# *In Vitro* Studies of Transcript Initiation by *Escherichia coli* RNA Polymerase. 3. Influences of Individual DNA Elements within the Promoter Recognition Region on Abortive Initiation and Promoter Escape†

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

*Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3202, and Program in Biochemistry, Mount Holyoke College, South Hadley, Massachusetts 01075-6456*

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

ABSTRACT: Abortive initiation and promoter escape are two principal biochemical reactions occurring in the latter stage of transcript initiation. We have analyzed the influences of individual DNA elements within the promoter recognition region (PRR) on these reactions by measuring the quantitative initiation parameters that describe abortive initiation and promoter escape; these parameters are the abortive rate, the productive rate, the abortive:productive ratio, the abortive probability, and the maximum size of abortive transcripts. Changes in the individual DNA elements within the PRR can have a substantial effect on each of these parameters. The discriminator region and the −10 element primarily influence the abortive probability at positions 2−5 and 6−10, respectively, while the −10 and −35 conserved hexamers and the spacer region affect the abortive probability at positions 11−15. Surprisingly, transcription of a consensus promoter invariably gives a higher abortive yield, a higher abortive probability, a longer abortive ladder, and a lower productive rate than promoter variants carrying even a single deviation in the consensus hexamers. These results suggest that strong RNA polymerase−PRR interactions stall the polymerase at the promoter, thereby reducing the rate of promoter escape and consequently enhancing the extent of abortive initiation.

Transcription, like other macromolecular synthetic processes, can be delineated into three phases: initiation, elongation, and termination. Initiation begins with the binding of the RNA polymerase holoenzyme to the promoter and ends with the conversion of the enzyme−DNA complex into an elongation complex. Between the beginning and end, the polymerase−promoter complex undergoes a multistep series of conformational and reactivity changes that can be further distinguished as two stages, (1) promoter binding and activation and (2) RNA chain initiation and promoter escape, based on the availability of NTP substrates. The first stage deals with the binding of RNA polymerase to the promoter to form a catalytically active complex, and the second stage involves the commencement of RNA synthesis and the consequent escape of RNA polymerase from the promoter region. It is the balance of binding and activation, and initiation and escape, that defines the strength of a promoter in initiating productive transcription (*1*, *2*).

We are interested in understanding the reactivity of the polymerase−promoter complex during the stage of RNA chain initiation and promoter escape when NTP substrates are present. With formation of the first phosphodiester bond, RNA polymerase is poised to move downstream. However, on most promoters, promoter escape does not occur until the synthesis of RNA reaches ∼10 nucleotides (nt)[1] and beyond. In the meantime, the enzyme produces an abundant level of abortive transcripts ranging from 2 to 15 nucleotides due to repetitive cycling of the enzyme back to the open complex conformation after releasing the short RNA (*3*−*6*). Abortive initiation ceases when RNA polymerase succeeds in translocating away from the promoter. This was thought to occur concurrently with the release of the $\sigma$ factor and the conversion of the core enzyme into a stable elongation complex in which the RNA transcript is tightly bound (*5*, *7*−*9*), although recent evidence indicates that the $\sigma$ factor may not be automatically released (*10*, *11*). Likely, at this transition, the binding of $\sigma$ factor to promoter DNA elements is broken to yield the reduced DNase I footprint of elongating complexes (*5*, *7*−*9*), but $\sigma$ binding to the core enzyme only weakened such that subsequent physical manipulation (e.g., by gel chromatography or washing and dilution, etc.) facilitates its release (*12*, *13*). Promoter escape, therefore, represents the transition from transcript initiation to elongation, marked by the breaking of $\sigma$−promoter contacts. The ease with which this transition occurs can be

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

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

‡ University of California.

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

|| Mount Holyoke College.

[1] Abbreviations: PRR, promoter recognition region; ITS or ITR, initial transcribed sequence region; bp, base pair(s); DIS, discriminator sequence; UPS or UAS, upstream (activating) sequence; APR, abortive:productive ratio; MSAT, maximum size of abortive transcript; nt, nucleotide(s).

10.1021/bi026962v CCC: $25.00 
Published on Web 03/14/2003

assessed by the relative extent of abortive versus productive synthesis from a given promoter.

Abortive initiation and promoter escape are universally observed in all RNA polymerases studied *in vitro*, but the molecular mechanism accounting for these reactivities is not clearly known. Because of their complexities, a set of quantitative parameters has been put forth to characterize various aspects of these reactions (*14*, *15*, *80*). Using these parameters, one can begin to compare the promoter escape efficiency among promoters and examine the role of various factors—promoter sequence composition, nucleotide substrate concentration, accessory proteins, etc.—on transcript initiation (*5*, *6*, *16*−*19*).

