA template in 10 mM DTT, 40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 10 mM NaCl, 0.5 mM ATP, 0.5 mM GTP, 0.5 mM UTP, 10 μM CTP, 10 μCi of [α-^{32}P]CTP, 3 units of ribonuclease inhibitor, and 20 units (1.2 pmol) of SP6 RNA polymerase. RNA size marker ladders were generated by alkaline hydrolysis of ^{32}P-labeled full length transcripts (multiple round) as previously described (Liu et al., 1995).

RNA Sequencing. SP6 RNA polymerase-generated transcripts for sequence analysis were produced as described above from multiple round transcription experiments except that 10 μM CTP and 10 μCi [α-^{32}P]CTP (3000 Ci/mmol) were used in place of 0.5 mM CTP. Unlabeled transcripts were treated with calf intestinal phosphatase to remove the 5′-terminal phosphate. Following purification from a 15% denaturing gel using the ^{32}P-labeled runoff transcript as a marker, phosphatase-treated RNA transcripts were 5′-end labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (Maniatis et al., 1982). RNA sequence analysis was carried out with base-specific ribonucleases as described in the RNA Sequencing Kit (United States Biochemical Corp.).

E. coli RNA polymerase-generated transcripts were produced under single round transcription conditions as described above, and cDNAs were generated with MMLV reverse transcriptase and 3′ antisense primer (5′-ATAC-GACTCACTATAGGGA-3′) under conditions recommended by the supplier (Novagen). The cDNAs were PCR-amplified with the 3′ antisense primer and the 5′ sense primer (5′-TTGCAGAATACACGGAATT-3′) using Fisher Taq DNA polymerase and sequenced using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical).

Quantitation of Bypass Efficiency of SP6 and E. coli RNA Polymerases. Three separate single round transcription experiments for each gap-containing template were carried out, and the relative amounts of stalled, short terminated, and runoff RNA transcripts at each time point were determined by phosphorimager analysis of the transcription

[1] Abbreviations: DHU, 5,6-dihydrouracil; HPLC, high-performance liquid chromatography; ST, stalled transcripts; STT, short terminated transcripts; RO, runoff transcripts; NT, nontemplate strand; MMLV, Moloney murine leukemia virus.

