In Vitro Studies of Transcript Initiation by *Escherichia coli* RNA Polymerase. 1. RNA Chain Initiation, Abortive Initiation, and Promoter Escape at Three Bacteriophage Promoters†

Lilian M. Hsu,*,‡ Nam V. Vo,§,|| Caroline M. Kane,§ and Michael J. Chamberlin§

Program in Biochemistry, Mount Holyoke College, South Hadley, Massachusetts 0l075-6456, and Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3202

Received October 7, 2002; Revised Manuscript Received January 27, 2003

ABSTRACT: RNA chain initiation and promoter escape is the latter stage of transcription initiation. This stage is characterized by several well-defined biochemical events: synthesis and release of short RNA products ranging 2 to 15 nucleotides in length, release of the σ subunit from the enzyme−promoter complex, and initial translocation of the polymerase away from the promoter. In this paper, we report the use of a steady-state transcription assay with [γ-^{32}P]ATP labeling to subject the RNA chain initiation−promoter escape reaction to quantitative analysis. The specific parameters we follow to describe the chain initiation−promoter escape process include the abortive and productive rates, the abortive probability, the abortive:productive ratio, and the maximal size of the abortive product. In this study, we measure these parameters for three bacteriophage promoters transcribed by *Escherichia coli* RNA polymerase: T7 A1, T5 N25, and T5 N25$_{antiDSR}$. Our studies show that all three promoters form substantial amounts of abortive products under all conditions we tested. However, each of the promoters shows distinct differences from the others when the various parameters are compared. At 100 μM NTP, in a 10 min reaction, the abortive and productive yields are 87 and 13%, respectively, for T7 A1; 97 and 3%, respectively, for T5 N25; and 99.4 and 0.6%, respectively, for T5 N25$_{antiDSR}$. These values correspond to approximately 7, 32, and 165 abortive transcripts per productive transcript for the three promoters, respectively. The yield of most of the abortive products is not affected by the elevated concentration of the NTP substrate corresponding to the next template-specified nucleotide; hence, abortive products are not normally formed through a simple process of "kinetic competition". Instead, formation of abortive products appears to be determined by intrinsic DNA signals embedded in the promoter recognition region and the initial transcribed sequence region of each promoter.

The efficiency of promoter function is dependent on its interaction with RNA polymerase at two distinct stages of transcription initiation: (1) promoter binding and activation and (2) RNA chain initiation and promoter escape. The former has been studied extensively in both prokaryotic and eukaryotic organisms and involves interaction of the RNA polymerase with general transcription factors, promoter specific factors, and DNA sequences of the recognition regions of promoters (*1*−*5*). For *Escherichia coli* RNA polymerase, the promoter binding−activation phase of transcription leads to the formation of the open promoter complex, the essential intermediate that interacts with the NTP substrates to begin actual transcription.

The second stage of transcription initiation (RNA chain initiation and promoter escape) involves synthesis by the open promoter complex of short RNA transcripts from 2 to 15 nucleotides (nt)[1] in length, yielding transient intermediates which we shall call initial transcribing complexes (ITCs). The nascent RNA in the ITCs can be further elongated, or can be released to form what are termed abortive transcripts (*6*−*8*). During abortive transcription, RNA polymerase molecules do not usually dissociate from the promoter (*7*, *9*), although this depends entirely on the actual stability of the open promoter complex (*10*, *11*). Having released the nascent RNA, the initial transcribing complex reassumes the open complex conformation and can undergo reinitiation (*12*). The footprints of the ITC along the DNA strands are quite similar to those of the open promoter complex with only minor differences in the downstream boundary (*9*). Completion of the RNA chain initiation−promoter escape phase of transcription is signaled by several distinct biochemical changes in the enzyme−promoter complex: relinquishing the upstream promoter contacts by RNA polymerase, release of the σ subunit from the complex (see refs

† This work was supported by NSF Grant 0077941 and NIH Grant GM 49478 to L.M.H. and NIH Grant GM 12010 to M.J.C.

* To whom correspondence should be addressed. Phone: (413) 538-2609. Fax: (413) 538-2327. E-mail: lhsu@mtholyoke.edu.

‡ Mount Holyoke College.

§ University of California.

|| Present address: Department of Molecular Microbiology and Immunology, University of Southern California Keck School of Medicine, Los Angeles, CA 90033-1054.

[1] Abbreviations: nt, nucleotide(s); ITC, initial transcribing complex; EC, elongation complex; ITS, initial transcribed sequence; bp, base pair(s); PCR, polymerase chain reaction; APR, abortive:productive ratio; MSAT, maximum size of the abortive transcript.

10.1021/bi026954e CCC: $25.00
Published on Web 03/14/2003

13 and *14* for a discussion of the σ release controversy), translocation of the polymerase downstream of the promoter, and formation of a highly stable elongation complex (EC) in which the RNA is tightly held (*15*).

The chain initiation and promoter escape phase of transcription directly affects the efficiency of promoter function and, therefore, promoter strength. In addition, these processes can be regulated by DNA sequences and protein factors that are distinct from those that regulate promoter binding and activation. Kammerer et al. (*16*) first reported several examples where replacement of the first 20 nt of the transcript (the initial transcribed sequence, or ITS) affects the rate of promoter escape from several strong promoters both *in vitro* and *in vivo*. Interference with promoter escape can dramatically weaken promoters that are otherwise highly efficient in the promoter binding and activation steps. The reduced rate of escape from one such promoter can be overcome *in vitro* and *in vivo* by transcript cleavage factors GreA and GreB (*17*). Regulation of promoter escape has also been demonstrated for the *E. coli malT* promoter where CRP protein activates the promoter by facilitating promoter escape (*18*, *19*). At the *E. coli pyrB* promoter, the ITS imposes UTP concentration-dependent slippage of nascent transcripts as a means of regulating promoter escape (*20*, *21*). Regulation of promoter escape may also be involved in the promoter proximal pausing of RNA polymerase II (*22*−*26*). Thus, while promoter escape is controlled in part by the template DNA sequence, it can also be regulated by small effector molecules and protein factors.

The synthesis of abortive transcripts at promoters in the presence of all four regular NTPs is still a particularly puzzling reaction. It is quite evident that at different promoters there can be considerable differences in the abortive yield, defined as the percentage of abortive products among total products (*15*, *27*). In addition, with different promoters, there are significant differences in the molar yields of different sizes of abortive transcripts (*7*−*9*, *28*), an aggregate parameter that we call abortive pattern (*15*). Finally, there are marked differences in the maximum size of abortive products made from different promoters, ranging from 8 to ≥15 nt. This parameter appears to correlate in some manner with the position at which the σ subunit dissociates from the complex (*9*, *29*, *30*).

A number of important questions remain unanswered with regard to the synthesis of abortive products. Do all *E. coli* promoters form significant amounts of abortive transcripts *in vitro* and *in vivo* in the presence of all four nucleoside triphosphates? Do *in vitro* reaction conditions that eliminate abortive initiation exist? What DNA sequences control the rate of promoter escape, the abortive yield, the abortive pattern, and the maximum size of abortive transcripts?