Relative to the transcription start site +1, a typical prokaryotic promoter generally spans nucleotides −60 to 20, in which the sequence from position −60 to −1 is hereby termed the promoter recognition region (PRR) and the sequence from position 1 to 20 the initial transcribed sequence region (ITS or ITR). The PRR is further composed of several distinct DNA elements. Central to PRR is the core region of −35 and −10 conserved hexamers joined by a spacer of the appropriate length that is required for positioning these two elements for direct interaction with the $\sigma$ subunit of the holoenzyme (*20*−*25*). Upstream of the core region, many promoters contain an AT-rich sequence, known as the UP element, which has been shown to interact with the carboxyl terminus of $\alpha$ subunits during promoter binding (*26*, *27*). Downstream of the core region, a 6−8 bp stretch connects the −10 box to the transcribed sequence. This region is involved in the melting isomerization to form the open promoter complex (*28*) and has been dubbed the discriminator (DIS) based on sequence characteristics of promoters subject to stringent control (*29*).

Although promoter sequence serves as the primary determinant of promoter binding, past studies did not address how the individual PRR and ITS elements might affect abortive initiation and promoter escape. In this paper, we systematically investigated the effect of the individual PRR elements on transcript initiation. The results presented here shed important light on understanding the molecular mechanism of abortive initiation and promoter escape.

## MATERIALS AND METHODS

*Enzymes.* *Escherichia coli* DNA-dependent RNA polymerase was purified from *E. coli* DG156 according to Gonzalez et al. (*30*); preparations used in this study contained 18−65% active molecules (*31*). The $His_6$-tagged RNA polymerase holoenzyme was purified from *E. coli* strain RL712 (R. Landick) by the method of Uptain (*32*); preparations used in this study contained 40−75% active molecules (*31*).

*Bacterial Strains and Plasmids*. The pBluSK-PrmT20/$T_1T_2$ plasmid (Figure 1) was constructed by excising a 500 bp *Hin*dIII DNA fragment containing the *rrnB* $T_1T_2$ tandem terminators from pKK5-1 (*33*) and cloning it into the *Hin*dIII site of pBluescriptSK (Stratagene) to create first the pBluSK-$T_1T_2$ plasmid. Next, a 116 bp DNA fragment containing the PrmT20 promoter was cloned into the *Sac*I and *Eco*RV restriction sites of pBluSK-$T_1T_2$ to make pBluSK-PrmT20/$T_1T_2$. This 116 bp insert, designed to contain convenient restriction sites between promoter elements, was generated by extension of two long DNA primers overlapped by 15 bp.

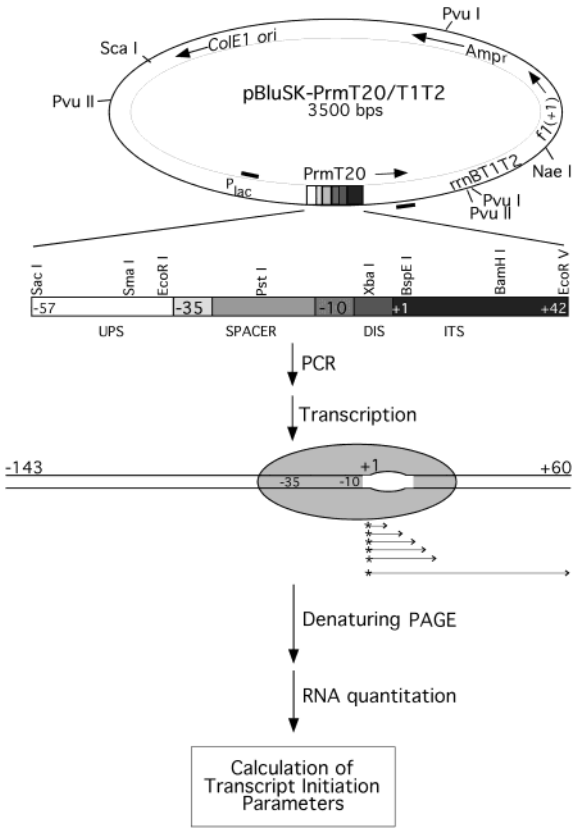


FIGURE 1: Experimental protocol for investigating the abortive initiation and promoter escape pattern of various PRR mutants. At the top, the parent plasmid pBluSK-PrmT20/$T_1T_2$ was constructed as described in Materials and Methods and used to generate the promoter variants by cassette mutagenesis using conveniently placed restriction sites. In the middle part, DNA fragments (nucleotides −143 to 60) containing a single promoter were generated by PCR and subjected to steady state transcription analysis. [$\gamma$-$^{32}$P]ATP was used to end-label all RNA transcripts (indicated with asterisks) from 0 to 60 min at 37 °C. At the bottom, abortive and productive products were resolved on a single 23% (19:1) polyacrylamide−7 M urea gel. Each RNA product, from dinucleotide tetraphosphate to the full-length RNA, was quantified for calculation of the quantitative initiation parameters.