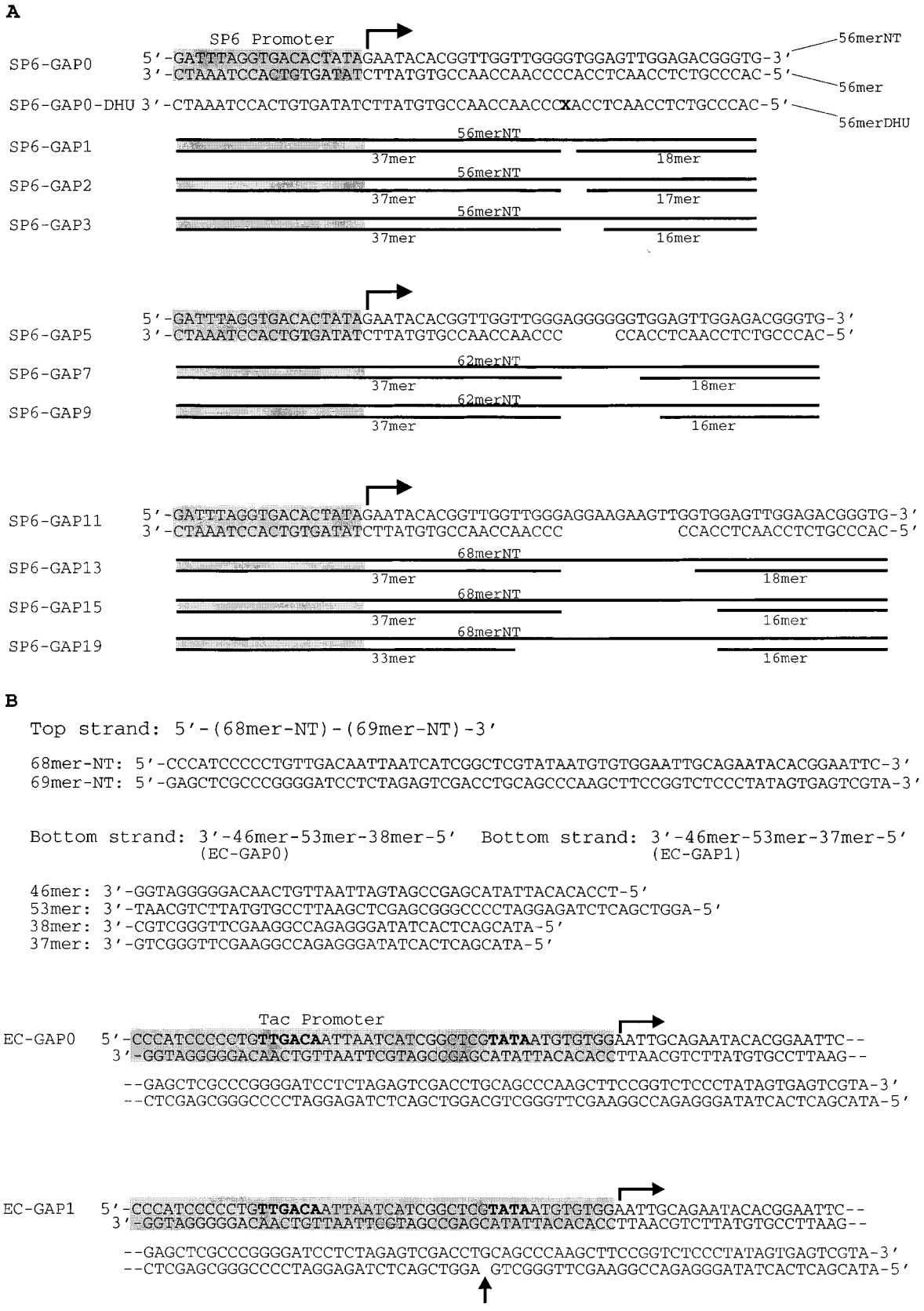

FIGURE 1: (A) SP6 RNA polymerase transcription templates SP6-GAP0, SP6-GAP1, SP6-GAP2, SP6-GAP3, SP6-GAP5, SP6-GAP7, SP6-GAP9, SP6-GAP11, SP6-GAP13, and SP6-GAP15 contain gaps of 0, 1, 2, 3, 5, 7, 9, 11, 13, and 15 nucleotides in length, respectively, on the template strand starting 20 nucleotides downstream from the transcription start site. Template GAP0-DHU contains dihydrouridine (DHU) at position "X", 20 nucleotides downstream from the transcription start site. Template SP6-GAP19 contains a gap 19 nucleotides in length starting 16 nucleotides downstream from the transcription start site. The shaded blocks indicate the SP6 RNA polymerase promoter. The horizontal arrow indicates the transcription start site and direction of transcription. (B) *E. coli* RNA polymerase transcription templates and oligonucleotides from which these templates were synthesized. EC-GAP0 and EC-GAP1 contain an intact (EC-GAP0) or a single nucleotide gap (EC-GAP1) on the template strand starting 55 nucleotides downstream from the transcription start site as indicated by the vertical arrow. Shaded blocks indicate the tac promoter.

products on denaturing polyacrylamide gels as previously described (Liu et al., 1995). For templates transcribed by SP6 RNA polymerase, there are 3 cytosine residues in each runoff transcript (RO) and 2 cytosine residues in each short terminated transcript (STT19−21). The percentage of STT (all species) was calculated as (STT) (3/2)/[(STT) (3/2) + (RO)] × 100, and the percentage of gap bypass was calculated as (RO)/[(STT) (3/2) + (RO)] × 100 (Zhou & Doetsch, 1994). For templates transcribed by *E. coli* RNA polymerase, there are 28 cytosine residues in each full length transcript and 16 cytosine residues in each short terminated transcript (STT). The percentage of STT was calculated as (STT) (28/16)/[(STT)(28/16) + (RO)] × 100, and the percentage of gap bypass was calculated as (RO)/[(STT) (28/16) + (RO)] × 100.

RESULTS

Generation of Gapped DNA Templates for in Vitro Transcription. Unbroken, control DNA templates SP6-GAP0 and SP6-GAP0-DHU were constructed by chemical synthesis of an SP6 RNA polymerase promoter-containing template strand from oligonucleotides 56mer and 56mer-DHU (containing a single DHU 19 nucleotides from the 5′-end) and annealing to the complementary nontemplate strand oligonucleotide 56mer-NT (Figure 1A). The resulting duplex templates contained the SP6 promoter and a transcription region of 38 nucleotides in length. In template SP6-GAP0-DHU, the DHU is located at a position 20 nucleotides downstream from the start of transcription. DHU on the template strand has previously been shown to cause brief pausing of SP6 RNA polymerase (Liu et al., 1995) and thus was used to provide a size marker for short terminated transcripts generated from the gapped templates. Templates SP6-GAP1, SP6-GAP2, and SP6-GAP3 were constructed by the annealing of 56mer-NT with two other oligonucleotides, a 37mer which is complementary to the 5′ end of 56mer-NT and a 18mer, 17mer, or 16mer which are complementary to the 3′-end of 56mer-NT (Figure 1A). Templates SP6-GAP5, SP6-GAP7, and SP6-GAP9 were constructed by annealing of 62mer-NT with the 37mer plus a 20mer, 18mer, or 16mer, respectively (Figure 1A). Templates SP6-GAP11, SP6-GAP13, and SP6-GAP15 were constructed by annealing of oligonucleotide 68mer-NT with the same set of oligonucleotides used in the construction of SP6-GAP5, SP6-GAP7, and SP6-GAP9. SP6-GAP19 was constructed by annealing of 68mer-NT with a 33mer and a 16mer. In templates SP6-GAP1 through SP6-GAP15, the start site of the gap is 20 nucleotides downstream from the start of transcription. In template SP6-GAP19, the start of the gap is 16 nucleotides downstream from the start of transcription. This collection of templates allows for the study of SP6 RNA polymerase transcription of gapped templates which can be directly compared to the previous studies of T7 RNA polymerase behavior on gapped templates (Zhou et al., 1995). Transcription of the control template SP6-GAP0 (unbroken, containing no gaps) should generate full length, runoff transcripts 38 nucleotides in length (RO38), and transcription of SP6-GAP0-DHU should generate a 19 nucleotide stalled transcript (ST19) at early times following the start of elongation with subsequent efficient bypass and generation of RO38 (Liu et al., 1995). The inclusion of this template provides a transcript size marker 19 nucleotides in length and allows direct comparison of SP6 RNA polymerase elongation complexes when its active site encounters a damaged base (DHU) versus a gap on the template strand.

To prepare the templates EC-GAP0 and EC-GAP1 used in the *E. coli* RNA polymerase experiments, five different oligonucleotides were annealed, ligated, and purified from a 20% nondenaturing gel (Materials and Methods). The resulting template EC-GAP1 contains a single nucleotide gap on the template strand 55 nucleotides downstream from the start of transcription (Figure 1B). Transcription of EC-GAP0 should generate a full length, runoff transcript of 92 nucleotides in length (RO92).

