Synthesis of in Vitro ColE1 Transcripts with 5′-Terminal Ribonucleotides That Exhibit Noncomplementarity with the DNA Template†

Ronald C. Parker*
Department of Chemistry, Indiana University, Bloomington, Indiana 47405
Received March 4, 1986; Revised Manuscript Received July 2, 1986

ABSTRACT: A region that forms the S1 nuclease site in ColE1 DNA is shown to code for an in vitro transcript, called S1 RNA-B, which contains a 5′-terminal GTP residue that exhibits noncomplementarity with the template's DNA sequence. The synthesis of S1 RNA-B initiates four bases upstream from the start point for S1 RNA-C. The initial four bases in S1 RNA-B and S1 RNA-C are identical. The relative synthesis of S1 RNA-B to S1 RNA-C is sensitive to the concentration of GTP, a substrate that is required for elongation past the +4 position in S1 RNA-C. Dinucleotides that are expected to only initiate synthesis of S1 RNA-C yield two transcripts that appear to initiate from the S1 RNA-C and S1 RNA-B start sites. In vitro studies involving other ColE1 transcripts, RNA-B and RNA-C, provide similar observations concerning the noncomplementary initiation phenomenon. A model involving transcriptional slippage is suggested to explain the noncomplementary initiation phenomenon. The model proposes that the cycling reaction of *Escherichia coli* RNA polymerase produces tetranucleotides that are transposed to nearby upstream sequences for priming transcription.

The initiation of transcription by *Escherichia coli* RNA polymerase is a complex process involving initial binding of RNA polymerase to promoter sequences (closed-complex formation), isomerization of the closed complex, resulting in localized melting of DNA surrounding the transcription start point (open-complex formation), and polymerization of template-directed ribonucleotides, leading to the formation of a stable elongation complex (promoter clearance) (Chamberlin, 1976; McClure, 1985). Productive initiation reactions are determined by several factors, including escape from abortive synthesis of oligonucleotides in the presence of all four ribonucleotides (the cycling reaction) (Carpousis & Gralla, 1980; Munson & Reznikoff, 1981), pausing of RNA polymerase near the transcription start point (Dahlberg & Blattner, 1973; Gilbert, 1976), and release of the σ subunit from RNA polymerase at the end of the initiation process (Travers & Burgess, 1969; Hansen & McClure, 1980).

This model must be reconciled with findings that characterize three ColE1 transcripts to initiate in vitro with 5′-terminal ribonucleotides that do not show complementarity with the DNA template (Parker, 1983a; Patient, 1979; R. Patient, personal communication). A 5′-terminal GTP residue begins the synthesis of each of these transcripts, despite the presence of thymidylate residues on the coding strand at the assigned start points for transcription. Since ATP does not initiate synthesis of these transcripts (Parker, 1983a; Patient, 1979), it is difficult to envision how T·G base pairing via enolization of either residue or "wobble-like" pairing (Paetkau et al., 1972) could produce the phenomenon.

The current study characterizes the synthesis of another in vitro ColE1 transcript that initiates with a noncomplementary 5′-terminal ribonucleotide, an event that bears several similarities with a previous observation (Parker, 1983a). The evidence does not substantiate results (Patient, 1979; R. Patient, personal communication) that suggest two overlapping transcripts of ColE1 DNA initiate with noncomplementary 5′-terminal ribonucleotides. The data suggest that the initiation phenomenon is explained by a transcriptional slippage model, wherein the cycling reaction of *E. coli* RNA polymerase produces oligonucleotides that prime transcription from nearby upstream sequences. This model has been offered to explain the noncolinearity of DNA template sequences with 5′-terminal sequences for some in vitro and in vivo transcripts of polyoma virus DNA (Cowie et al., 1982). With regards to *E. coli* RNA polymerase activities, a slippage model was first developed to explain the synthesis of poly(adenylic acid) from single-stranded DNA in the presence of ATP alone (Chamberlin & Berg, 1962). The data support a reiterative copying mechanism involving short runs of thymidylate residues in the DNA template (Chamberlin & Berg, 1964). Reiterative copying by *E. coli* RNA polymerase has been observed in studies concerning the synthesis of long ribopolynucleotides from short oligomers containing repeating sequences of one, two, or three deoxyribonucleotides (Falaschi et al., 1963; Nishimura et al., 1964).

Experimental Procedures

E. coli RNA polymerase either was obtained from a preparation based on the method of Burgess and Jendrisak (1975) or was supplied by Enzo Biochemicals. ColE1 DNA and DNA fragments were prepared as previously described (Parker, 1983a). Nucleoside triphosphates, S1 nuclease, and calf intestine alkaline phosphatase were purchased from Boehringer/Mannheim. Dinucleotides, heparin, rifampicin, ribonuclease (RNase)[1] T2, and RNase A were obtained from Sigma. T4 polynucleotide kinase and RNase *Phy*I were provided by P-L Biochemicals. RNase T1 and RNase U2 were obtained from Calbiochem. [γ-^{32}P]GTP (3200 Ci/mmol) and [γ-^{32}P]ATP (7000 Ci/mmol) were supplied by ICN, while [α-^{32}P]UTP (7000 Ci/mmol) and [α-^{32}P]CTP

† This work was supported by National Institutes of Health Postdoctoral Fellowship F32 GM09358 and by U.S. Public Health Service Grant R01 AI 10142 to J. P. Richardson.
* Address correspondence to this author at the Department of Biological Sciences, Columbia University, New York, NY 10027.