Cassette mutagenesis was performed on the pBluSK-PrmT20/$T_1T_2$ plasmid to make changes in individual DNA elements in the promoter region. First, two complementary DNA primers containing the desired sequence modifications were annealed in 10 mM Tris-HCl (pH 8) and 1 mM $Na_2$-EDTA by heating at 65 °C for 5 min, followed by slow cooling to 25 °C to form a short double-stranded DNA fragment with the appropriate restriction enzyme sites. The fragment was then cloned into pBluSK-PrmT20/$T_1T_2$ plasmid at the complementary blunt and/or cohesive sites. *Sac*I and *Eco*RI sites were used to clone in the various upstream sequence elements; *Eco*RI and *Pst*I, the −35 element; *Eco*RI and *Xba*I, the spacer region; *Pst*I to *Xba*I, the −10 element; *Xba*I and *Bsp*EI, the discriminator region; and *Xba*I or *Bsp*EI and *Bam*HI, the ITS. Bacterial strain XL1 Blue (*34*) was the host cell for transformation. All sequence alterations in the promoter were verified by DNA sequencing (*35*).

*Promoter Templates.* Single-promoter DNA fragments were amplified by polymerase chain reaction (PCR) (*36*) from plasmid DNA templates. Promoter fragments generally

span nucleotides −150 to 60, giving rise to short productive runoff transcripts that can be resolved on the same gel with abortive transcripts. DNA products from PCR were purified by one round each of phenol/chloroform extraction, ammonium acetate (2.5 M)/ethanol precipitation, and sodium acetate (0.3 M)/ethanol precipitation. This procedure removed >99.9% of the free dNTP and >80% of the residual primers (*15*). The concentration of the PCR promoter was determined spectrophotometrically.

*Steady State Transcription Assay.* Single-promoter fragment templates were transcribed under the extensive synthesis condition (*37*, *38*). All reactions were tested to ensure the enzyme recycled quantitatively (data not shown). Transcription reaction mixtures (20−50 $\mu$L) containing 30 nM promoter template, 150 mM KCl, all four NTPs (100 $\mu$M each), and [$\gamma$-$^{32}$P]ATP (~5 cpm/fmol) in buffer [40 mM Tris-HCl (pH 8), 10 mM $MgCl_2$, 10 mM $\beta$-mercaptoethanol, and 10 $\mu$g/mL acetylated BSA] were prepared on ice. *E. coli* RNA polymerase was then added to a final concentration of 30 nM, and the reaction mixtures were incubated at 37 °C. At desired points (0−60 min), samples of 5−10 $\mu$L were removed and mixed with an equal volume of the formamide loading buffer (FLB) {80% (v/v) freshly deionized formamide, 1× TBE [89 mM Trisma base, 89 mM boric acid, and 2.5 mM $Na_2EDTA$ (pH 8.3)], 10 mM $Na_2EDTA$, and 0.025% (w/v) xylene cyanol (XC)}. RNA products were fractionated by gel electrophoresis. To determine the quantitative initiation parameters, steady state transcription was performed at least three times independently for each promoter variant to obtain the average value with one standard deviation as reported in the tables.

*Gel Electrophoresis.* RNA products from all transcription reactions were resolved on 23% (38:2) polyacrylamide−7 M urea gels. Gel electrophoresis was performed in a salt gradient buffer (*39*). RNA samples were heated at 100 °C for 3 min before being loaded onto a prerun gel. Electrophoresis was continued at a constant power of 1.5 W/cm until XC has migrated ~17 cm from the wells. The gels were then exposed and scanned in a phosphorimager (Molecular Dynamics Storm) and RNA signals quantified using the ImageQuant software. Conversion of phosphorimager counts (in IQ volume units) to femtomoles of RNA was performed as described (*80*).

*Nearest-Neighbor Analysis*. Nearest-neighbor analysis (*40*) was carried out as described in an accompanying paper (*80*) to verify the identity of short RNA transcripts.

## RESULTS

*Rationale and Experimental Approach.* Previous studies have demonstrated that different promoters give rise to greatly different levels of abortive initiation and promoter escape (*4*, *6*, *17*, *18*, *41*, *80*). We sought to determine the sequence contribution of PRR and/or ITS in determining the extent of abortive initiation and promoter escape at different promoters. In this study, we investigated the role of various PRR elements.

Distinct DNA elements within the PRR have been shown to mediate the binding of RNA polymerase to a promoter sequence and its subsequent activation. These DNA elements include the −10 and −35 consensus hexamers, the spacer region, the upstream (activating) sequence (UPS or UAS), and the downstream discriminator sequence (DIS). We were interested in knowing whether the same DNA elements that participate in promoter recognition, binding, and melting play a functional role in transcript initiation. Our approach involves systematically varying the sequence of individual PRR elements and subjecting the promoter variants to steady state transcription analysis followed by the determination of quantitative initiation parameters (*80*).