To address these questions, we examined some of the parameters of RNA chain initiation and abortive synthesis using three bacteriophage promoters, T7 A1, T5 N25, and T5 $N25_{antiDSR}$. The T7 A1 and T5 N25 sequences represent two of the strongest promoters utilized by *E. coli* RNA polymerase Eσ^{70} (*31*); while T7 A1 appears to bind RNA polymerase more rapidly than T5 N25, T5 N25 forms a more stable open promoter complex (*32*, *33*). Finally, T5 $N25_{antiDSR}$ was formed by replacing the ITS of T5 N25, from position 3 to 20, with an antisense sequence; this change in the ITS has a dramatic negative effect on promoter escape. In fact, this template provides the best current example of a strong promoter rendered weak by a limitation of the promoter escape reaction.

MATERIALS AND METHODS

Materials. Plasmid pAR1707 was obtained from A. Rosenberg (*34*); pDS3/PN25 and pDS3/$PN25_{antiDSR}$ were from H. Bujard (*31*). RNA polymerase holoenzyme, 100% saturated with the σ^{70} subunit, was isolated from *E. coli* DG156 cells according to the procedure of Burgess and Jendrisak (*35*) as modified by Gonzalez et al. (*36*). At the time of use, the enzyme preparations contained ∼30−60% active molecules (*37*).

In Vitro Transcription. Template DNA fragments of approximately 200 base pairs (bp) containing single promoters were prepared by polymerase chain reactions (PCR) as described previously (*27*). Transcription reactions were performed under steady-state conditions in a 10−50 μL volume containing 20−100 nM template DNA and an equimolar amount of RNA polymerase in 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 10 μg/mL acetylated BSA. The KCl concentration was optimized for each promoter to maximize the yield of productive RNA and to permit extensive RNA synthesis (*38*). HPLC-grade NTP was used at 100 μM, unless specified otherwise. For quantitation studies, transcripts were labeled with [γ-^{32}P]ATP at ∼10 cpm/fmol. For the nearest-neighbor analysis, individual [α-^{32}P]NTP was also used at ∼10 cpm/fmol. Reactions were routinely started with the addition of RNA polymerase, and mixtures were incubated at 37 °C for varying lengths of time and analyzed by denaturing polyacrylamide gel electrophoresis as described previously (*27*). After electrophoresis, the gels were exposed directly to a phosphorimager screen and scanned using a Molecular Dynamics model 425E instrument. To obtain a factor for conversion between phosphorimager counts and (femto)moles of RNA, known scintillation counts of the labeling nucleotide of a given specific activity were spotted on the gel and exposed to the phosphor screen for the same length of time.

Nearest-Neighbor Analysis. Nearest-neighbor analysis (*39*) was carried out to verify the identity of the short oligonucleotide transcripts. Briefly, a single band of [α-^{32}P]NMP-labeled transcript was excised from the gel and incubated overnight at 37 °C in 0.33 M KOH. The alkaline extract was recovered and neutralized with 60% perchloric acid (1 μL per 25 μL of KOH extract), cooled on ice, and spun for 15 min at 4 °C to pellet the insoluble $KClO_4$ precipitate. The clear supernatant was brought to dryness in a rotary desiccator under vacuum. The residue was redissolved in distilled H_2O and analyzed by high-voltage paper electrophoresis in 50 mM sodium citrate buffer (pH 3.5) against known nucleoside 3′-monophosphate standards.

RESULTS

Rationale. Our immediate goal in these studies was to establish several quantitative and qualitative parameters for the RNA chain initiation reaction using the three promoters T7 A1, T5 N25, and T5 $N25_{antiDSR}$, under a variety of *in vitro* reaction conditions. The sequences of these promoters, showing the promoter recognition (positions −35 and −10)

FIGURE 1: Sequences of the three promoters used in these studies. Linear duplex DNA templates were prepared by PCR amplification such that the runoff RNA size is ~50 nucleotides long. Actual lengths of the individual DNA templates are as follows: positions −149 to 50 for T7 A1, positions −160 to 50 for T5 N25, and positions −160 to 64 for T5 $N25_{antiDSR}$. The sequence of the nontemplate promoter DNA strand is shown with the −35 and −10 consensus elements and the +1 transcription start site indicated in bold. The 18 bp substitution yielding the initial transcribed sequence of T5 $N25_{antiDSR}$ is shaded.

and ITS regions, are shown in Figure 1. The parameters we have followed for each promoter include the abortive rate, the productive rate, the abortive:productive ratio, and the maximum size of aborted transcripts. These experiments were performed to reveal whether abortive products are always found, or only under some conditions, and whether different reaction conditions can alter the abortive yield and the abortive pattern.

Our experimental system employs [γ-^{32}P]ATP to 5′-end label all nascent RNAs formed in the RNA chain initiation reaction. The DNA templates that were employed were truncated using PCRs to generate short runoff transcripts from the different promoters (ca. 50 nt). This approach allows us to resolve both abortive and productive transcripts on a single polyacrylamide gel and, therefore, to obtain direct measurement of the amount of each product from the same gel.

An important feature of these experiments involves the use of extensive synthesis conditions for *E. coli* RNA polymerase (*38*, *40*). In such reactions, there is efficient recycling of RNA polymerase through many rounds of the transcription cycle. As a consequence, the only major products are abortive transcripts and completed productive transcripts, and few or no intermediates (i.e., paused transcripts) are detected. Another benefit of the extensive synthesis reaction is the relative ease in obtaining high levels of radioactive products for analysis. A preliminary description of this steady-state assay system has been reported (*27*), although in this paper, we have redefined the quantitative initiation parameters more rigorously.

Optimizing the Solution Conditions for Productive Initiation. We have varied the concentrations of several different reaction components and have determined the abortive pattern and abortive yield for each reaction. In particular, the concentrations of KCl (0−500 mM) and $MgCl_2$ (1−20 mM) were titrated, and the effect of replacing KCl with potassium acetate (KAc) or potassium glutamate (KGlu) was tested over the same concentration range (*41*). Variation of the anion in potassium salts led to significant differences in the total amounts of transcription products, but did not affect the abortive pattern or abortive yield (data not shown). Compared to transcription in KCl which is set at 100%, KAc gave ~20% and KGlu ~10% total products. Because of the greater yields obtained with KCl, we used this salt throughout our studies.

For further studies, it was important to optimize the KCl concentration for each promoter. At lower KCl concentrations, minor amounts of transcripts corresponding to initiation at the ends of the DNA were seen; these disappeared at high KCl concentrations. Optimal KCl concentrations for the three promoters were as follows: 190 mM for T7 A1, 250 mM for T5 N25, and 150 mM for T5 $N25_{antiDSR}$. The unusually high KCl concentration used for T5 N25 transcription suppressed the formation of spurious long RNAs without altering the basic pattern of abortive versus productive synthesis obtained at 150 mM.