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane; RNase, ribonuclease; PEI, poly(ethylenimine); tRNA, transfer RNA; mRNA, messenger RNA.

0006-2960/86/0425-6593$01.50/0

(7000 Ci/mmol) were prepared by the method of Walseth and Johnson (1979).

The transcription conditions for the initial identification and purification of transcripts were 10 mM Tris·HCl (pH 8.0), 10 mM $MgCl_2$, 125 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA, 50–100 μg/mL bovine serum albumin, 50 μM each of ATP, GTP, CTP, and UTP, 10 pmol of *E. coli* RNA polymerase, 1 μCi of [α-^{32}P]CTP or 225 μCi of [γ-^{32}P]GTP, and either 1.2 pmol of ColE1 DNA or 4 pmol of *Eco*RI$_1$–*Alu*I$_{219}$ DNA fragment (see Figure 1). Following 40-min reactions at 37 °C, the products were electrophoresed on polyacrylamide–urea gels, and [γ-^{32}P]GTP-labeled transcripts were purified as previously described (Parker, 1983a).

The 5′-terminal RNA sequences were analyzed by enzymatic techniques (Donis-Keller et al., 1977; Wurst et al., 1978; Donis-Keller, 1980; Swerdlow & Guthrie, 1984). Partial alkaline hydrolysis was done by combining aliquots of RNA that had been treated in 25 mM $NaHCO_3$ and 25 mM Na_2CO_3 (pH 9.0) for 1, 4, and 8 min at 92 °C. Two aliquots of RNA were dissolved in 20 mM sodium citrate (pH 5.0), 7 M urea, 1 mM EDTA, 0.025% bromphenol blue, and 0.025% xylene cyanol, then heated to 50 °C for 5 min, and incubated at 50 °C for 15 min with either RNase U2 (0.3 unit/μg of RNA) or RNase T1 (0.03 unit/μg of RNA). RNA was dissolved in 10 mM sodium acetate (pH 4.5) and 1 mM EDTA and then treated at 23 °C for 10 min with RNase *Phy*I (0.33 unit/μg of RNA). A portion of RNA was dissolved in 100 mM Tris·HCl (pH 7.5) and 1 mM EDTA, and then treated at 23 °C for 10 min with RNase A (2.5 pg/μg of RNA). Reactions with S1 nuclease were done in 50 mM sodium acetate (pH 4.5), 200 mM NaCl, 20 mM $MgCl_2$, and 10 mM $ZnSO_4$ and incubated for 1.5–6.0 min at 37 °C with S1 nuclease (15 units/μg of RNA), followed by the addition of 10 mM EDTA and 5 μg of tRNA. Reactions with alkali, RNase *Phy*I, RNase A, and S1 nuclease were terminated with an equal volume of 10 M urea, 0.1% bromphenol blue, and 0.1% xylene cyanol. The samples were electrophoresed on 20% polyacrylamide–50% urea sequencing gels. As previously described (Parker, 1983a), PEI–cellulose thin-layer chromatography was done with complete digests of RNA to chemically identify the 5′-terminal ribonucleotide.

To assess the effect of nucleotide concentrations on RNA synthesis, transcription assays were done in the manner described by Carpousis and Gralla (1980). The reactions contained 30 mM Tris·HCl (pH 8.0), 100 mM KCl, 4 mM $MgCl_2$, 0.1 mM EDTA, 0.2 mM dithiothreitol, 100 μg/mL bovine serum albumin, 0.8–1.2 pmol of RNA polymerase, and 0.4 pmol of DNA fragment. RNA polymerase–DNA complexes were allowed to form at 37 °C for 10 min, then heparin (125 μg/mL) was added, and the incubation was continued for 5 min. Transcription was begun by adding the indicated concentrations of ribonucleotides and 1 μCi of [α-^{32}P]UTP. Following a 10-min incubation at 37 °C, the reactions were terminated with 40 mM EDTA, 4 M urea, 0.05% bromphenol blue, and 0.05% xylene cyanol and directly electrophoresed on a 20% polyacrylamide–50% urea sequencing gel.

Transcription assays with dinucleotides were done in a manner similar to the method of Minkley and Pribnow (1973). The reactions contained 30 mM Tris·HCl (pH 8.0), 100 mM KCl, 4 mM $MgCl_2$, 0.1 mM EDTA, 0.2 mM dithiothreitol, 100 μg/mL bovine serum albumin, 200 μM dinucleoside monophosphate, 20 μM each of ATP, GTP, CTP, and UTP, 1 μCi of [α-^{32}P]UTP, 0.8 pmol of RNA polymerase, and 0.4 pmol of DNA fragment. Following a 15-min reaction at 37 °C, rifampicin (10 μg/mL) was added and incubation continued for 5 min. Finally, the reactions were treated with 400 μM each of ATP, GTP, CTP, and UTP and incubated for 15 min at 37 °C. The reactions were terminated and electrophoresed on a sequencing gel as described above.