*Quantitative Initiation Parameters and Their Interpretation.* Quantitative initiation parameters are derived from measurement of individual RNA products, all of the abortive RNAs from two nucleotides up and the productive runoff RNA, made at a promoter during a steady state transcription assay (*15*). In this study, we calculated a slightly different set of parameters to ensure that they accurately reflect the steady state situation and are not compromised by the formation of unproductive initial transcribing complexes (see the Discussion). The parameters we compared include the abortive rate, the productive rate, the abortive:productive ratio (APR), the abortive probability, and the maximum size of abortive transcripts (MSAT). The abortive rate, expressed as moles of abortive RNA per mole of enzyme per hour, measures the rate of total abortive RNA synthesis. The productive rate, also expressed as moles of productive RNA per mole of enzyme per hour, measures productive RNA synthesis and is indicative of promoter strength. The sum of abortive and productive rates corresponds to the frequency of total initiation from a promoter during a given time period. The APR is the ratio of abortive to productive rate (or yield) and indicates the extent of abortive versus productive initiation occurring at a promoter. The MSAT monitors the longest abortive RNA that accumulates with time and is released from the transcription complex during a steady state assay. It is indicative of the length of the abortive initiation−promoter escape program and has been shown to vary with individual promoters (see the results presented below). The abortive probability, expressed in percentage, represents the likelihood of an initial transcribing complex (ITC) aborting its RNA at a particular template position (*15*). The greater the abortive probability is at a template position, the more prone the ITC (at that position) is to release its RNA and, therefore, the more unstable is that ITC. The abortive probability profile is the display of abortive probability associated with each ITC during abortive initiation. It gives an overview of the process a polymerase molecule must undergo to achieve promoter escape at a given promoter. The abortive probability profile raises the following questions regarding promoter escape from a promoter. How many high-barrier steps (i.e., steps with high abortive probabilities) are involved? Where do these steps occur? When is promoter escape accomplished?

*Natural E. coli Promoters Exhibit Vastly Different Patterns of Abortive Initiation and Promoter Escape.* We started our investigation by studying the effect of different promoter sequences on the quantitative initiation parameters. We chose five natural bacteriophage promoters: $\lambda$ $P_R$, $P_{R'}$, and $P_L$, T5 N25, and T7 A1. Their sequences from nucleotide −60 to 20 are shown in Figure 2A. T5 N25 and T7 A1 templates produced runoff transcripts 50 nt in length, and $P_{R'}$ produced runoff transcripts 60 nt in length. Both the $P_R$ and $P_L$ promoter templates used in this study contain the $\rho$-independent terminator *rrnB*$T_1$ near the 3′-end of the template