SP6 RNA Polymerase Behavior at Small Template Strand Gaps. To determine the interaction of SP6 RNA polymerase with template strand gaps during elongation, comparative single round transcription experiments were carried out initially on templates SP6-GAP0, SP6-GAP0-DHU, SP6-GAP1, and SP6-GAP2. Each molecule of template is utilized only once by a single molecule of RNA polymerase, and the transcription products generated reflect a single, promoter-dependent elongation event (Zhou & Doetsch, 1993). Single round transcription experiments with the control template SP6-GAP0 produced full length, runoff transcripts (RO38) as expected (Figure 2A, lanes 2−7). Under these conditions, the transcription process was essentially complete after 60 s, followed by one and two nucleotide nontemplated additions to the 3′-end of RO38 which resulted the generation of species 39 and 40 nucleotides in length. Such nontemplated additions are a property of several RNA polymerases, including SP6 RNA polymerase (Jacques et al., 1991; Zhou & Doetsch, 1993). Transcription of template SP6-GAP0-DHU (containing DHU in place of cytosine on the template strand 20 nucleotides downstream from the start of transcription) resulted in the initial production of stalled transcript ST19 which rapidly disappeared at later times and was converted into a full length transcript (RO38) (Figure 2A, lanes 8−13). Such pausing/elongation at sites of DHU has been previously observed with both SP6 and T7 RNA polymerases (Liu et al, 1995). In contrast, a single round of transcription with template SP6-GAP1 generated short terminated transcripts of 19 nucleotides (STT19) and runoff transcripts (Figure 2A, lanes 14−19). However, with the gapped template nontemplated additions also occurred with STT19 to generate short terminated transcripts of 20 and 21 nucleotides in length (STT20−21). Unlike ST19 resulting from the transcription of template SP6-GAP0-DHU in which essentially all of ST19 was elongated into full length RO38, the total amount of STT19−21 generated from template SP6-GAP1 did not change with time, indicating that a fraction of SP6 RNA polymerase is permanently arrested at the gap site and has lost the ability for further chain elongation. However, the majority of template-engaged SP6 RNA polymerase was able to continue RNA synthesis and generate full length runoff transcripts. The runoff transcript generated from template SP6-GAP1 was 37 nucleotides in length (RO37) and was one nucleotide shorter than RO38 produced from templates SP6-GAP0 and SP6-GAP0-DHU. Nontemplated additions of one and two nucleotides also occurred for RO37, resulting in the generation of species 38 and 39 nucleotides in length. In addition, for some templates, a family of minor bands can be observed just below the region of the RO transcripts and may result from premature release as the elongation complex approaches the end of the template.