Is Abortive Initiation the Result of Kinetic Competition? Abortive transcripts are formed during the early steps of transcription initiation. In the simplest treatment of abortive initiation, one might suppose that there is a direct competition between the addition of an NMP residue to a nascent RNA and the release of that nascent RNA as an aborted product (*1*, *15*). In this "kinetic competition" model for RNA chain initiation, it might be expected that the abortive yield would decrease, with increasing concentrations of all four NTPs (*42*). Further, it might be expected that the level of formation of a particular abortive product would decrease, as the concentration of the particular nucleoside triphosphate substrate providing the next NMP residue is increased. Thus, in this simple model, both the abortive yield and abortive pattern should be altered by proper variations of the NTP concentration.

In the initial experiment, all four NTPs were varied together in concentrations from 10 to 100 μM (Figure 2). Interestingly, as the NTP concentration is increased, there is an increase in the amount of both abortive and productive transcripts; this is contrary to the first prediction of the kinetic competition model. Furthermore, a distinct abortive pattern is seen for each of the three promoters at an NTP concentration that is high enough to yield the full set of abortive RNA (ca. 100 μM). The abortive ladders derived from each of the three promoters also differed in the maximum size of detected abortive transcripts, ranging from 8 nt for T7 A1 and 10 nt for T5 N25 to 15 nt for T5 $N25_{antiDSR}$. In addition, there are substantial differences in the size of the prominent abortive products, with 2- and 3-mers being in excess for T7 A1, 2-, 4-, and 8-mers for T5 N25, and 4-, 6-, 7-, 13-, and 14-mers for T5 $N25_{antiDSR}$.

Characterization of Abortive Transcripts from the Three Promoters. Formation of short transcripts up to 14 or 15 nt does not, in itself, indicate that abortive initiation is occurring at a promoter, nor does it ensure that these transcripts are accurate copies of the template sequence. This caveat is pertinent because *E. coli* RNA polymerase can pause in the region of positions 15−25 to yield transcripts that are considered paused products, despite the fact that the σ subunit has not been released from these promoters (*30*). Additionally, the phenomenon known as "primer shifting" can lead

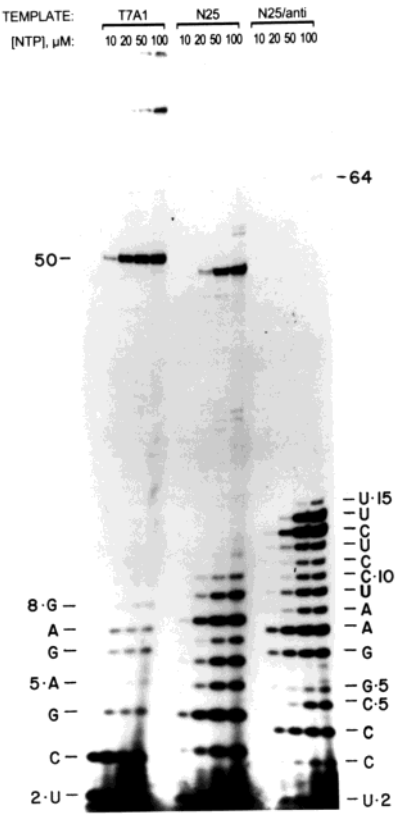


FIGURE 2: Productive and abortive initiation as a function of NTP concentration. Each promoter was transcribed in reaction mixtures containing the four NTPs (10, 20, 50, and 100 μM each). Recovered transcripts were fractionated in a denaturing 20% (19:1) polyacrylamide gel in an electrolyte gradient buffer (*71*). Each abortive transcript is designated by its size and the identity of its 3′-terminal nucleotide. The size of the abortive products from T7 A1 is indicated along the left border and that of T5 $N25_{antiDSR}$ along the right border. For T5 $N25_{antiDSR}$, two 5 nt abortive products were identified (see the text). The G5 product is the true abortive product specified by the template sequence, and C5 is a slippage product (pppAUCCC) produced in the initiation reaction. The productive RNA for T7 A1 and T5 N25 is 50 nucleotides in length, and that for T5 $N25_{antiDSR}$ is 64 nucleotides in length.

to longer abortive products that are no longer precisely complementary to the initial transcribed region of the template sequence (*43*−*45*). These findings prompted us, especially in the case of T5 $N25_{antiDSR}$ where abortive products extend to 15 nt, to obtain further verification as to whether the short RNAs were true abortive transcripts.

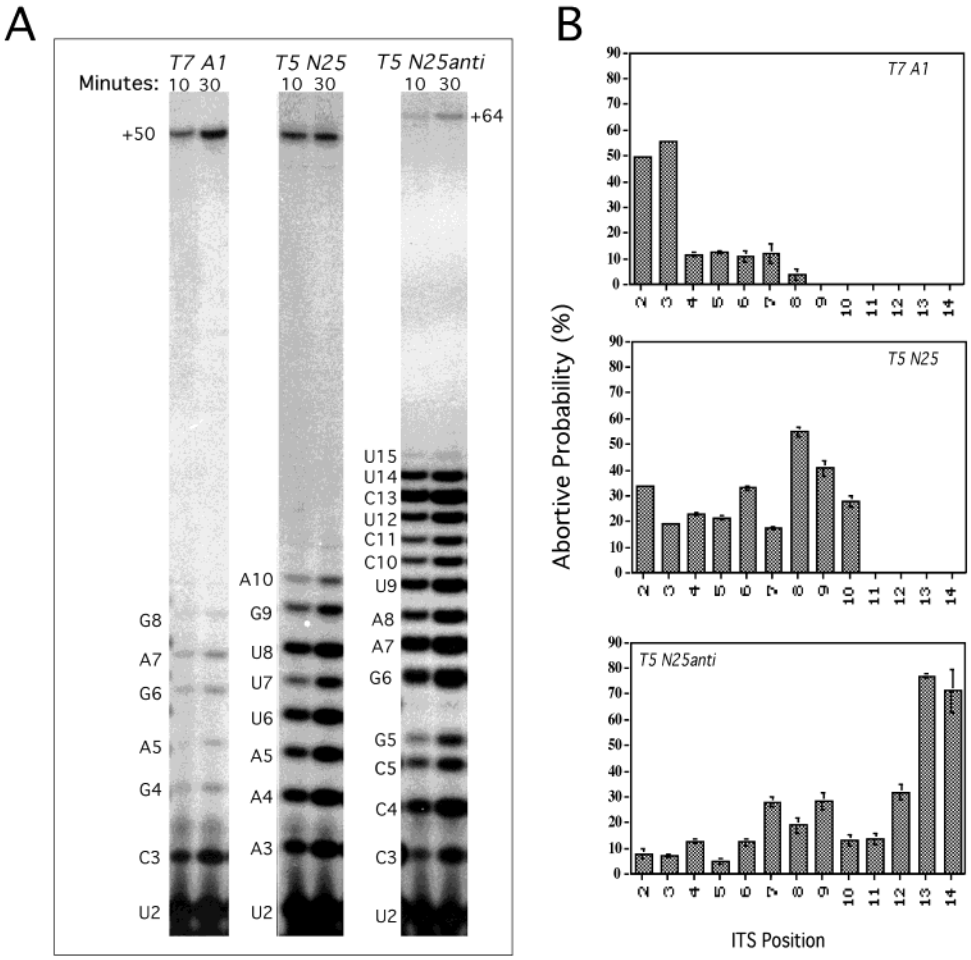
Several criteria convinced us that the short transcripts seen with each of the three promoters are true abortive products. (1) The short transcripts accumulate continuously during transcription for at least 30 min, suggesting that they are not paused products. Furthermore, when transcription reactions are chased at various times with a high concentration (i.e., 1 mM) of unlabeled NTP, no discernible reduction of any short transcripts was observed (data not shown). (2) All of the short RNAs were deemed to have been released from RNA polymerase complexes by their failure to be retained on nitrocellulose filters (*46*) or to elute in the excluded volume when chromatographed on molecular sieve columns (ref *9* and data not shown). (3) Finally, nearest-neighbor analysis using different [α-^{32}P]NMP labels (see Materials and Methods) was employed to probe the sequence of each abortive transcript, ensuring that primer shifting did not take place. According to all of these criteria, the short transcripts seen for the three promoters are true abortive transcripts.