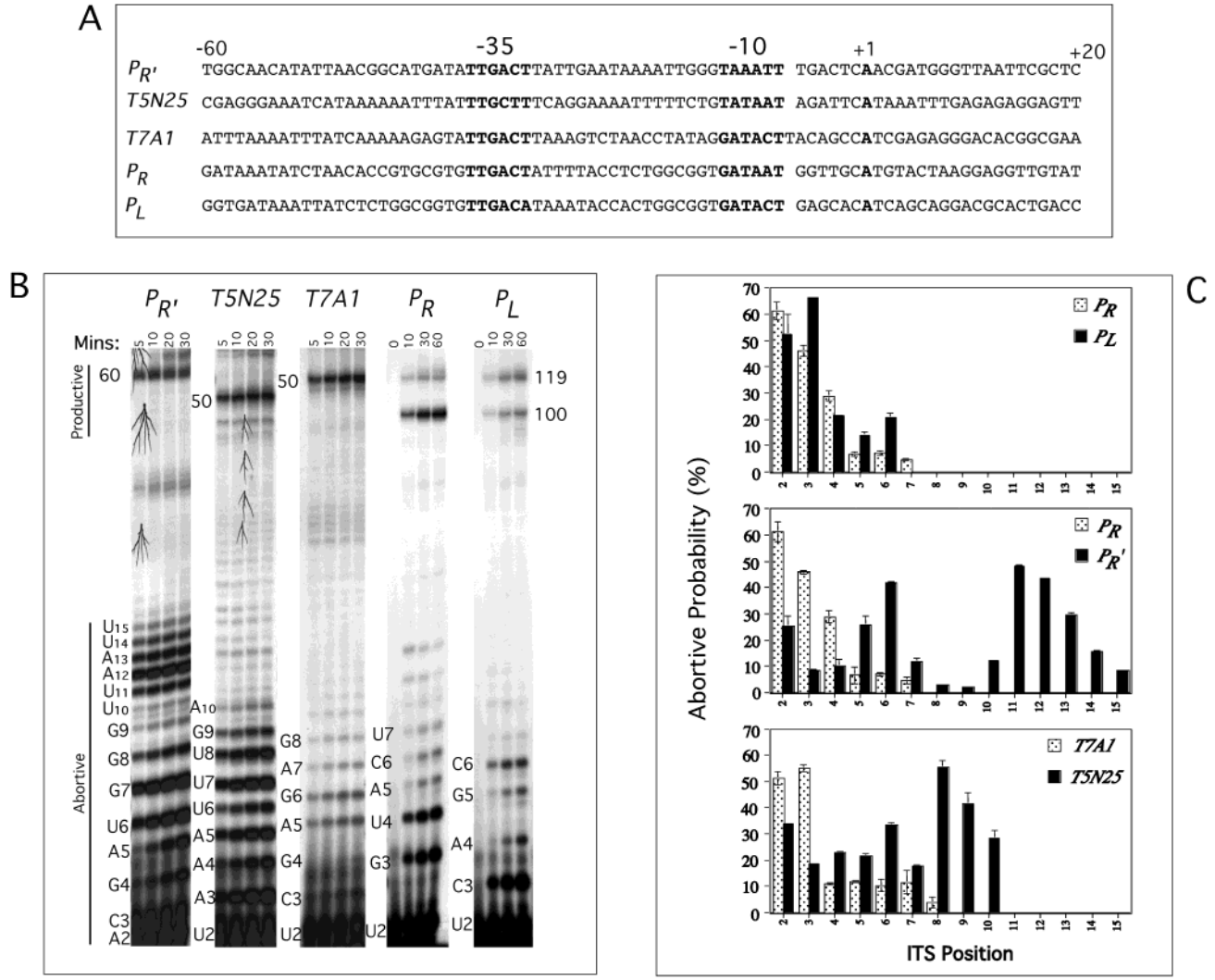


FIGURE 2: Quantitative analysis of transcript initiation from natural bacteriophage promoters. Five promoters were examined: $\lambda$ $P_{R'}$, T5 N25, T7 A1, $\lambda$ $P_R$, and $\lambda$ $P_L$. (A) Their promoter DNA sequences from position −60 to 20. (B) Gel profiles of abortive and productive RNA synthesis as a function of time. Abortive RNAs are represented by a letter and number combination. The number indicates the length of the RNA (in nucleotides), and the letter is the identity of the 3′-nucleotide. Productive transcripts are denoted with a positive number. For $P_R$ and $P_L$, the 100 band is the terminated RNA and the 119 band the runoff product. (C) Pairwise comparison of the abortive probability profiles.

Table 1: Quantitative Initiation Parameters for Several Bacteriophage Promoters[a]

| promoter | rate of RNA synthesis[b] total | productive | abortive | APR[c] | MSAT (nt) |
|---|---|---|---|---|---|
| $P_R'$ | 294 ± 62 | 14 ± 4 | 280 ± 59 | 20 ± 5 | 15 |
| T5 N25 | 888 ± 103 | 28 ± 5 | 860 ± 100 | 31 ± 5 | 10 |
| T7 A1 | 242 ± 20 | 31 ± 6 | 212 ± 14 | 7 ± 1 | 8 |
| $P_R$ | 74 ± 8 | 8 ± 1 | 65 ± 10 | 8 ± 1 | 7 |
| $P_L$ | 56 ± 9 | 4 ± 1 | 52 ± 9 | 13 ± 2 | 6 |

[a] The mean values are derived from at least three independent experiments and shown with one standard deviation. [b] The rate of RNA synthesis is expressed as the femtomoles of RNA molecules produced per enzyme per hour. The total rate is the sum of the abortive and productive rates. [c] APR represents the molar ratio of abortive vs productive transcripts derived from each promoter.

fragment, and therefore generate both a terminated transcript of 100 nt and a runoff transcript of 119 nt (Figure 2B).

Transcription of these five natural E$\sigma^{70}$ promoter sequences elucidated very different abortive properties (see Table 1). Each promoter gives rise to a distinct collection of abortive RNAs with different abortive probabilities at different ITS positions (Figure 2C). To illustrate their uniqueness, we compare the position of highest abortive probability in each. The highest abortive probability value from the $P_R$ promoter is 61% at position 2, 67% at position 3 for $P_L$, 50% at position 11 for $P_{R'}$, 55% at position 8 for T5 N25, and 53% at position 3 for T7 A1. Thus, the highest abortive probability and the position at which this high barrier occurs vary from one promoter to the next.