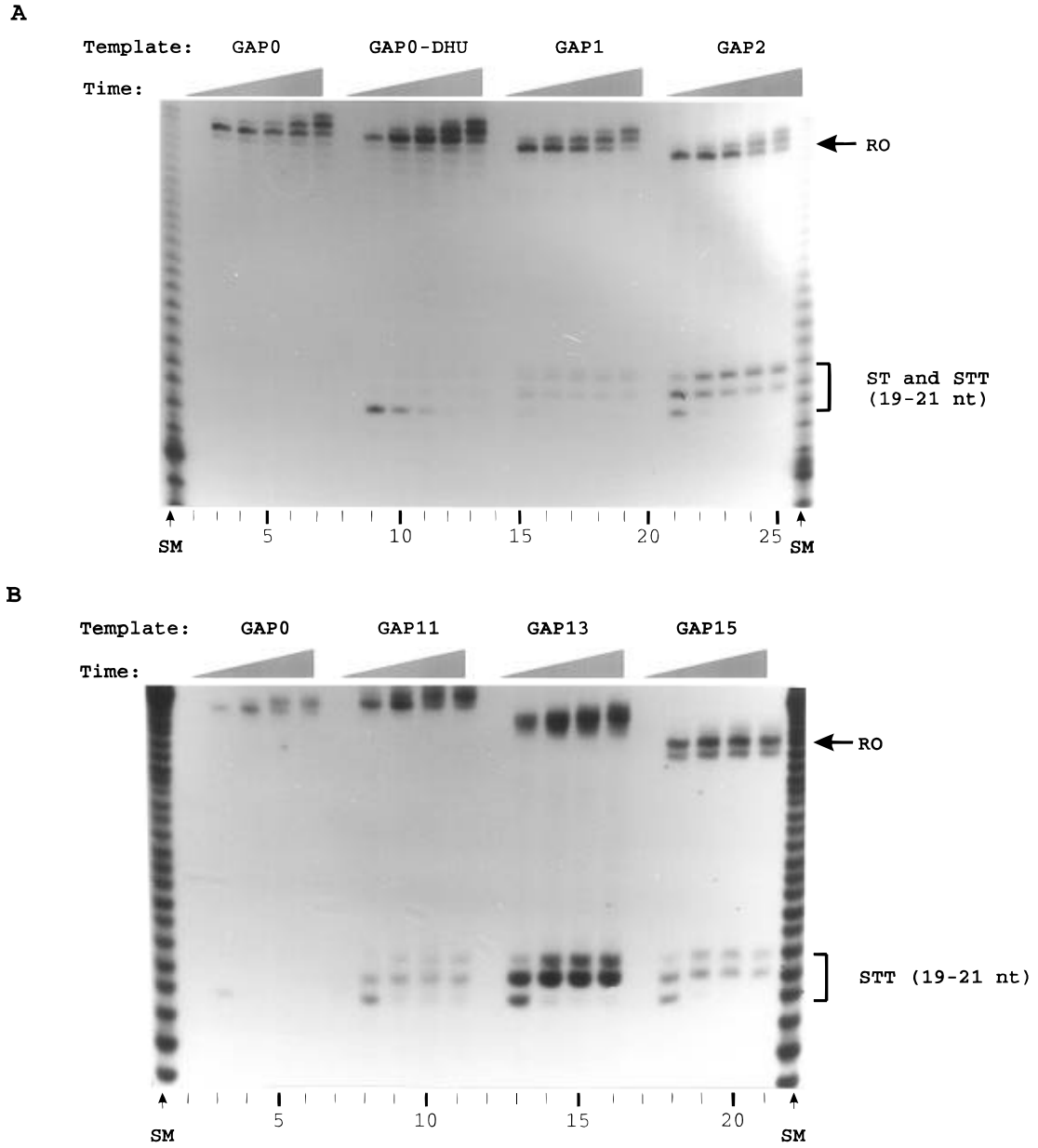

FIGURE 2: (A) Single round transcription experiments with SP6 RNA polymerase and templates SP6-GAP0, SP6-GAP0-DHU, SP6-GAP1, and SP6-GAP2 templates. Region ST and STT indicates either stalled transcripts (ST19) or short terminated transcripts (STT19−21), and arrow RO indicates runoff transcripts. Lanes 2−7, 8−13, 14−19, and 20−25 correspond to transcription products generated at 10 s, 20 s, 30 s, 1 min, and 2 min following the start of elongation, respectively. RNA size marker ladder is included in the end lanes (SM). (B) Single round transcription experiments with SP6 RNA polymerase with templates SP6-GAP0, SP6-GAP11, SP6-GAP13, and SP6-GAP15. Region STT indicates short terminated transcripts (STT19−21), and arrow RO indicates runoff transcripts. Lanes 2−6, 7−11, 12−16, and 17−21 represent transcription products generated at 10 s, 30 s, 1 min, and 2 min after the start of transcription, respectively.

Quantitation of the relative amounts of transcripts RO37 and STT19−21 indicated that the efficiency of a one nucleotide gap bypass by SP6 RNA polymerase is about 80% (Figure 3). The result of a single round of transcription on a two nucleotide gap-containing template (SP6-GAP2) is similar to the result obtained with template SP6-GAP1. Both short terminated and runoff transcripts were produced (Figure 2A, lanes 20−25), and the gap bypass efficiency was approximately 72% (Figure 3). Runoff transcripts 36 nucleotides in length were generated and are two nucleotides shorter than those produced with template SP6-GAP0. Single round transcription experiments with template SP6-GAP3 generated runoff transcripts 35 nucleotides in length which are three nucleotides shorter than RO38 (not shown). Templates SP6-GAP1, SP6-GAP2, and SP6-GAP3 were constructed with the same nontemplate strand 56mer-NT and the same 37mer which is complementary to the upstream portion of 56mer-NT. The difference with these three templates is in the distal portion of the template strand corresponding to 18mer (SP6-GAP1), 17mer (SP6-GAP2), or 16mer (SP6-GAP3). Thus, the size of the runoff transcripts from one, two, and three nucleotide gap bypass events corresponds to RO38 minus the gap size. Single round transcription experiments with templates SP6-GAP5, SP6-GAP7, and SP6-GAP9 (Figure 1A) were also carried out (not shown), and runoff transcripts 39, 37, and 35 nucleotides in length, respectively, were generated with decreasing efficiencies (Figure 3). In each case of template gap bypass, the sizes of the runoff transcripts are equal to the theoretical length of the runoff transcript that would result from transcription of the corresponding control, unbroken template minus the gap size (Figure 1A). We conclude from

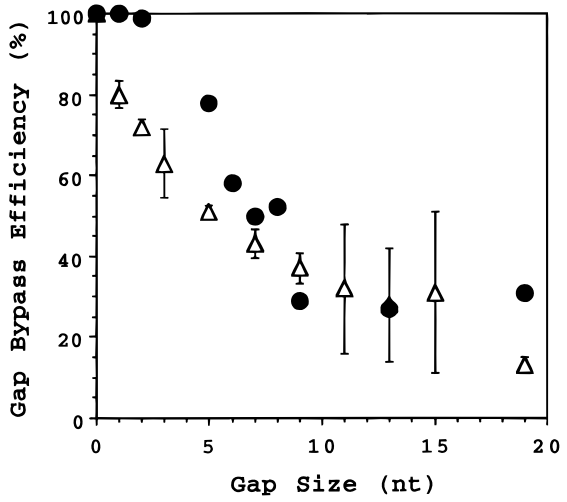

FIGURE 3: Bypass efficiency of SP6 RNA polymerase at sites of gaps on the transcribed strand. The relative amounts of short terminated transcripts (STT19−21) and runoff (RO) transcripts were determined as described in Materials and Methods. This figure compares the bypass efficiency of SP6 RNA and T7 RNA polymerases on template strand gaps 1−19 nucleotides in length. For SP6 RNA polymerase, three separate single round transcription experiments were conducted for each template. Error bars represent standard deviations. The data for the bypass of gaps by T7 RNA polymerase is taken from Zhou et al. (1995). Open triangles correspond to SP6 RNA polymerase, and the filled circles correspond to T7 RNA polymerase.

these results that SP6 RNA polymerase is capable of bypassing small gaps on the template strand and that the efficiency of gap bypass decreases as the gap size increases.

SP6 RNA Polymerase Behavior at Large Template Strand Gaps. We wished to determine whether SP6 RNA polymerase exhibited stalling or bypass behavior at large gaps (greater than 10 nucleotides in length) on the template strand. Single round transcription experiments with SP6 RNA polymerase were carried out on templates SP6-GAP11, SP6-GAP13, SP6-GAP15, and SP6-GAP19. Transcription of the control template SP6-GAP0 generated a full length runoff transcript RO38 as expected (Figure 2B, lanes 2−7) and which, at later times, was extended by two additional nucleotides via nontemplated additions. The results of SP6 RNA polymerase single round transcription of templates SP6-GAP11, SP6-GAP13, and SP6-GAP15 (Figure 2B, lanes 7−11, 12−16, and 17−21) were similar to those obtained with SP6 RNA polymerase with templates SP6-GAP1 and SP6-GAP2 except that a lower overall efficiency of bypass was observed (Figure 3). Runoff transcripts of 31 nucleotides in length were generated as a result of bypass of a 19 nucleotide gap, and the efficiency of bypass was 12% (data not shown). For all of the large gap-containing templates (SP6-GAP11 through SP6-GAP19), the size of the corresponding runoff transcripts is the size of the theoretical full length RO generated from unbroken template minus the gapped region. The efficiencies of gap bypass by SP6 and T7 RNA polymerases are compared in Figure 3. In general, the gap bypass efficiencies for SP6 and T7 RNA polymerases are similar. SP6 RNA polymerase stalls temporarily at sites of DHU, and the stalled ternary complexes are still transcriptionally engaged with subsequent base insertion opposite to DHU and further efficient chain elongation into runoff transcripts (RO38). However, arrested elongation complexes from gap-containing templates are no longer transcriptionally engaged. These arrested elongation complexes are converted into terminated complexes after one or two nucleotide nontemplated additions occur (Figure 2).