The nearest-neighbor analysis, however, did reveal an unexpected complexity among the short transcripts derived from T5 $N25_{antiDSR}$. While the transcribed sequence of this promoter starts with AUCCGG, the transcript that migrated just above the 4 nt product (pppApUpCpC, abbreviated C4) was found to contain three C residues, and has the sequence pppApUpCpCpC (abbreviated C5). This discrepancy was confirmed by repeat sequencing of the DNA template, and by quantitative nearest-neighbor analysis of the C5 abortive transcript which shows transfer of α-^{32}P from CMP to UMP and CMP in a ratio of 33:67. Experiments in which only ATP, CTP, and UTP were added to the reaction mixture showed that at 1−5 μM CTP, only the C4 transcript was obtained, but above 10 μM CTP, there was extensive formation of the C5 transcript and even traces of C6 and C7 transcripts bearing four and five C residues as well (N. V. Vo and M. J. Chamberlin, unpublished results).

The incorporation of additional C residues into the nascent transcript apparently does not affect the synthesis of abortive transcripts that are template-specified. Thus, migrating slightly above the C5 band is the expected 5 nt abortive product G5 (pppApUpCpCpG). G5 is clearly identified by its distinct mobility and by the characteristic ratios of nearest-neighbor transfer from [α-^{32}P]CMP and [α-^{32}P]GMP. Thus, the incorporation of additional C residues appears to be a branch reaction that occurs at the 4−5 junction only on the T5 $N25_{antiDSR}$ template and yields a dead-end product that is unable to support further elongation on the template sequence.

Quantitative Parameters of Abortive Initiation. To develop quantitative description of abortive versus productive transcription from each promoter, we measure (or derive) the following parameters: the abortive rate, the productive rate, the abortive:productive ratio (APR), the abortive probability, and the maximum size of the abortive transcript (MSAT). The abortive and productive rates, expressed as femtomoles of transcripts per femtomole of RNA polymerase per hour, are derived from time course experiments performed under steady-state conditions where the rate of synthesis remains linear for 1 h. The APR is simply the abortive rate over the productive rate; this parameter gives a quick assessment of the relative extent of abortive versus productive transcription. The abortive pattern refers to the overall distribution of abortive products; each transcript is given a parameter called the abortive probability calculated from its molar yield. From the gel profile, the MSAT is obtained by visual inspection; however, this size can range over a few nucleotides depending on the specific activity of the label that is used. The time course determination for obtaining these parameters is only valid when steady-state transcription is performed at an NTP concentration $\geq$100 μM (*72*). A typical set of data and the ensuing analyses are shown in panels A−C of Figure 3.

Abortive probability is calculated for each initial template position from the molar yield of each abortive transcript further corrected by the fraction of RNA polymerase that reaches a given template position (*27*). Abortive probability is indicative of the instability of an ITC and its tendency to release its RNA; the higher the abortive probability, the more unstable the ITC.



C

PROMOTER	RATE OF RNA SYNTHESIS			APR	MSAT
	TOTAL	PRODUCTIVE	ABORTIVE		
T5N25	888 ± 103	28 ± 5	860 ± 99	31 ± 5	10
T5N25anti	1049 ± 130	6 ± 2	1043 ± 129	174 ± 27	15
T7A1	243 ± 17	31 ± 6	212 ± 14	7 ± 2	8

FIGURE 3: Quantitative parameters of RNA chain initiation and promoter escape for the three promoters. (A) A time course investigation was performed to obtain abortive and productive rates. Here, only two time points for each template are shown. (B) Profile of abortive probabilities for the three promoters. Abortive probability was calculated as described previously (*27*). (C) Tabulation of the various quantitative parameters. Productive and abortive rates are in units of (femto)moles of RNA per (femto)mole of RNA polymerase per hour.

Figure 3B plots the abortive probability associated with the early template positions in the three promoters, and several features are worth noting. For one, there is no obvious sequence dependence of abortive probability; that is, the 3′-terminal nucleotide of a highly abortive product can be any of the four normal bases. This is unlike the case of T7 RNA polymerase that undergoes preferential abortive release after incorporating a U residue (*10*). In fact, the position with the highest abortive probability among the three promoters is a C residue at position 13 of $N25_{antiDSR}$. We note that, of the three promoters examined, $N25_{antiDSR}$ contains the most GC-rich ITS, yet it aborts to the highest degree and produces the longest abortive ladder. Thus, there is no correlation of the abortive probabilities with the ITS sequence composition and the presumed strength of the RNA−DNA heteroduplex formed during initial transcription (*47*, *48*). Furthermore, there appears to be no positional dependence of abortive probability. The three promoters we studied all begin with pppApU; however, the abortive probability of the 2 nt RNA varies widely: 50% for T7 A1, 33% for T5 N25, and 8% for T5 $N25_{antiDSR}$. Finally, the highest abortive probabilities at different promoters occur at distinct positions: at positions 2 and 3 for T7 A1, at positions 8 and 9 for T5 N25, and at positions 13 and 14 for T5 $N25_{antiDSR}$.

Figure 3C summarizes the various quantitative parameters measured for the three promoters. From these results, we conclude that T7 A1, while the least active promoter of the three in terms of the total number of initiations per polymerase per hour, is nevertheless the most facile at promoter escape, showing an APR of only 7. As expected, T5 N25 and T5 $N25_{antiDSR}$ initiated at approximately the same efficiency (*16*), but in $N25_{antiDSR}$, sequence alteration in the initial transcribed region created a hybrid promoter that is now compromised at promoter escape. T5 $N25_{antiDSR}$ con-