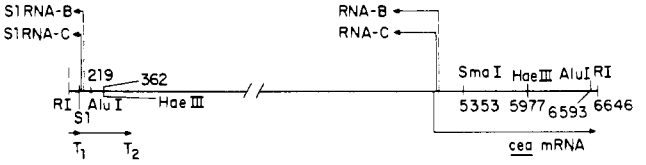

FIGURE 1: Coding regions for in vitro ColE1 transcripts that contain noncomplementary 5′-terminal ribonucleotides. A map of the ColE1 plasmid is shown with restriction sites listed by the nucleotide numbering system of Chan et al. (1985). The unique site in supercoiled ColE1 DNA that is cleaved by S1 nuclease is indicated (Lilley, 1980). Arrows denote directions of transcription, and dashed lines indicate transcripts that contain noncomplementary 5′-terminal ribonucleotides. The 5′-termini of the transcripts map to the following nucleotide positions: S1 RNA-C, 120; S1 RNA-B, 124; RNA-C, 5091; RNA-B, 5096; colicin E1 mRNA (*cea* mRNA), 5065. The ρ-dependent and ρ-independent termination sites for *cea* mRNA are denoted by T_1 and T_2, respectively (Ebina & Nakazawa, 1983).

RESULTS

An attempt was made to verify results that suggest two in vitro transcripts (about 31 and 37 bases long) initiate with 5′-terminal GTP residues that are not complementary to the ColE1 DNA template (Patient, 1979; R. Patient, personal communication). According to Patient's RNA and DNA sequence data, the *Eco*RI$_1$–*Alu*I$_{219}$ DNA fragment of ColE1 DNA (see Figure 1) contains overlapping coding sequences for these transcripts. Two small transcripts are identified following gel electrophoresis of [α-^{32}P]CTP-labeled transcription products from covalently closed circular and linear ColE1 DNA templates (Figure 2). These transcripts were also identified following transcription of ColE1 DNA in the presence of [γ-^{32}P]GTP (data not shown). The transcripts contain approximately 46 and 42 bases, and they are designated S1 RNA-B and S1 RNA-C, respectively.

Transcriptional mapping data indicate that S1 RNA-B and S1 RNA-C are coded within the *Eco*RI$_1$–*Alu*I$_{219}$ DNA fragment (data not shown). This DNA fragment provided a template for the preparative synthesis of S1 RNA-B and S1 RNA-C using [γ-^{32}P]GTP as the radiolabel. The 5′-terminal sequences and structures of these transcripts were analyzed by methods involving limited digestion with ribonucleases (Donis-Keller et al., 1977; Donis-Keller, 1980) and digestion with S1 nuclease (Wurst et al., 1978; Swerdlow & Guthrie, 1984). Analysis of data from the RNA sequencing gels (Figure 3A,B) indicates (1) pppGpUpUpU is present at the 5′ terminus of both S1 RNA-B and S1 RNA-C, (2) the entire sequence for S1 RNA-C overlaps the S1 RNA-B sequence, and (3) the transcripts contain an extensive, internal sequence that is protected from ribonuclease and S1 nuclease digestion. Internal bases in the transcripts were identified by the chemical method of Peattie et al. (1979) (data not shown).

The template's DNA sequence (data not shown) is aligned with the RNA sequences (Figure 4a) to indicate (1) the 5′-terminal ribonucleotide of S1 RNA-B does not exhibit complementarity with the DNA sequence, (2) the coding sequence for S1 RNA-B and S1 RNA-C contains the S1 nuclease site in ColE1 DNA, and (3) S1 RNA-B and S1 RNA-C begin synthesis at ColE1 nucleotide positions 124 and 120, respectively, and terminate synthesis at approximately position 78 (see Figure 1).

The DNA coding sequence for S1 RNA-B and S1 RNA-C does not agree with the sequence suggested by Patient's data

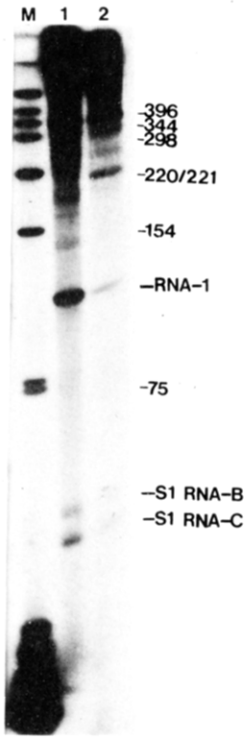

FIGURE 2: In vitro identification of two small ColE1 transcripts. The transcription products of covalently closed circular ColE1 DNA (lane 1) and *Sma*I-linearized ColE1 DNA (lane 2) include S1 RNA-B and S1 RNA-C, which are composed of approximately 46 and 42 bases, respectively. RNA-1 denotes a 108-base transcript, whose synthesis is preferentially stimulated from supercoiled ColE1 DNA (Levine & Rupp, 1978) and which functions to repress priming of ColE1 DNA synthesis (Tomizawa & Itoh, 1981). The [α-^{32}P]CTP-labeled transcripts and 5′ end-labeled pBR322–*Hin*fI size markers (lane M) were heat denatured in urea and electrophoresed on a 6% polyacrylamide–9 M urea gel. The number of bases contained in the size markers is listed in the margin.