Sequence Analysis of SP6 RNA Polymerase Transcripts. To determine the nature of the transcript that was generated as a result of gap bypass by SP6 RNA polymerase, we utilized multiple round transcription conditions to generate runoff transcripts from the 5 nucleotide gap-containing template SP6-GAP5. This template possesses an intermediate gap size and is bypassed with a relatively high efficiency to produce a sufficient amount of RNA for direct sequence analysis using base-specific ribonucleases (Materials and Methods). Sequence analysis of the runoff transcript produced from the control template SP6-GAP0 revealed the expected nucleotide sequence complementary to the template strand (Figure 4A,B). The sequences obtained from the runoff transcript produced from polymerase bypass of the template strand gap in SP6-GAP5 indicated that SP6 RNA polymerase faithfully transcribed the broken template strand flanking the gap site. From the 5'-end of the transcripts generated from template SP6-GAP5, there are five consecutive Gs at nucleotide positions 16−20. The first three Gs (16−18) are complementary to the three Cs at the 5′-end of the proximal portion of the broken template strand, whereas the last two Gs (19 and 20) are complementary to the first two 3′ Cs of the distal portion of the broken template (Figure 4A). Thus, the sequence of this runoff transcript reflected the full template sequence with the RNA deleted precisely for the gapped region on the template strand. We conclude from the gap bypass and sequence analysis results that SP6 RNA polymerase can efficiently transcribe templates with gaps on the template strand and generate faithfully templated, internally deleted transcripts.

E. coli RNA Polymerase Bypasses a Gap in the Template Strand. Both SP6 and T7 RNA polymerases are single polypeptide phage enzymes. Therefore, it was important to determine whether gap bypass was also a property of multiple subunit RNA polymerases such as *E. coli* RNA polymerase (5 subunits). Single round transcription experiments were carried out with templates EC-GAP0 (control, unbroken template) and EC-GAP1 (single nucleotide gap template). Transcription of template EC-GAP0 resulted the generation of full length runoff transcripts 92 nucleotides in length (RO92) as expected (Figure 5A, lanes 7−12). Under the same conditions, transcription of EC-GAP1 generated short terminated transcripts in the range of 54 nucleotides in length as well as runoff transcripts (Figure 5A, lanes 1−6). It appeared that the mobility of this runoff transcript (RO91) was slightly greater than that of RO92. This mobility difference was further substantiated by subjecting these runoff transcripts to electrophoresis for 20 h on a 20% denaturing polyacrylamide gel which confirmed that RO generated from EC-GAP1 is one nucleotide shorter in length than RO generated from EC-GAP0 (Figure 5B). The gap bypass efficiency of *E. coli* RNA polymerase is 50 ± 3% (Materials and Methods). We conclude that *E. coli* RNA polymerase can also bypass at least a one nucleotide gap on the template strand and generate a runoff transcript which is one nucleotide shorter than the runoff transcript resulting from the unbroken, control template. Based on these results, we also conclude that gap bypass is a general property of prokaryotic RNA polymerases.

Sequence Analysis of E. coli RNA Polymerase Transcripts. Transcripts from templates EC-GAP0 and EC-GAP1 were