MSAT values from these promoters varied from 6 to 15 nt. $P_{R'}$ has the longest abortive ladder (15 nt), and $P_L$ has the shortest (6 nt); the MSAT values are 10, 8, and 7 nt for N25, A1, and $P_R$, respectively. Abortive:productive ratios from these promoters also varied more than 4-fold (see Table 1). A1 gave the lowest APR, while $P_{R'}$ gave the highest. Thus, the extent of abortive initiation is highly dependent on the promoter.

The difference in abortive rates among the five promoters can be as high as 16-fold. N25 has the highest abortive rate and $P_L$ the lowest. Productive rates of these promoters also varied considerably; the A1 promoter is the most productive while $P_L$ the least. The low abortive and productive rates

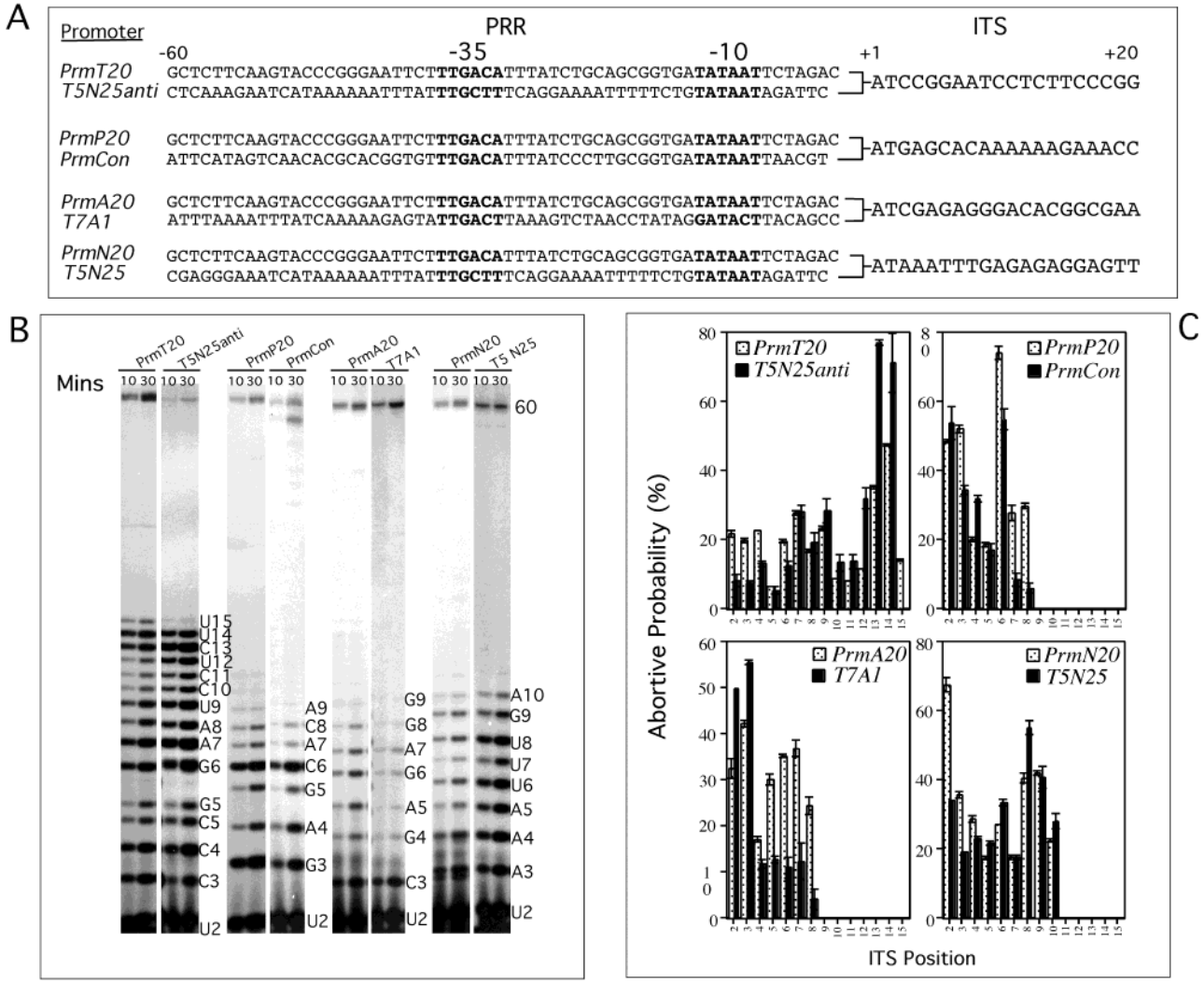


FIGURE 3: Role of the promoter recognition region (PRR) in transcript initiation. (A) Four pairs of promoters are compared; each pair shares an identical ITS from position 1 to 20, but differs in their PRR, from position −1 to −60. (B) Gel profiles of abortive and productive RNA synthesis as a function of time. (C) Pairwise comparison of the abortive probability profiles.

observed for $P_L$ are likely the result of poor polymerase binding to a linear template (*42*).