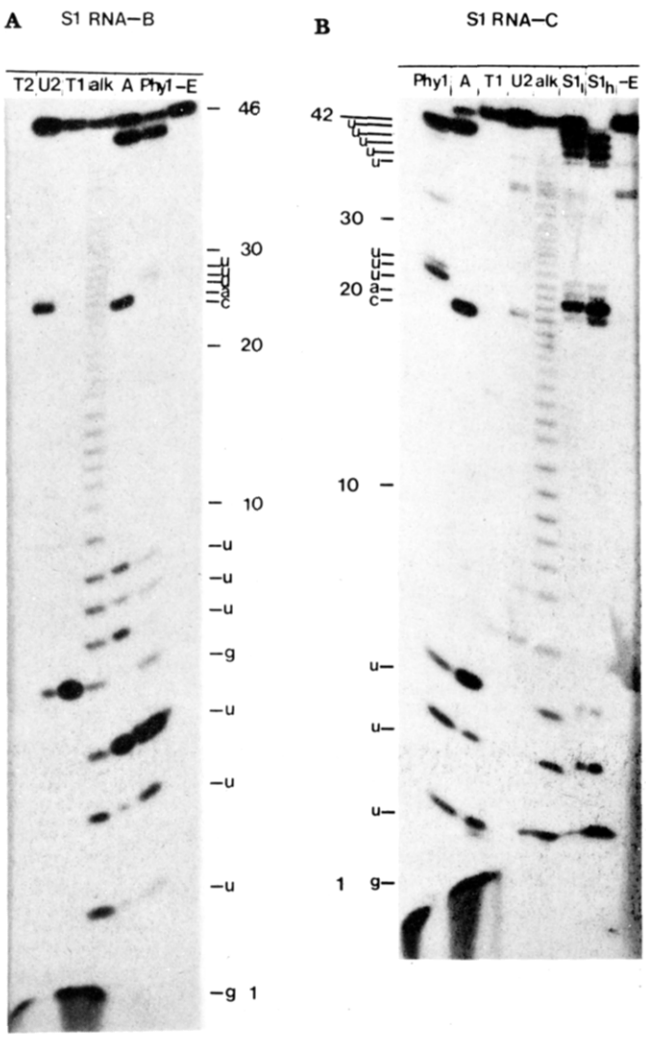

FIGURE 3: 5′-Terminal sequence determinations for S1 RNA-B and S1 RNA-C. The [γ-^{32}P]GTP-labeled transcripts were synthesized from the *Eco*RI$_1$–*Alu*I$_{219}$ DNA template (see Figure 1) and subsequently purified. Partial digests were done with the following: RNase T1, G specific (lanes T1); RNase U2, A specific (lanes U2); RNase A, C + U specific (lanes A); RNase *Phy*I, A + G + U specific (lanes Phy1); and alkali, nonspecific (lanes alk). Complete digestion to mononucleotides was done with a mixture of RNase T2, RNase T1, and RNase A (lanes T2). An unhydrolyzed sample of RNA is included (lanes –E). Lanes S1$_l$ and S1$_h$ contain S1 nuclease reactions for 1.5 and 6 min, respectively. The identities of bases in S1 RNA-B (A) and S1 RNA-C (B) are listed in the margins. The numbers represent base positions relative to the 5′-terminal nucleotide at position +1. The presence of GTP at the 5′ termini was established by PEI–cellulose thin-layer chromatography of transcripts that were hydrolyzed to mononucleotides (data not shown).

(Patient, 1979; R. Patient, personal communication) to code for two small, overlapping transcripts that initiate with noncomplementary 5′-terminal ribonucleotides. These sequences are located on opposite DNA strands, and the coding sequence assigned by Patient does not contain the S1 nuclease site. No small transcripts, other than S1 RNA-B and S1 RNA-C, were observed when transcription reactions employed conditions described by Patient (data not shown).

Transcription assays were done with variable nucleotide concentrations, since elevated ATP concentrations notably reduce the in vitro synthesis of RNA-B (Parker, 1983a,b), a ColE1 transcript that initiates with a noncomplementary 5′-terminal ribonucleotide (see Figure 4b). According to the method of Carpousis and Gralla (1980), *E. coli* RNA polymerase was preincubated with the *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment (see Figure 1), and then heparin was added to inactivate free RNA polymerase, followed by the addition of ribonucleotides to initiate transcription. To detect full-length transcripts and oligonucleotide products, the mixtures were loaded directly on sequencing gels. The data indicate that high concentrations of GTP (200–800 μM) notably and specifically reduce the relative synthesis of S1 RNA-B to S1 RNA-C (Figure 5A). The reduced relative synthesis is accompanied by a decrease in the production of several oligonucleotides, which persist following a chase with high concentrations of ribonucleotides (see Figure 5A, lanes 1–5).

Similar studies indicate that GTP concentrations between 10 and 800 μM produce pronounced effects on the relative synthesis of S1 RNA-B to S1 RNA-C, while little effect is observed on the net synthesis of these transcripts (Figure 5B)—suggesting that S1 RNA-B and S1 RNA-C synthesis may be linked. It is not surprising that S1 RNA-C synthesis increases as the GTP concentration is raised (Figure 5B), since the rate of RNA chain initiation is directly proportional to the concentration of the 5′-terminal nucleotide, up to fairly high concentrations (McClure et al., 1978; Nierman & Chamberlin, 1979); however, S1 RNA-B synthesis peaks at relatively low GTP concentration—suggesting that S1 RNA-B initiates synthesis by a nonconventional route. Only DNA templates that contain an intact promoter region for S1 RNA-B and S1 RNA-C yield the characteristic oligonucleotide products (Figure 5B, lanes 1–4). These oligonucleotides are produced in greatest abundance when S1 RNA-B synthesis is optimized at 20 μM GTP (Figure 5B, lane 7).