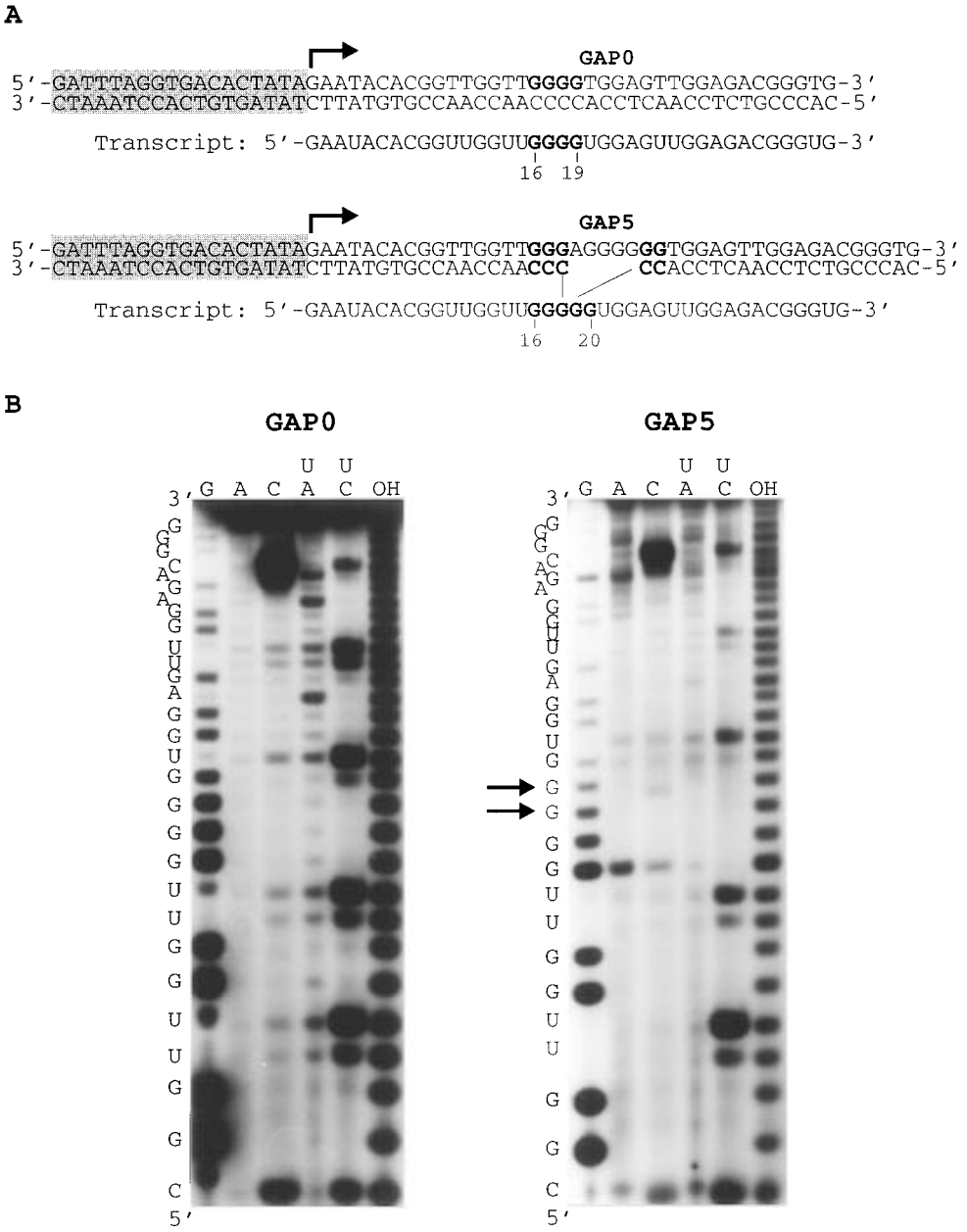

FIGURE 4: (A) Templates used in RNA sequence analysis. Template SP6-GAP0 encodes a full length transcript 38 nucleotides in length, and the predicted sequence of the transcript is shown. Template SP6-GAP5 contains a gap on the template strand 5 nucleotides in length corresponding to nucleotide positions 20−24 downstream from the transcription start site. Sequence for the deduced transcript generated from SP6-GAP5 is shown below the template. (B) Transcript sequence analysis. SP6 RNA polymerase-generated transcripts from template SP6-GAP0 and SP6-GAP5 were produced under multiple round transcription conditions and 5′-end labeled as described in Materials and Methods. RNA sequencing was carried out with base-specific ribonucleases (lanes G, A, C, U+A, and U+C). Arrows indicate the nucleotides inserted at positions 18 and 19 from the 5′-end of the RNA and correspond to insertions opposite to the two cytosines flanking the gap on the template strand of SP6-GAP5 (panel A). The RNA size ladder was generated by alkaline hydrolysis (OH lanes) of 5′-end labeled RO38 from template SP6-GAP0 (Liu et al., 1995).

generated under single round transcription conditions. Runoff transcripts were used to generate cDNA s by RT-PCR followed by DNA sequencing (Materials and Methods). Sequence analysis of the runoff transcripts generated from template EC-GAP0 revealed that *E. coli* RNA polymerase faithfully transcribed the broken template strand flanking the gap (Figure 6). The sequence of the transcripts as a result of gap bypass reflected the full template sequence with the RNA deleted for the gapped nucleotide. From these results we conclude that *E. coli* RNA polymerase can bypass a one nucleotide gap on the template strand and generate correctly templated, internally deleted transcripts.

DISCUSSION

The results of these studies indicate that both SP6 and *E. coli* RNA polymerases are able to continue elongation when there are gaps of various sizes present on the template strand. Both RNA polymerases generate internally deleted transcripts from such templates. This suggests that gap bypass is a general property of prokaryotic RNA polymerases. The decrease of gap bypass efficiency as gap size increases probably represents the competition between gap closing, which supports further chain elongation, and nontemplated additions leading to termination of transcription (Zhou et al., 1995). This notion is further supported by the different

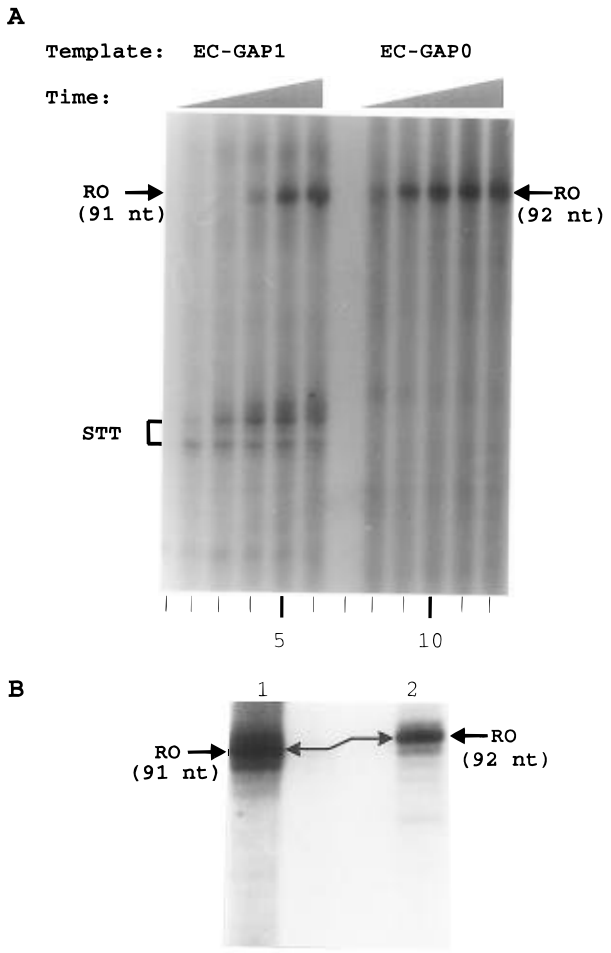

FIGURE 5: Template strand gap bypass by *E. coli* RNA polymerase. (A) Single round transcription experiments with templates on EC-GAP0 and EC-GAP1. Region STT corresponds to short terminated transcripts. Arrow RO indicates runoff transcripts 91 and 92 nucleotides in length. Lanes 1−6 and 7−12 correspond to transcripts generated at 10 s, 30 s, 1 min, 2 min, and 4 min, respectively. (B) Size difference of RO transcripts generated from templates EC-GAP0 and EC-GAP1. Lanes 1 and 2 are the runoff transcripts from template EC-GAP1 and EC-GAP0, respectively. Runoff transcripts (RO) produced at 1 min following start of elongation for both templates were subjected to electrophoresis on a 20% denaturing polyacrylamide gel for 20 h followed by autoradiography. The size difference between the two transcripts is one nucleotide and is indicated by the horizontal lines between lanes 1 and 2.