*The PRR Sequence Influences Abortive Initiation and Promoter Escape.* The effects of the general promoter sequence discussed above reflect the combined roles of the PRR and ITS. To distinguish the contribution of the PRR versus ITS, we examined the abortive properties of four pairs of templates; each pair shares an identical ITS but differs in the PRR (Figure 3A). The four pairs include T7 A1, T5 N25, T5 $N25_{antiDSR}$ (abbreviated as N25anti), and a consensus variant of $\lambda$ Prm (i.e., PrmCon), each with a counterpart containing its ITS adapted to the Prm promoter in the cloning vector pBluSK-PrmT20/$T_1T_2$ (Figure 3B), giving rise to PrmA20, PrmN20, PrmT20, and PrmP20, respectively. As constructed, PrmP20 and PrmCon contain identical ITS and consensus hexamers at −35 and −10, but differ in the sequence of the other PRR elements (see Figure 3A).

Sequence variation in the PRR caused large changes in the abortive probability profile (Figure 3C). Changes in abortive probability occurred at most of the ITS positions and can vary as much as 6-fold; for example, the abortive probability at position 8 is 5% for PrmCon and 30% for PrmP20, and at position 15, 15% for PrmT20 but 70% for T5 $N25_{antiDSR}$. Note that PrmCon and PrmP20 differ only at a few nucleotides in the spacer region and in having a 6 or 7 bp DIS, but they differ significantly in sequences upstream of the −35 element, although neither contains an apparent AT-rich UP element. The above results provide the first strong evidence linking the PRR to abortive probability at each ITS position, and the difference of a few nucleotides in the PRR is sufficient to alter the abortive properties.

Table 2: Quantitative Initiation Parameters for Pairs of Promoters that Differ in the PRR[a]

| | rate of RNA synthesis[b] | | | | |
|---|---|---|---|---|---|
| promoter | total | productive | abortive | APR[c] | MSAT (nt) |
| PrmT20 | 1125 ± 34 | 30 ± 2 | 1094 ± 33 | 36 ± 2 | 15 |
| T5 $N25_{anti}$ | 1049 ± 130 | 6 ± 2 | 1043 ± 129 | 174 ± 27 | 15 |
| PrmP20 | 774 ± 89 | 12 ± 1 | 762 ± 88 | 64 ± 5 | 8 |
| PrmCon | 452 ± 89 | 30 ± 4 | 422 ± 125 | 14 ± 4 | 8 |
| PrmA20 | 175 ± 6 | 12 ± 1 | 163 ± 5 | 13 ± 1 | 8 |
| T7 A1 | 242 ± 20 | 31 ± 6 | 212 ± 14 | 7 ± 1 | 8 |
| PrmN20 | 493 ± 55 | 10 ± 1 | 483 ± 49 | 48 ± 5 | 10 |
| T5 N25 | 888 ± 103 | 28 ± 5 | 860 ± 100 | 31 ± 5 | 10 |

[a] See the footnotes of Table 1. [b] The rate of RNA synthesis is expressed as the femtomoles of RNA molecules produced per enzyme per hour. The total rate is the sum of the abortive and productive rates. [c] APR represents the molar ratio of abortive vs productive transcripts derived from each promoter.