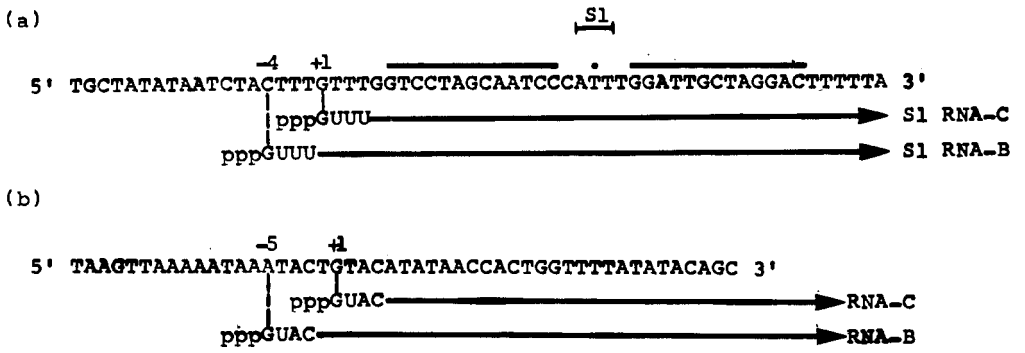

FIGURE 4: DNA sequences that code for ColE1 transcripts involved in the noncomplementary initiation phenomenon. The 5′-terminal sequences for S1 RNA-B and S1 RNA-C (a) are aligned with the sequence of their sense strand in ColE1 DNA. The DNA sequence of the coding region (see Figure 1) was determined by the method of Maxam and Gilbert (1980) (data not shown). The start point for S1 RNA-C at +1 codes for a 5′-terminal GTP residue that is complementary to the DNA template, while the 5′-terminal GTP residue of S1 RNA-B is not complementary to the −4 position. The initial four nucleotides in S1 RNA-B and S1 RNA-C are identical (pppGpUpUpU). The coding sequence contains an extensive, inverted repeat (see overlining). The bracketed S1 symbol denotes the site in supercoiled ColE1 DNA that is cleaved by S1 nuclease (Lilley, 1980). Similarly, the 5′-terminal sequences of RNA-B and RNA-C (b) are aligned with the sequence of their sense strand (see Figure 1) in ColE1 DNA (Parker, 1983a). The start point for RNA-C at +1 codes for a 5′-terminal GTP residue that is complementary to the DNA template, while the 5′-terminal GTP residue of RNA-B is not complementary to the −5 position. The initial four bases in RNA-B and RNA-C are identical (pppGpUpApC).

To further examine the relationship between the synthesis of S1 RNA-B and S1 RNA-C, transcription assays were conducted with dinucleotides that are complementary to regions near the transcription start points. Usually, *E. coli* RNA polymerase initiates an RNA chain with dinucleotides that are complementary to the DNA sequence at the start site for transcription, although some cases exist where noncomplementary dinucleotides prime transcription (Downey & So, 1970; Minkley & Pribnow, 1973; Dausse et al., 1975; Dunn & Studier, 1981). According to the method of Minkley and Pribnow (1973), the *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment was transcribed with RNA polymerase in the presence of nucleoside triphosphates (20 μM each) and various dinucleotides (200 μM each), and then rifampicin was added to prevent reinitiation, followed by a chase with nucleoside triphosphates. In a control reaction, the 5′-terminal phosphates of S1 RNA-B and S1 RNA-C were removed by phosphatase treatment to enable a comparison of the electrophoretic mobilities of these transcripts with the dinucleotide initiated products. The data (Figure 6) indicate that dinucleotides (UpU, UpG, GpU), which are expected to prime transcription only from the S1 RNA-C start site (see Figure 4a), initiate the synthesis of two major transcripts that appear to be changed in length by the same number of bases and which correlate with transcripts from *both* the S1 RNA-C and S1 RNA-B start sites. With the possible exception of ApC, dinucleotides that are complementary to the DNA sequence at the S1 RNA-B start site were generally ineffective at priming transcription of S1 RNA-B. The assay suggests that the synthesis of S1 RNA-B is directly dependent upon the synthesis of S1 RNA-C.

Discussion

The results characterize the in vitro synthesis of two small transcripts, S1 RNA-B and S1 RNA-C, which initiate within four bases of each other and contain a coding region that forms the S1 nuclease site in supercoiled ColE1 DNA. In vitro, this region contains a bidirectional terminator, comprised of a ρ-dependent terminator for colicin E1 mRNA (Ebina & Nakazawa, 1983) and a ρ-independent terminator for S1 RNA-B and S1 RNA-C (see Figure 1).

This study indicates that the 5′-terminal ribonucleotide of S1 RNA-B does not exhibit complementarity with the sequence of the DNA template. This observation also applies to the RNA-B transcription product of ColE1 DNA (Parker, 1983a). No evidence was obtained to substantiate that other transcripts initiate with noncomplementary 5′-terminal ribonucleotides from a region near the S1 nuclease site in ColE1 DNA (Patient, 1979; R. Patient, personal communication). A comparison of sequence data for ColE1 transcripts involved in the noncomplementary initiation phenomenon (see Figure 4) reveals (1) transcripts with noncomplementary 5′-terminal ribonucleotides begin 4–5 base pairs upstream from transcripts that initiate with complementary ribonucleotides, (2) the initial four bases in S1 RNA-B and S1 RNA-C are identical (pppGpUpUpU), as are the initial four bases in RNA-B and RNA-C (pppGpUpApC), and (3) the noncomplementary 5′-terminal GTP residues in S1 RNA-B and RNA-B align, respectively, with a guanylate and a thymidylate residue on the DNA coding strand—hence, "wobble-like" base pairing schemes cannot explain the initiation phenomenon.