pattern of transcripts generated from SP6 RNA polymerase bypass of DHU versus the gaps on the template strand (Figure 2A). DHU only briefly stalls the RNA polymerase, and the stalled elongation complex is still transcriptionally engaged. After SP6 RNA polymerase inserts an adenine opposite to the site of DHU, efficient elongation is resumed (Liu et al., 1995). Nontemplated additions do not occur because of the presence of an intact unbroken DNA template strand, and the stalled elongation complexes are able to continue RNA synthesis with a high degree of efficiency. When SP6 RNA polymerase is arrested at a gap site, nontemplated additions leading to termination occur at the same time the gap is being closed. As a result, the cumulative amount of STT19−21 did not change over time during a single round of transcription (Figure 2A,B). Thus it can be predicted that SP6 RNA polymerase bypasses large gaps with a lower efficiency because a longer time would be required to close large gaps compared to small gaps or, alternatively, reflects an increased probability of adding a nontemplated nucleotide to the stalled transcript, resulting in a dead end complex.

RNA sequence analysis revealed that both SP6 and *E. coli* RNA polymerases are able to bring the distal portion of the broken template to the active site and continue the elongation process. When there is a 13, 15, or 19 nucleotide gap present in the template strand, we speculate that the distal portion of the template strand is likely to be outside of the leading edge of the SP6 RNA polymerase molecule when its active site is stalled at the beginning of the gap. The fact that both SP6 and T7 RNA polymerases bypass large gaps on the template strand suggests that it is the RNA polymerase−nontemplate strand interaction that pulls the distal portion of template strand into the polymerase and closes the gap (Zhou et al., 1995). This model is further supported by the observation that when certain regions of the nontemplate strand are deleted, SP6 RNA polymerase cannot efficiently generate full length runoff transcripts (J. Liu and P. Doetsch, unpublished results). These gap bypass results suggest a critical role for a RNA polymerase−nontemplate strand interaction during the normal elongation process. It is possible that it is this polymerase−nontemplate strand interaction that helps maintain efficient movement of polymerase along the DNA template strand.

Although only prokaryotic RNA polymerases were investigated in this and previous studies of template strand gap bypass, the results with *E. coli* RNA polymerase may have implications for eukaryotic RNA polymerases as well. *E. coli* RNA polymerase and eukaryotic RNA polymerase II are responsible for the regulated expression of thousands of genes during various cellular process. Many similarities exist between these two polymerases and their transcription processes. General initiation factors TBP (Horikoshi et al., 1989), TFIIB (Ha et al., 1991), and TFIIE (Ohkuma et al., 1991) share different regions of sequence homology with σ70 initiation factor. The β and β' subunits of *E. coli* RNA polymerase also share sequence homology with the two largest subunits of RNA polymerase II (Allison et al., 1985; Sweetser et al., 1987). The ability of *E. coli* RNA polymerase to bypass a gap on the template strand indicates that gap bypass can be achieved by a multiple subunit RNA polymerase and thus underscores the importance to investigate the behavior of RNA polymerase II at a gap site.

The efficiency of *E. coli* RNA polymerase bypass of a single nucleotide gap is lower than that of SP6 and T7 RNA polymerases. A relatively lower overall transcription efficiency of *E. coli* RNA polymerase compared with the phage RNA polymerases could account for this difference. A slower moving *E. coli* RNA polymerase elongation complex would pull the template strand in at a slower rate (compared to more efficient phage RNA polymerases) when encountering a gap on the template strand. We propose that such slower processing would increase the chance for nontemplated additions resulting in transcription termination. It could also be predicted that gap bypass will be less efficient with eukaryotic RNA polymerases because of a lower overall transcription efficiency compared to that of prokaryotic RNA polymerases.

We have shown that small to large size gaps in the template strand can be bypassed by prokaryotic RNA polymerases. The resulting transcripts are internally deleted for the gapped region. Single nucleotide gaps can be generated in cells by free radical damage to DNA as a