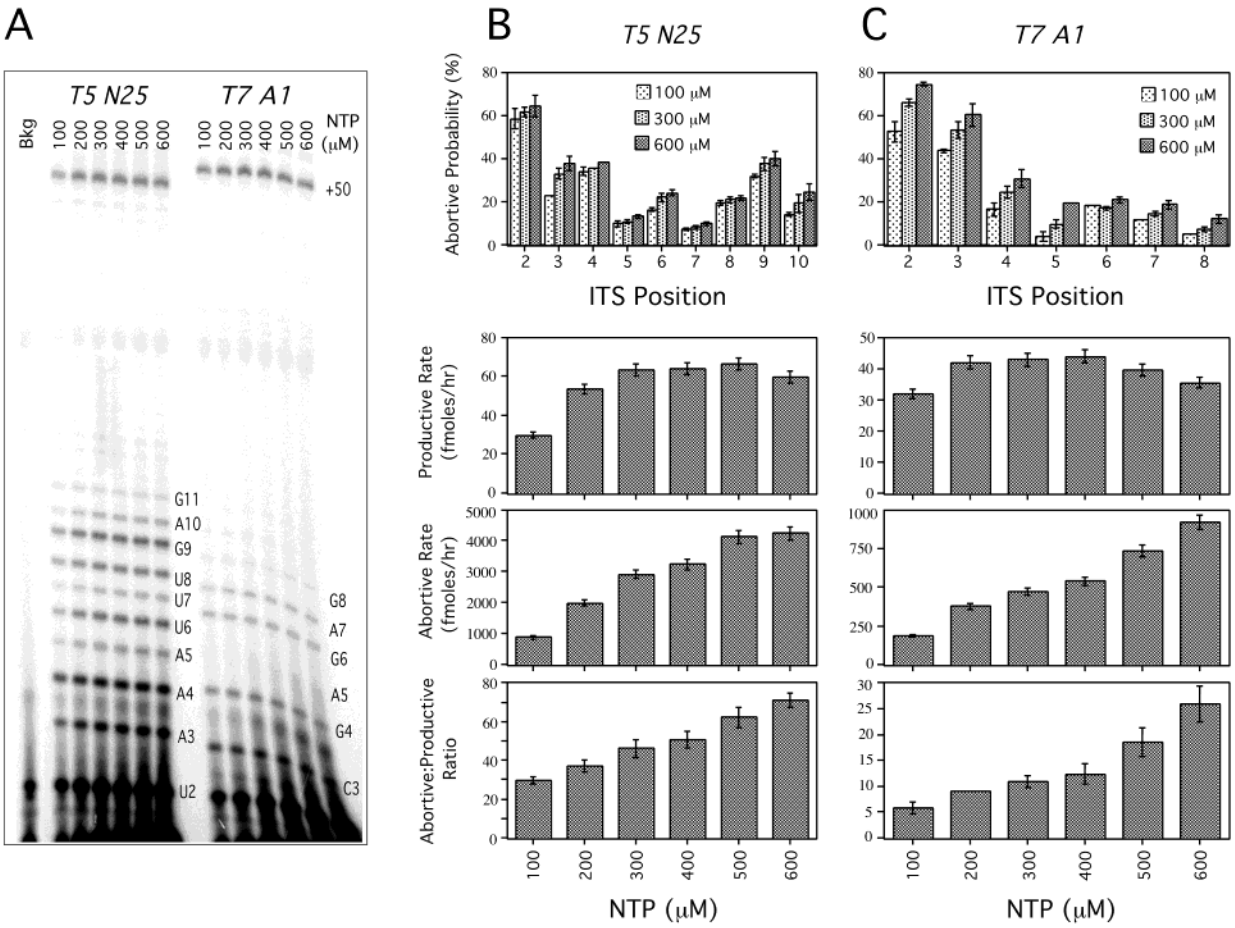


FIGURE 4: Titration by varying all four NTPs. For each set of titrations, the concentration of all four NTPs in each reaction was varied together from 100 to 600 μM. The specific activity of [γ-^{32}P]ATP was kept constant for all NTP concentrations, at ~10 cpm/fmol. (A) Gel profiles. (B) Graphs of quantitative parameters for T5 N25. (C) Graphs of quantitative parameters for T7 A1.

stitutes the first report of a promoter that is rendered weak by changes in the ITS and raises the caveat that, in studying promoter activity where sequence swapping involves both the upstream and downstream promoter regions, it is inadequate to make conclusions only about the upstream region.

Effect of High NTP Concentrations on Abortive Initiation and Promoter Escape. We tested the effect of high NTP concentrations, from 100 to 600 μM of all four nucleoside triphosphates, on abortive initiation and promoter escape from T5 N25 and T7 A1 promoters. The results are shown in panels A−C of Figure 4. Qualitatively, both promoters show little change in the pattern of abortive and runoff products over this range of NTP concentrations (Figure 4A). Quantitatively, there is a positive dependence of both productive and abortive rates with increasing NTP concentrations (Figure 4B,C, middle panels). However, the dependence of the productive rate on NTP concentration is limited; as a result, T5 N25 and T7 A1 both reach a productive rate plateau at 300 and 200 μM NTP, respectively. At higher NTP concentrations, there appears to be a small degree of substrate inhibition of productive RNA synthesis. In contrast, the abortive rates increase linearly and a plateau is not reached even at 600 μM. While an increasing NTP concentration stimulates both productive and abortive synthesis, the increase in the abortive rate is disproportionately large, as indicated by the high APR, which itself exhibits a linear dependence on NTP concentration (Figure 4B,C, bottom panels).

Interestingly, an NTP concentration above 100 μM had little effect on the abortive probabilities overall (Figure 4B,C, top panels). Except for position 3 of T5 N25 and positions 3 and 5 of T7 A1 where >10% change was observed, all other variations are within the standard deviation of the measurement. For the three positions that are mentioned, the direction of the change was toward higher abortive probability with a higher NTP concentration; this was unexpected on the basis of the kinetic competition model of abortive RNA formation (see below). This result suggests that abortive probability is likely an intrinsic property of the ITC and unrelated to kinetic competition.

Effect of Varying Individual NTP Concentrations on Abortive Pattern. A second prediction of the kinetic competition model for initiation is that addition of large amounts of the next NTP needed in the ITS sequence should decrease the amount of aborted transcript to which that NMP is to be added. In simple terms, it is expected that increasing the concentration of the next NTP will increase the rate of addition, much as raising the concentration of substrate increases the rate of an enzyme reaction in Michaelis−Menten kinetics. For this reason, where there is a reduction in the molar yield of an abortive product in response to an increase in the nucleotide substrate concentration, we shall call this the "K_S" effect, and a "K_S site" is one where the