A model involving transcriptional slippage is suggested to explain the noncomplementary initiation phenomenon. The model proposes that a pppGpUpUpU oligonucleotide, synthesized during the cycling reaction at the S1 RNA-C promoter, is transposed by *E. coli* RNA polymerase (without dissociation of the enzyme from the template) to a position four bases upstream for priming the synthesis of S1 RNA-B (see Figure 4a). Similarly, a pppGpUpApC oligonucleotide, produced at the RNA-C promoter, is envisioned to prime transcription of RNA-B (see Figure 4b). In both cases, three out of four nucleotides in the proposed primers are complementary to the DNA template at sites that initiate synthesis of S1 RNA-B and RNA-B. The priming event may resemble transcriptional priming by dinucleotides that are not exactly complementary to DNA sequences at some transcription start sites in phage T7 DNA (Minkley & Pribnow, 1973; Dausse et al., 1975; Dunn & Studier, 1981).

The slippage model is supported by the observation that low concentrations of GTP result in increased relative yields of S1 RNA-B to S1 RNA-C (Figure 5). This effect can be understood if one postulates that RNA polymerase hesitates after synthesizing the initial four bases in S1 RNA-C; at that point, a GTP residue is scheduled for condensation (see Figure 4a). A low GTP concentration is envisioned to decrease the ability of the ternary complex (DNA–RNA polymerase–pppGpUpUpU) to translocate and bind GTP, thereby enhancing abortive synthesis of pppGpUpUpU and increasing the relative synthesis of S1 RNA-B to S1 RNA-C. Consistent with this view is the observation that low GTP concentrations yield enhanced synthesis of abortive oligonucleotides (Figure 5),

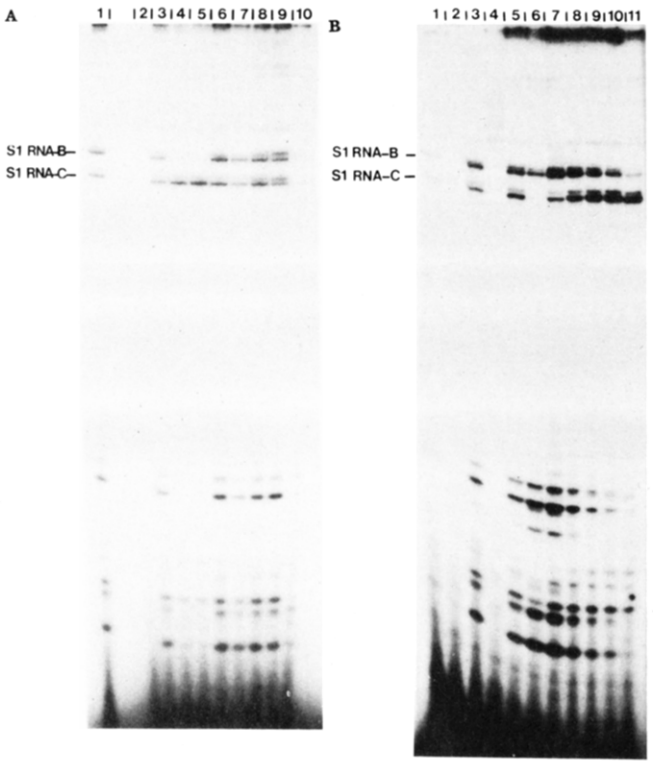

FIGURE 5: Effect of ribonucleotide concentrations on the synthesis of S1 RNA-B and S1 RNA-C. The reaction and electrophoresis conditions were according to the method of Carpousis and Gralla (1980), as described under Experimental Procedures. (A) Reactions included 1.2 pmol of RNA polymerase and 0.4 pmol of *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment (see Figure 1). Lane 1, reaction with a 50 μM aliquot of each ribonucleotide (ATP, GTP, CTP, and UTP) that was completed by a 5-min chase with a 800 μM aliquot of each ribonucleotide. Lane 2, reaction without DNA. Lane 3, reaction with a 50 μM aliquot of each ribonucleotide. Other lanes contain reactions identical with that of lane 3, with the following variations: lane 4, 200 μM GTP; lane 5, 800 μM GTP; lane 6, 200 μM CTP; lane 7, 800 μM CTP; lane 8, 200 μM ATP; lane 9, 800 μM ATP; lane 10, 200 μM UTP. The autoradiograph indicates that the S1 RNA-B and S1 RNA-C transcripts are each composed of two bands, probably arising from heterogeneous 3′ termini. Abortive oligonucleotides are shown in the lower half of the figure. (B) Reactions included 0.8 pmol of RNA polymerase and 0.4 pmol of DNA fragment. Lanes 1–4, reactions with a 50 μM sample of each ribonucleotide and the following DNA templates: lane 1, *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment digested with *Alu*I; lane 2, *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment digested with *Ava*II; lane 3, *Alu*I$_{6593}$–*Alu*I$_{219}$ DNA fragment; lane 4, *Hae*III$_{5977}$–*Ava*II$_{115}$ DNA fragment (*Ava*II cleaves between +5 and +6 in the coding region for S1 RNA-C) (see Figures 1 and 4a). Lane 5, reaction with the *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment and a 50 μM aliquot of each ribonucleotide, followed by a 10-min chase with a 1 mM sample of each ribonucleotide. Other lanes contain reactions with the *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment and 50 μM each of ATP, CTP, and UTP with the following GTP concentrations: lane 6, 10 μM GTP; lane 7, 20 μM GTP; lane 8, 50 μM GTP; lane 9, 100 μM GTP; lane 10, 200 μM GTP; lane 11, 800 μM GTP. Transcripts and abortive oligonucleotides are identified as in (A).