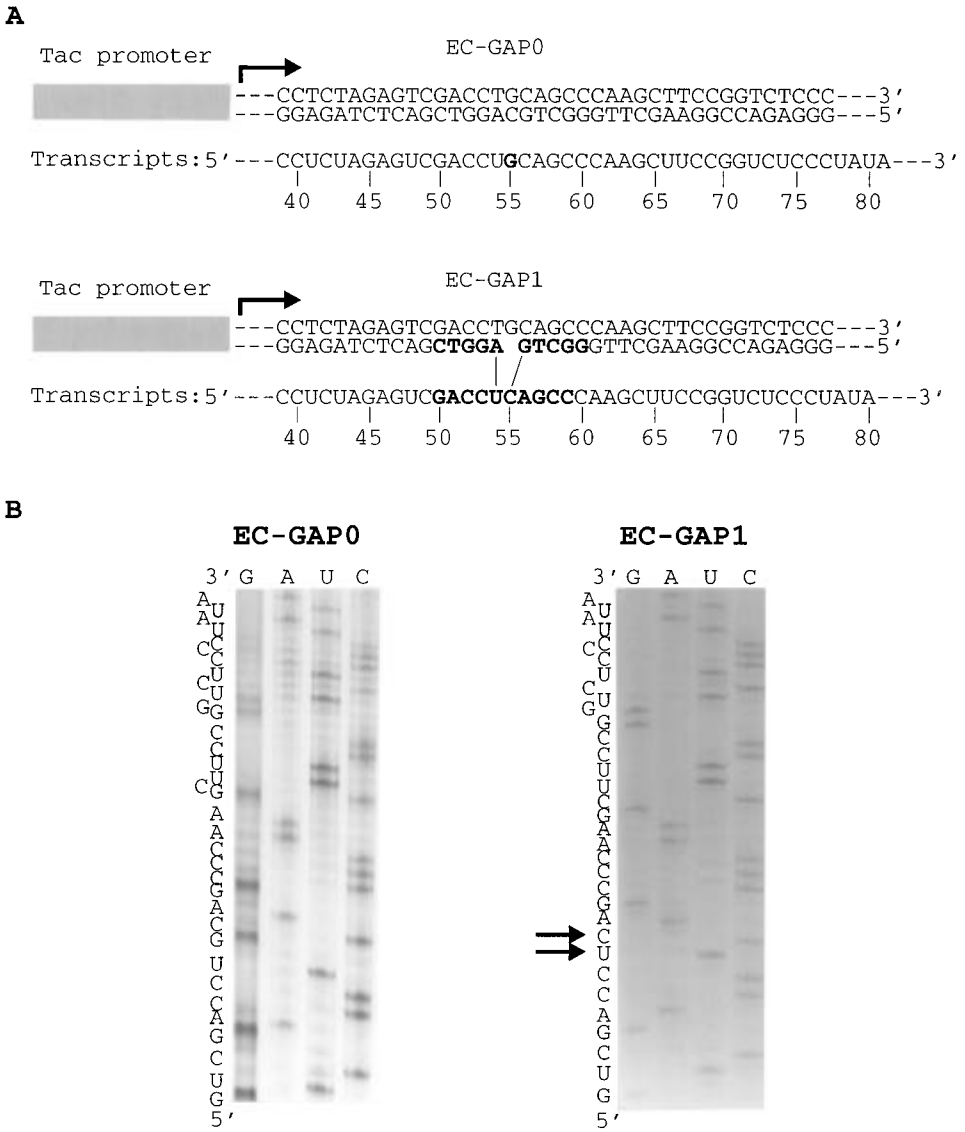

FIGURE 6: Sequence analysis of transcripts generated from templates EC-GAP0 and EC-GAP1. (A) Template EC-GAP0 encodes a full length transcript 92 in length, and part of the predicted transcripts is shown. Template EC-GAP1 contains a one nucleotide gap at nucleotide position 55 downstream from the transcription start site, and part of the deduced transcript sequence is shown. (B) Transcript sequence analysis. RT-PCR and DNA sequencing were carried out as described in Materials and Methods. Arrows indicate nucleotides U and C inserted at positions 54 and 55 downstream from the 5′-end of the transcript, respectively, and correspond to insertions opposite to the two nucleotides flanking the gap on the template strand of EC-GAP1 (panel A).

consequence of exposure to ionizing radiation and certain chemicals (Ward, 1990). Therefore, it is possible that RNA elongation complexes could encounter and bypass single nucleotide gaps with various degrees of efficiency and produce transcripts containing the equivalent of a deletion mutation. It would be quite rare that gaps larger than one nucleotide would exist in cells, but if such larger gaps were encountered by RNA polymerase elongation complexes, the gap bypass event would result in large deletion mutations in the resulting transcript. For example, the *E. coli* nucleotide excision repair system removes a 12−13 nucleotide oligonucleotide containing the damage and generates a gapped duplex repair intermediate (Sancar, 1994). However, it would be counterproductive to permit an elongation complex to engage such a region before the DNA repair process has been completed. Whether or not gap bypass occurs *in vivo* is not known, and its potential biological relevance remains to be elucidated.

ACKNOWLEDGMENT

We thank the members of the Doetsch laboratory for technical assistance and helpful discussions. We also thank Dr. Daniel Reines for helpful discussions.

REFERENCES

Allison, J. A., Moyle, M., Shales, M., & Ingles C. J. (1985) *Cell 42* (2), 599−610.

Chamberlin, M. J. (1982) in *The Enzymes* (Boyer, P. D., Ed.) Vol. 15, pp 61−108, Academic, New York.

Chamberlin, M. J. (1995) *The Harvey Lectures 88*, 1−21.

Daube, S. S., & von Hippel, P. H. (1992) *Science 258*, 1320−1324.

Eick, D., Wedel, A., & Heumann, H. (1994) *Trends Genet. 10* (8), 292−296.

Ha, I., Lane, W. S., & Reinberg, D. (1991) *Nature 352*, 689−695.

Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil, P. A., & Roeder, R. G. (1989) *Nature 341*, 299−303.

Jacques, J., & Kolakofsky, D. (1991) *Genes Dev. 5*, 707−713.

Kassavetis, G. A., Butler, E. T., Roulland, D., & Chamberlin, M. J., B. (1981) *J. Biol. Chem. 257*, 577−5788.
Liu, J., Zhou, W., & Doetsch, P. D. (1995) *Mol. Cell. Biol. 15*, 6729−6735.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 449−560.
Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M., & Roeder, R. G. (1991) *Nature 354*, 398−401.
Ring, B. Z., & Roberts, J. W. (1994) *Cell 78*, 317−324.
Sancar, A. (1994) *Science 266*, 1954−1956.
Sweetser, D. J., Nonet, M., & Young, R. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1192−1196.
Ward, J. F. (1990) *Int. J. Radiat. Biol. Biol. 57*, 1141−1150.
Yager, T. D., & von Hippel, P. H. (1987) in *E. coli and S. typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., Ed.) American Society for Microbiology Publications, Washington, DC. pp 1241−1275.
Zhou, W., & Doetsch, P. W. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6601−6606.
Zhou, W., & Doetsch, P. W. (1994) *Biochemistry 33*, 14926−14934.
Zhou, W., Reines, D., & Doetsch, P. W. (1995) *Cell 82*, 577−585.

BI961455X