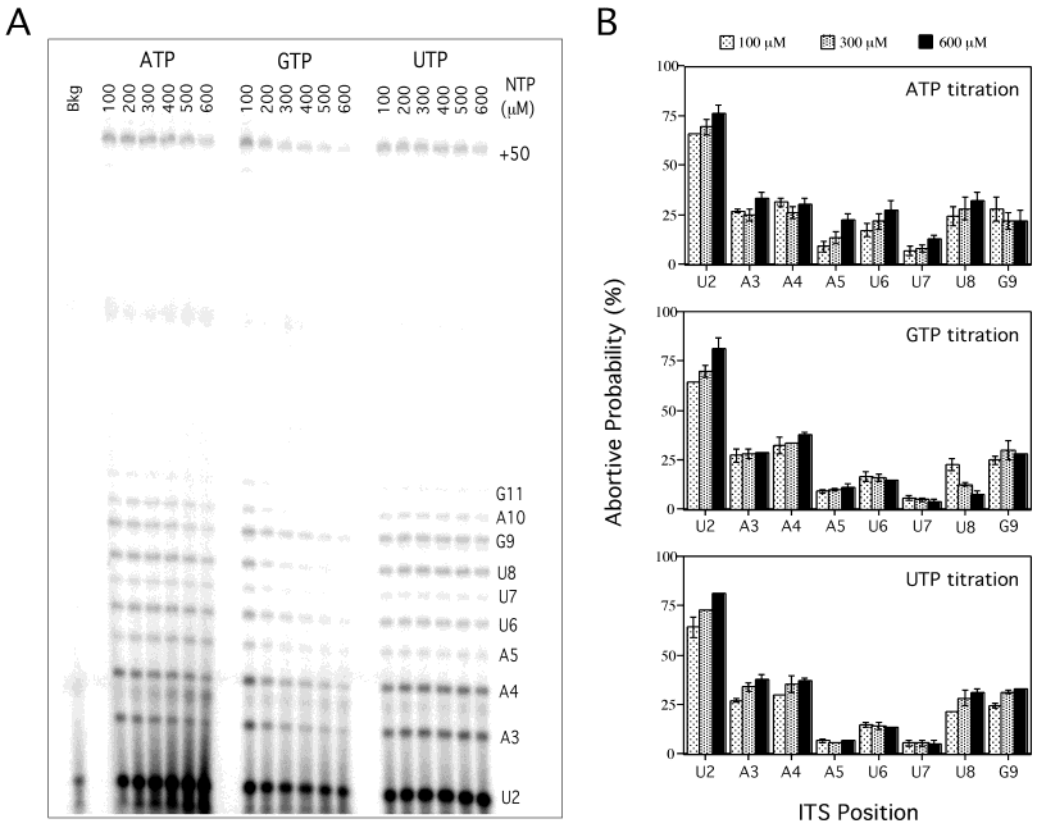


FIGURE 5: Analysis to identify K_S sites. Titration of individual NTPs for the T5 N25 promoter. For each set of titrations, the concentration of one NTP was varied from 100 to 600 μM while the concentration of the other three NTPs was kept at 100 μM. (A) Gel profiles. (B) Abortive probabilities associated with the initial positions: (top panel) ATP titration, (middle panel) GTP titration, and (bottom panel) UTP titration. In each panel, the abortive probabilities for a given position at 100, 300, and 600 μM variable nucleotide are juxtaposed for ease of comparison.

preceding abortive product is reduced with an increasing concentration of the NTP corresponding to the site.

It must be pointed out that, although a plausible analogy, both the initiation and elongation reactions of RNA polymerase on duplex DNA cannot be analyzed by steady-state kinetics models. The apparent values of K_S vary by a factor of up to 1000 for different positions in the template (*49*), thereby rendering the classical application of the ping-pong kinetic treatment invalid (*50*, *51*). However, we will retain the nomenclature for convenience and refer to K_S sites as described above.

To detect the existence of K_S sites within the initial transcribed region, we carried out titration experiments in which one of the NTP substrates was varied from 100 to 600 μM, while the other three were kept at 100 μM. Such a titration revealed no K_S sites within the T7 A1 initial transcribed region. With T5 N25, however, there appears to be a K_S site at G9; the results are shown in panels A and B of Figure 5. Since the initial transcribed sequence of this promoter is AUAAAUUUGA, we performed titration of only ATP, GTP, and UTP. In the GTP titration, the abortive probability of the U8 product decreased from 22 to 8% as the GTP concentration increased from 100 to 600 μM. This K_S effect at G9 was even more pronounced at GTP concentrations below 100 μM (results not shown).

With a K_S site at G9, it is expected that increasing the GTP concentration would lead to an increased amount of full-length product. On the contrary, the yield of full-length RNA decreased with increasing GTP concentrations. We attribute this to substrate level inhibition elicited at high substrate concentrations noted for *E. coli* RNA polymerase under *in vitro* conditions (*52*).

Taken together, T7 A1 and T5 N25 contain very few K_S sites in the initial template region. These results suggest that the rate-limiting step leading to the formation of an abortive product is normally not affected by NTP concentration.

DISCUSSION

Basic Features of the RNA Chain Initiation Reaction by E. coli RNA Polymerase. We have studied the qualitative and quantitative parameters of RNA chain initiation and promoter escape for three bacteriophage promoters. Employing a steady-state reaction in which RNA chains are synthesized and released in many rounds of transcription, this assay allows one to directly determine the molar yields of productive and abortive RNA, and to calculate many of the important parameters for the initiation phase.

In our studies and those of others, certain common elements in the RNA chain initiation reaction that are shared by all of the prokaryotic promoters that have been carefully examined emerge. These include the formation of abortive products that are released by RNA polymerase in the ITC to regenerate the open promoter complex, and the persistence of RNA polymerase at the promoter recognition region until promoter escape is achieved (*7*, *53*, *72*, *73*). Promoter escape represents the culmination of at least several well-orchestrated biochemical events: the relinquishment of promoter DNA contacts by the polymerase (*9*, *12*), the translocation

of RNA polymerase downstream of the promoter recognition region (*9*, *54*), and possibly the release of the σ subunit (*9*, *29*, *55*, *56*; see further discussion below).

Despite these common elements in initiation, there are considerable differences among different promoters during the initiation reaction. Although we have only studied three promoters in detail, the differences among them are so striking that their diversity would seem to serve as a paradigm for the RNA chain initiation reaction. Thus, there are large differences in the rate of promoter escape, in the molar yield of abortive products, and in the abortive pattern and the maximal size of abortive RNA. The differences in these parameters must reflect major biochemical differences in how RNA polymerase moves through the initial transcribed region, and carries out the process of promoter escape. These differences in mechanism are, in turn, due to differences in the DNA sequence of the promoters. Vo (*57*) has undertaken the considerable challenge to dissect the role played by different promoter regions in the various biochemical steps during RNA chain initiation. Their results appear in the subsequent papers in this series (*72*, *73*).

Abortive Initiation Appears To Be an Essential Feature of RNA Chain Initiation. Although we have only studied a limited number of promoters in detail (see also refs *9*, *28*, *72*, and *73*), we hypothesize that all *E. coli* promoters give rise to abortive products during RNA chain initiation in the presence of all four NTP substrates. RNA polymerase II is also prone to aborting nascent RNAs 10 nt or shorter when all four NTP substrates are present; however, the enzyme does not appear to cycle extensively at the promoter (*58*, *59*). These differences aside, we believe that abortive initiation is probably an essential element of the initiation reaction for all promoters with all DNA-dependent RNA polymerases (*15*). In short, we know of no promoters, nor *in vitro* reaction conditions, that permit specific initiation of productive RNA transcripts without the prior synthesis of abortive products. Our hypothesis is supported by the revelation from the crystal structures of *Taq* and *Tth* RNA polymerase holoenzyme; both structures reveal an unexpected disposition of σ region 3.2 in the open complex, implicating a role of this domain of σ^{70} in abortive initiation (*60*, *61*).