produced under conditions that demonstrated the cycling reaction for *E. coli* RNA polymerase (Carpousis & Gralla, 1980). This idea also explains the observation that low ATP concentrations increase the relative synthesis of RNA-B to RNA-C (Parker, 1983a,b); in this case, ATP is scheduled for condensation after synthesis of the proposed primer, pppGpUpApC, at the RNA-C promoter (see Figure 4b). The analysis is supported by observations that nucleotide concentrations affect the strength of RNA polymerase pausing at specific DNA regions (Maizels, 1973; Darlix & Horaist, 1976; Kassavetis & Chamberlin, 1981); especially noteworthy is the finding that the strength of a pause site is increased if the concentration of one of the nucleotides contained in the site is lowered (Reisbig & Hearst, 1981).

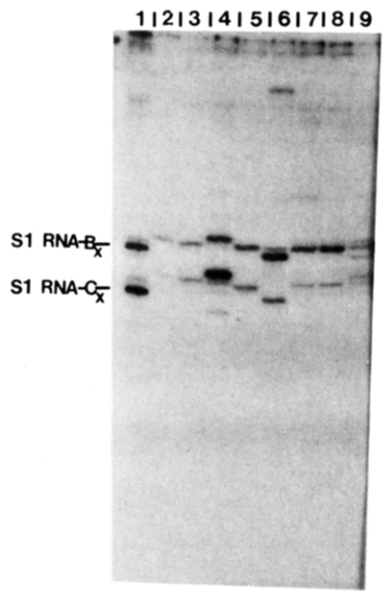

FIGURE 6: Effect of dinucleotides on priming synthesis of S1 RNA-B and S1 RNA-C. As described under Experimental Procedures, reactions were done with 0.4 pmol of *Eco*RI$_1$–*Hae*III$_{362}$ DNA fragment, 0.8 pmol of RNA polymerase, 20 μM each of ATP, GTP, CTP, and UTP, and the following additions: lane 2, no dinucleotides; lane 3, 30 μM GTP (therefore, 50 μM GTP final concentration); lane 4, 200 μM UpU; lane 5, 200 μM UpG; lane 6, 200 μM GpU; lane 7, 200 μM UpA; lane 8, 200 μM ApC; lane 9, 200 μM CpU. A reaction containing 50 μM GTP and 20 μM each of ATP, CTP, and UTP was subsequently treated with alkaline phosphatase to remove 5′-terminal phosphates from S1 RNA-B and S1 RNA-C (lane 1). The phosphatase-treated, major transcripts in lane 1 are denoted as S1 RNA-B$_x$ and S1 RNA-C$_x$ in the margin.

The slippage model is supported by transcription assays with dinucleotides (Figure 6), since transcripts correlated to both S1 RNA-B and S1 RNA-C are primed by dinucleotides that are complementary to bases at and adjacent to the transcription start point for S1 RNA-C—suggesting that the S1 RNA-C promoter is responsible for initiating synthesis of both transcripts.

The data suggest that the noncomplementary initiation phenomenon is mediated in vitro by a slippage reaction, involving priming of transcription from nearby upstream sequences by oligonucleotides produced in the cycling reaction of *E. coli* RNA polymerase. This phenomenon may not occur to any significant extent in vivo, since it should be suppressed by the high concentrations of ribonucleotides (1–4 mM) in *E. coli* (Cashel & Gallant, 1968; Moses & Sharp, 1972; Beck et al., 1973). Electrophoretic analysis of [^{32}P]orthophosphate-labeled RNA from a colicinogenic culture reveals no transcripts with sizes comparable to S1 RNA-B and S1 RNA-C (data not shown). Also, primer extension and S1 mapping analysis of RNA from colicinogenic cells fail to identify RNA-B, despite the presence of RNA-C in vivo (unpublished experiments).

ACKNOWLEDGMENTS

I thank Dr. John Richardson for several helpful discussions regarding this study.

Registry No. GTP, 86-01-1; UpU, 2415-43-2; UpG, 3474-04-2; GpU, 4785-07-3; UpA, 3256-24-4; ApC, 4833-63-0; CpU, 2382-64-1; RNA polymerase, 9014-24-8.

REFERENCES

Beck, K. C., Ingraham, J., Maaloe, O., & Neuhard, J. (1973) *J. Mol. Biol. 78*, 117.