What Factors Determine the Molar Yield of Different Abortive Products and the Differences in Abortive Probability? The simple kinetic competition model for RNA chain initiation predicts that the yield of an abortive transcript should decrease as the concentration of the next template-specified nucleotide substrate is increased. Thus, at saturation, it is expected that there would be little or no formation of that particular abortive product. For convenience, we refer to this type of process as revealing a K_S site, for which the concentration of the particular NTP controls formation of the abortive product. We have identified only one K_S site within the initial transcribed region of the T5 N25 promoter. The molar yield of most abortive products is not affected by the levels of the subsequent NTP at concentrations above 100 μM. This suggests that the rate-limiting step in abortive initiation does not involve binding of NTP at most sites.

The dilemma of how the pattern of abortive products is generated is made more puzzling by the dramatic differences we observe among the three promoters we have studied. The T7 A1 promoter is clearly most unstable at the 2 and 3 positions, with associated abortive probabilities of more than 50%. Once beyond positions 2 and 3, the complexes are sufficiently stable to lead to promoter escape, yielding low levels of the longer abortive RNA and a high level of productive RNA synthesis. For T5 N25, the major block exists at position 8. Finally, for the T5 $N25_{antiDSR}$ promoter, the major block lies at positions 13 and 14, with abortive probabilities of 75 and 70%, respectively. Since T5 N25 and T5 $N25_{antiDSR}$ promoters have identical promoter recognition sequences, and share transcribed sequences to position 2, the ITS from position 3 to 20 is clearly responsible for this change. It will be interesting, therefore, to make alterations in these two initial sequence regions to study their effects.

Maximal Size of Abortive Transcripts and the Position of σ Release. We have summarized the evidence that the 14 and 15 nt products formed at the T5 $N25_{antiDSR}$ are true abortive transcripts, and are released from the ITC. However, what is the evidence to implicate the correspondence between the MSAT and the position of σ release? In a previous study, it was shown that the *tac* promoter yields an abortive ladder of 8−9 nt; on that promoter, σ was released after the incorporation of either the ninth or tenth nucleotide (*9*). Similarly, the T7 A1 promoter yields an abortive ladder of 7−8 nt, and the σ factor was released during the addition of the eighth or ninth nucleotide (*29*). For T5 $N25_{antiDSR}$, the sequence composition of the ITS precludes a direct determination of the position of σ release. By inference from the earlier studies, however, σ release must not occur until after the formation of the longer abortive transcripts. In the case of T5 $N25_{antiDSR}$, the MSAT is 14−15 nt and the promoter escape transition likely occurs after that. That the promoter escape transition might take place such a long distance from the start site is no longer an unusual claim, since it has been shown on the λ $P_{R'}$ promoter that σ was retained even at position 16 (*30*). Indeed, the analysis of an increasing number of promoters indicates that σ release, if equated to take place right after the longest abortive transcript, must occur at different positions on different promoters (*57*; L. Hsu, unpublished results). Therefore, there does not appear to be a singular transition point for all promoters on any RNA polymerase, although such has been suggested from recent structural (*62*) and biochemical studies (*63*).

The point of σ release aside, the issue of whether σ is released at all during transcription or only partially released depending on the physiological state of the cell has become a controversy (*64*, *65*). The status of σ release at the promoter escape transition awaits further clarification (*13*, *14*).

Unusual Features of T5 $N25_{antiDSR}$. Our results show that T5 $N25_{antiDSR}$ is an enigmatic promoter. It is derived from T5 N25 by sequence substitution, A for C and T for G and vice versa, over an 18 bp stretch, not in the promoter recognition region, but in the ITS region from position 3 to 20. Kammerer et al. (*16*) showed that, compared to T5 N25, T5 $N25_{antiDSR}$ is 10-fold less active *in vitro* and *in vivo*. Both promoters give rise to the same high association constant with RNA polymerase, so the polymerase binding and activation aspect of promoter function is not compromised. Kammerer et al. concluded that T5 $N25_{antiDSR}$ is rate-limited at a step after the open complex formation, namely, the promoter escape step. The analysis presented here provides the biochemical evidence to support their conclusion; T5 $N25_{antiDSR}$ shows a 99.4% abortive yield versus a 0.6%

productive yield, giving rise to only one productive synthesis every 165 initiation events. The high abortive yield takes the unusual form of a long abortive ladder of 14−15 nt. Only one other Eσ^{70} promoter, i.e., λ $P_{R'}$, produces an equally long abortive ladder (*66*). The long abortive ladder of $P_{R'}$ requires σ rebinding to bring about the pausing of RNA polymerase at position 16 or 17; this pause is in turn necessary to achieve Q-mediated antitermination (*30*, *67*).

T5 $N25_{antiDSR}$ also undergoes an unusual reaction in that it adds an extra one to three C residues beginning at position 5 in the nascent RNA, leading to transcripts that are not exactly complementary to the template sequence. None of these abortive transcripts appears to re-enter the template-directed transcription reaction. This is an unprecedented reaction with DNA-dependent RNA polymerases. It does not resemble primer shifting, which requires a template region upstream of position 1 that is complementary to the initiating primer (*43*, *45*). It also differs from the "slipping" reaction demonstrated by others in two respects: (1) it does not form a long sequence of C residues (*21*), and (2) it slips over an internal CC dinucleotide repeat rather than the three identical template residues normally required (*68*).

Of the three promoters analyzed here, only T5 $N25_{antiDSR}$ was found in an earlier study to be significantly stimulated in promoter escape by the *E. coli* transcript cleavage factor GreA and/or GreB (*17*). At the time, we wondered about the basis of template specificity of GreA and/or GreB action. Given the quantitative parameters we now have, it is clear that GreA and GreB acted on T5 $N25_{antiDSR}$ to overcome the rate limitation at promoter escape. The GreA or GreB stimulation involves transcript cleavage of RNAs 7 nt or longer (*17*), suggesting that the rate limitation of promoter escape at T5 $N25_{antiDSR}$ might be intimately tied to transcriptional arrest involving backtracking of the polymerase active center (*69*, *70*). ITCs with RNA shorter than 7 nt are not susceptible to GreA- or GreB-stimulated cleavage, most likely because the RNA, especially after its 3′-OH end has been displaced, is too short to form a stabilized substrate for transcript cleavage; the displaced RNA is eventually released. By this account, transcriptional arrest during the initiation stage may be what leads to abortive initiation (*13*).

REFERENCES

1. McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171.
2. Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S. (1993) *Annu. Rev. Biochem. 62*, 749.
3. Zawel, L., and Reinberg, D. (1995) *Annu. Rev. Biochem. 64*, 533.
4. Record, M. T., Jr., Reznikoff, W. S., Craig, M. L., McQuade, K. L., and Schlax, P. J. (1996) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., and Umbarger, H. E., Eds.) 2nd ed., pp 792−821, ASM Press, Washington, DC.
5. deHaseth, P. L., Zupanic, M. L., and Record, M. T., Jr. (1998) *J. Bacteriol. 180*, 3019.
6. Cech, C. L., and McClure, W. R. (1980) *Biochemistry 19*, 2440.
7. Carpousis, A. J., and Gralla, J. D. (1980) *Biochemistry 19*, 3245.
8. Munson, L. M., and Reznikoff, W. S. (1981) *Biochemistry 20*, 2081.
9. Krummel, B., and Chamberlin, M. J. (1989) *Biochemistry 28*, 7829.
10. Martin, C. T., Muller, D. K., and Coleman, J. E. (1988) *Biochemistry 27*, 3966.
11. Whipple, F. W., and Sonenshein, A. L. (1992) *J. Mol. Biol. 223*, 399.
12. Carpousis, A. J., and Gralla, J. D. (1985) *J. Mol. Biol. 183*, 165.
13. Hsu, L. M. (2002) *Biochim. Biophys. Acta 1577*, 191.
14. Hsu, L. M. (2002) *Nat. Struct. Biol. 9*, 502.
15. Chamberlin, M. J., and Hsu, L. M. (1996) in *Regulation of Gene Expression in Escherichia coli* (Lin, E. C. C., and Lynch, A. S., Eds.) pp 7−25, Landes Press, Austin, TX.
16. Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986) *EMBO J. 5*, 2995.
17. Hsu, L. M., Vo, N. V., and Chamberlin, M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 11588.
18. Menendez, M., Kolb, A., and Buc, H. (1987) *EMBO J. 6*, 4227.
19. Eichenberger, P., Dethiollaz, S., Buc, H., and Geiselmann, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 9022.
20. Liu, C., and Turnbough, C. L., Jr. (1989) *J. Bacteriol. 171*, 3337.
21. Jin, D. J., and Turnbough, C. L., Jr. (1994) *J. Mol. Biol. 236*, 72.
22. Rasmussen, E. B., and Lis, J. T. (1995) *J. Mol. Biol. 252*, 522.
23. Krumm, A., Hickey, L. B., and Groudine, M. (1995) *Genes Dev. 9*, 559.
24. Dvir, A., Conaway, R. C., and Conaway, J. W. (1996) *J. Biol. Chem. 271*, 23352.
25. Pal, M., McKean, D., and Luse, D. S. (2001) *Mol. Cell. Biol. 21*, 5815.
26. Dvir, A. (2002) *Biochim. Biophys. Acta 1577*, 208.
27. Hsu, L. M. (1996) *Methods Enzymol. 273*, 59.
28. Levin, J. R., Krummel, B., and Chamberlin, M. J. (1987) *J. Mol. Biol. 196*, 85.
29. Metzger, W., Schickor, P., Meier, T., Werel, W., and Heumann, H. (1993) *J. Mol. Biol. 232*, 789.
30. Ring, B. Z., Yarnell, W. S., and Roberts, J. W. (1996) *Cell 86*, 485.
31. Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986) *EMBO J. 5*, 2987.
32. Brunner, M., and Bujard, H. (1987) *EMBO J. 6*, 3139.
33. Knaus, R., and Bujard, H. (1991) *Nucleic Acids Mol. Biol. 4*, 100.
34. Studier, F. W., Rosenberg, A. H., Simm, M. N., and Dunn, J. J. (1979) *J. Mol. Biol. 135*, 917.
35. Burgess, R. R., and Jendrisak, J. J. (1975) *Biochemistry 14*, 4634.
36. Gonzalez, N., Wiggs, J., and Chamberlin, M. J. (1977) *Arch. Biochem. Biophys. 182*, 404.
37. Chamberlin, M., Nierman, W., Wiggs, J., and Neff, N. (1979) *J. Biol. Chem. 254*, 10061.
38. Arndt, K. M., and Chamberlin, M. J. (1987) *J. Mol. Biol. 202*, 271.
39. Josse, J., Kaiser, A. D., and Kornberg, A. (1961) *J. Biol. Chem. 236*, 864.
40. Chamberlin, M. J., Kingston, R., Gilman, M., Wiggs, J., and DeVera, A. (1983) *Methods Enzymol. 101*, 540.
41. Leirmo, S., Harrison, C., Cayley, D. S., Burgess, R. R., and Record, M. T., Jr. (1987) *Biochemistry 26*, 2095.
42. Gralla, J. D., Carpousis, A. J., and Stefano, J. E. (1980) *Biochemistry 19*, 5864.
43. Borukhov, S., Sagitov, V., Josaitis, C. A., Gourse, R. L., and Goldfarb, A. (1993) *J. Biol. Chem. 268*, 23477.
44. Feng, G., Lee, D. N., Wang, D., Chan, C. L., and Landick, R. (1994) *J. Biol. Chem. 269*, 22282.
45. Severinov, K., and Goldfarb, A. (1994) *J. Biol. Chem. 269*, 31701.
46. Altmann, C. R., Solow-Cordero, D. E., and Chamberlin, M. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3784.
47. Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997) *Cell 89*, 33.
48. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001) *Science 292*, 1876.
49. Levin, J. R., and Chamberlin, M. J. (1987) *J. Mol. Biol. 196*, 61.
50. Rhodes, G., and Chamberlin, M. J. (1974) *J. Biol. Chem. 249*, 6675.
51. Kingston, R., and Chamberlin, M. J. (1981) *J. Biol. Chem. 256*, 2787.
52. Chamberlin, M., and Berg, P. (1964) *J. Mol. Biol. 8*, 708.
53. Kubori, T., and Shimamoto, N. (1996) *J. Mol. Biol. 256*, 449.
54. Zaychikov, E., Denissova, L., and Heumann, H. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 1739.
55. Hansen, U. M., and McClure, W. R. (1980) *J. Biol. Chem. 255*, 9564.
56. Straney, D. C., and Crothers, D. M. (1987) *J. Mol. Biol. 193*, 267.
57. Vo, N. V. (1998) Ph.D. Thesis, University of California, Berkeley, CA.
58. Jacob, G. A., Kitzmiller, J. A., and Luse, D. S. (1994) *J. Biol. Chem. 269*, 3655.
59. Pal, M., and Luse, D. S. (2002) *Mol. Cell. Biol. 22*, 30.
60. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) *Nature 417*, 712.

61. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) *Science 296*, 1280.
62. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000) *Science 289*, 619.
63. Kugel, J. F., and Goodrich, J. A. (2002) *Mol. Cell. Biol. 22*, 762.
64. Bar-Nahum, G., and Nudler, E. (2001) *Cell 106*, 443.
65. Mukhophadyay, J., Kapanidis, A. N., Meckler, V., Kortkhonjia, E., Ebright, Y. W., and Ebright, R. H. (2001) *Cell 106*, 453.
66. Marr, M. T., and Roberts, J. W. (2000) *Mol. Cell 6*, 1275.
67. Marr, M. T., Datwyler, S. A., Meares, C. F., and Roberts, J. W. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 8972.
68. Jacques, J.-P., and Kolakofsky, D. (1991) *Genes Dev. 5*, 707.
69. Arndt, K., and Chamberlin, M. J. (1990) *J. Mol. Biol. 213*, 79.
70. Rudd, M. D., Izban, M. G., and Luse, D. S. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8057.
71. Sheen, J.-Y., and Seed, B. (1988) *BioTechniques 6*, 942.
72. Vo, N. V., Hsu, L. M., Kane, C. M., and Chamberlin, M. J. (2003) *Biochemistry 42*, 3787−3797.
73. Vo, N. V., Hsu, L. M., Kane, C. M., and Chamberlin, M. J. (2003) *Biochemistry 42*, 3798−3811.

BI026954E