Burgess, R. R., & Jendrisak, J. J. (1975) *Biochemistry 14*, 4634.
Carpousis, A. J., & Gralla, J. D. (1980) *Biochemistry 19*, 3245.
Cashel, M. T., & Gallant, J. (1968) *J. Mol. Biol. 34*, 317.
Chamberlin, M. J. (1976) in *RNA Polymerase* (Losick, R., & Chamberlin, M., Eds.) p 17, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Chamberlin, M., & Berg, P. (1962) *Proc. Natl. Acad. Sci. U.S.A. 48*, 81.
Chamberlin, M., & Berg, P. (1964) *J. Mol. Biol. 8*, 708.
Chan, P. T., Ohmori, H., Tomizawa, J., & Lebowitz, J. (1985) *J. Biol. Chem. 260*, 8925.
Cowie, A., Jat, P., & Kamen, R. (1982) *J. Mol. Biol. 159*, 225.
Dahlberg, J. E., & Blattner, F. R. (1973) in *Virus Res., ICN–UCLA Symp. Mol. Biol., Proc., 2nd*, 533.
Darlix, J., & Horaist, M. (1976) *Nature (London) 256*, 288.
Dausse, J.-P., Sentenac, A., & Fromageot, P. (1975) *Eur. J. Biochem. 57*, 569.
Donis-Keller, H. (1980) *Nucleic Acids Res. 8*, 3133.
Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527.
Downey, K. M., & So, A. G. (1970) *Biochemistry 9*, 2520.
Dunn, J. J., & Studier, F. W. (1981) *J. Mol. Biol. 148*, 303.
Ebina, Y., & Nakazawa, A. (1983) *J. Biol. Chem. 258*, 7072.
Falaschi, A., Adler, J., & Khorana, H. G. (1963) *J. Biol. Chem. 238*, 3080.
Gilbert, W. (1976) in *RNA Polymerase* (Losick, R., & Chamberlin, M., Eds.) p 193, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Hansen, U. M., & McClure, W. R (1980) *J. Biol. Chem. 255*, 9564.
Kassavetis, G. A., & Chamberlin, M. J. (1981) *J. Biol. Chem. 256*, 2777.
Levine, A. D., & Rupp, W. D. (1978) in *Microbiology—1978* (Schlessinger, D., Ed.) p 163, American Society of Microbiology, Washington, D.C.
Lilley, D. M. J. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6468.
Maizels, N. M. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3585.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499.
McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171.
McClure, W. R., Cech, C. L., & Johnson, D. E. (1978) *J. Biol. Chem. 253*, 8941.
Minkley, E. G., & Pribnow, D. (1973) *J. Mol. Biol. 77*, 255.
Moses, V., & Sharp, P. B. (1972) *J. Gen. Microbiol. 71*, 181.
Munson, L. M., & Reznikoff, W. S. (1981) *Biochemistry 20*, 2081.
Nierman, W. C., & Chamberlin, M. J. (1979) *J. Biol. Chem. 254*, 7921.
Nishimura, S., Jacob, T. M., & Khorana, H. G. (1964) *Proc. Natl. Acad. Sci. U.S.A. 52*, 1494.
Paetkau, V., Coulter, M. B., Flintoff, W. G., & Morgan, A. R. (1972) *J. Mol. Biol. 71*, 293.
Parker, R. C. (1983a) *Gene 26*, 127.
Parker, R. C. (1983b) Ph.D. Dissertation, Yale University, New Haven, CT.
Patient, R. K. (1979) *Nucleic Acids Res. 6*, 2647.
Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1760.
Reisbig, R., & Hearst, J. E. (1981) *Biochemistry 20*, 1907.
Swerdlow, H., & Guthrie, C. (1984) *J. Biol. Chem. 259*, 5197.
Tomizawa, J., & Itoh, T. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6096.
Travers, A. A., & Burgess, R. R. (1969) *Nature (London) 222*, 537.
Walseth, T. F., & Johnson, R. A. (1979) *Biochim. Biophys. Acta 562*, 11.
Wurst, R. M., Vournakis, J. N., & Maxam, A. M. (1978) *Biochemistry 17*, 4493.

Site Specificity of Psoralen–DNA Interstrand Cross-Linking Determined by Nuclease *Bal*31 Digestion†

Wei-ping Zhen,‡ Ole Buchardt,§ Henrik Nielsen,‡ and Peter E. Nielsen*,‡

Department of Biochemistry B, The Panum Institute, University of Copenhagen, DK-2200 Copenhagen N, Denmark, and Chemical Laboratory II, The H. C. Ørsted Institute, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

Received April 29, 1986; Revised Manuscript Received July 7, 1986

ABSTRACT: A novel method for determination of psoralen photo-cross-linking sites in double-stranded DNA is described, which is based on a pronounced inhibition of *Bal*31 exonuclease activity by psoralen–DNA interstrand cross-links. The results using a 51 base pair fragment of plasmid pUC19 and a 346 base pair fragment of pBR322 show that 5′-TA sequences are preferred cross-linking sites compared to 3′-TA sequences. They also indicate that sequences flanking the 5′-TA site influence the cross-linking efficiency at the site. The DNA photo-cross-linking by 4,5′,8-trimethylpsoralen and 8-methoxypsoralen was analyzed, and these two psoralens showed identical site specificity. The 5′-TA preference is rationalized on the basis of the local DNA structure in terms of the π–π electronic interaction between the thymines and the intercalated psoralens, as well as on the base tilt angles of the DNA.

It is generally assumed that the light-induced formation of psoralen–DNA adducts, and especially DNA interstrand cross-links, is intimately associated with many of the photobiological effects exhibited by the psoralens [see Ben-Hur & Song (1984) for a review]. These adducts may also be highly relevant when discussing the beneficiary effects (plaque clearing) as well as the unwarranted side effects (e.g., skin phototoxicity and the suspected carcinogenicity) in the therapeutical applications of these compounds (Andersen &

† Supported by the Danish Natural Science Research Council. P.E.N. is a "Niels Bohr Fellow" through the Egmont Foundation.
* Correspondence should be addressed to this author.
‡ The Panum Institute, University of Copenhagen.
§ The H. C. Ørsted Institute, University of Copenhagen.

0006-2960/86/0425-6598$01.50/0