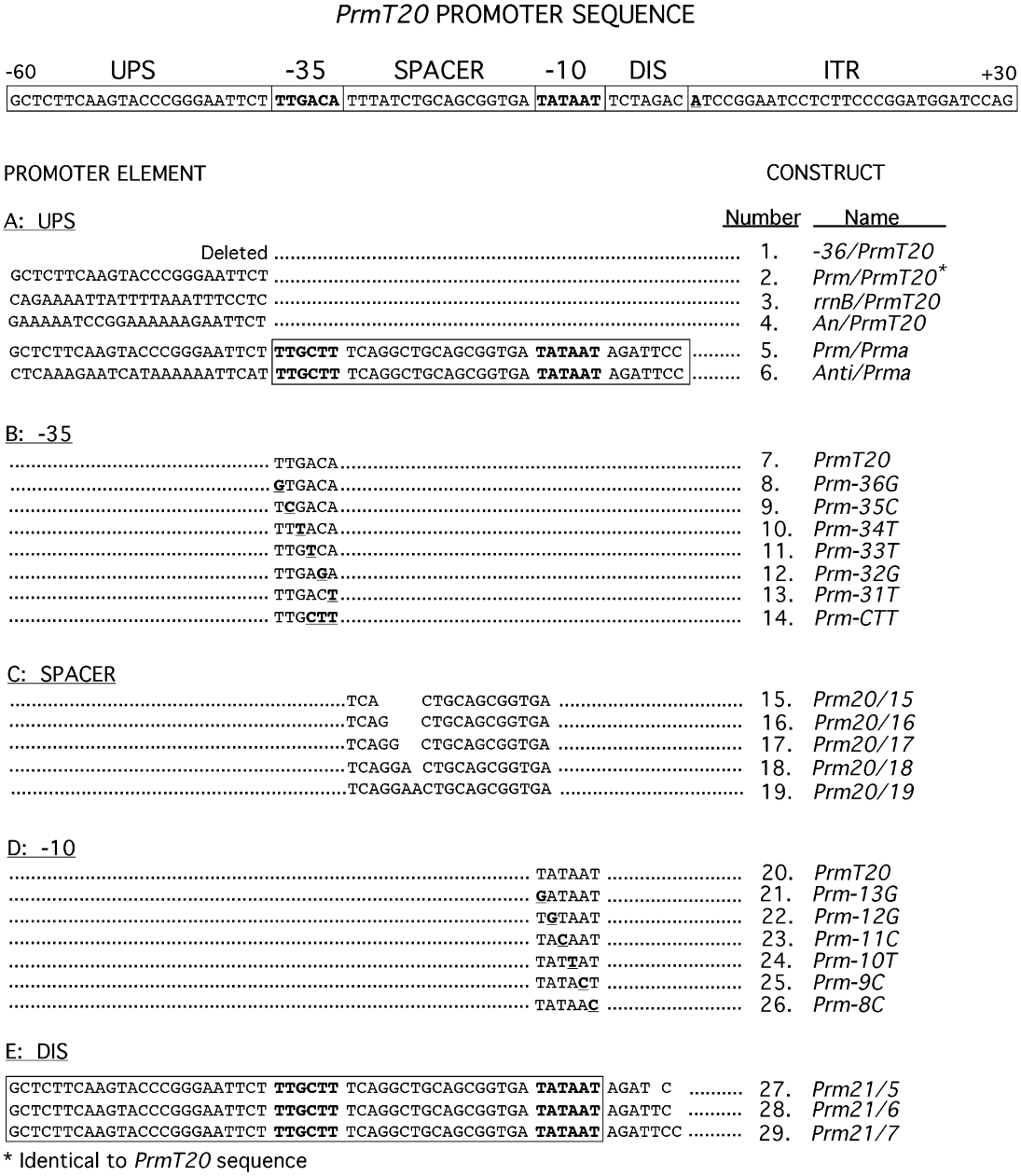


FIGURE 4: Sequence of promoter variants generated in this study. At the top is a sequence of the parental PrmT20 promoter, from position −60 to 30, as a contiguous string of discrete elements: from 5′ to 3′, UPS, −35, SPACER, −10, DIS, and ITS. The −10 and −35 consensus hexamers are highlighted in bold, and the +1 start site is underlined and bold. Below, each of five PRR elements (groups A−E) was altered systematically. Each variant is assigned a number and a name (right columns). Sequence changes in each are shown, with dots indicating a residue identical to that of PrmT20. Prm/Prma and Anti/Prma (constructs 5 and 6, respectively) share the same sequence from position −35 to −1 (boxed), but this sequence is different from that of PrmT20. The Prm21 variants are made from the Prm/Prma promoter.

The MSAT, however, was not significantly affected by the changes in the PRR that was tested. On the other hand, sequence variation in the PRR had a great impact on the productive rate but only moderately affected the abortive rate, resulting in large alterations in the APR for three of the four pairs (see Table 2). As an example, the APR of the PrmT20 promoter is 36 abortive transcripts per productive transcript while that of T5 $N25_{anti}$ is 174. Surprisingly, the difference in the APRs between PrmN20 and T5 N25 is less than 1-fold, while their PRR changes are identical to that of the PrmT20/$N25_{anti}$ pair (see Figure 3A), suggesting that ITS must also contribute to the APR.

Note that the $N25_{anti}$ ITS was considered a portable element capable of impeding promoter escape (*43*). However, in the results shown in Figure 3 (and Table 2), the anti ITS was the most optimal of the four sequences in supporting productive RNA synthesis from the Prm promoter (compare PrmT20 with PrmP20, PrmA20, and PrmN20). Thus, ITS sequences do not function independently, but exert their effect on promoter escape in a promoter-dependent manner.

*Effect of Individual PRR Elements on Abortive Initiation and Promoter Escape.* We next investigated each of the different DNA elements within the PRR, including the UPS, the −35 hexamer, the spacer DNA, the −10 hexamer, and the DIS, for their possible roles in abortive versus productive initiation, using PrmT20 as a reference promoter to make systematic changes. PrmT20 is a consensus derivative of the $\lambda$ Prm promoter outfitted with various restriction enzyme sites for cassette mutagenesis (see Figure 1) and adapted with the ITS of $N25_{antiDSR}$ (*1*, *16*); its sequence from position −60

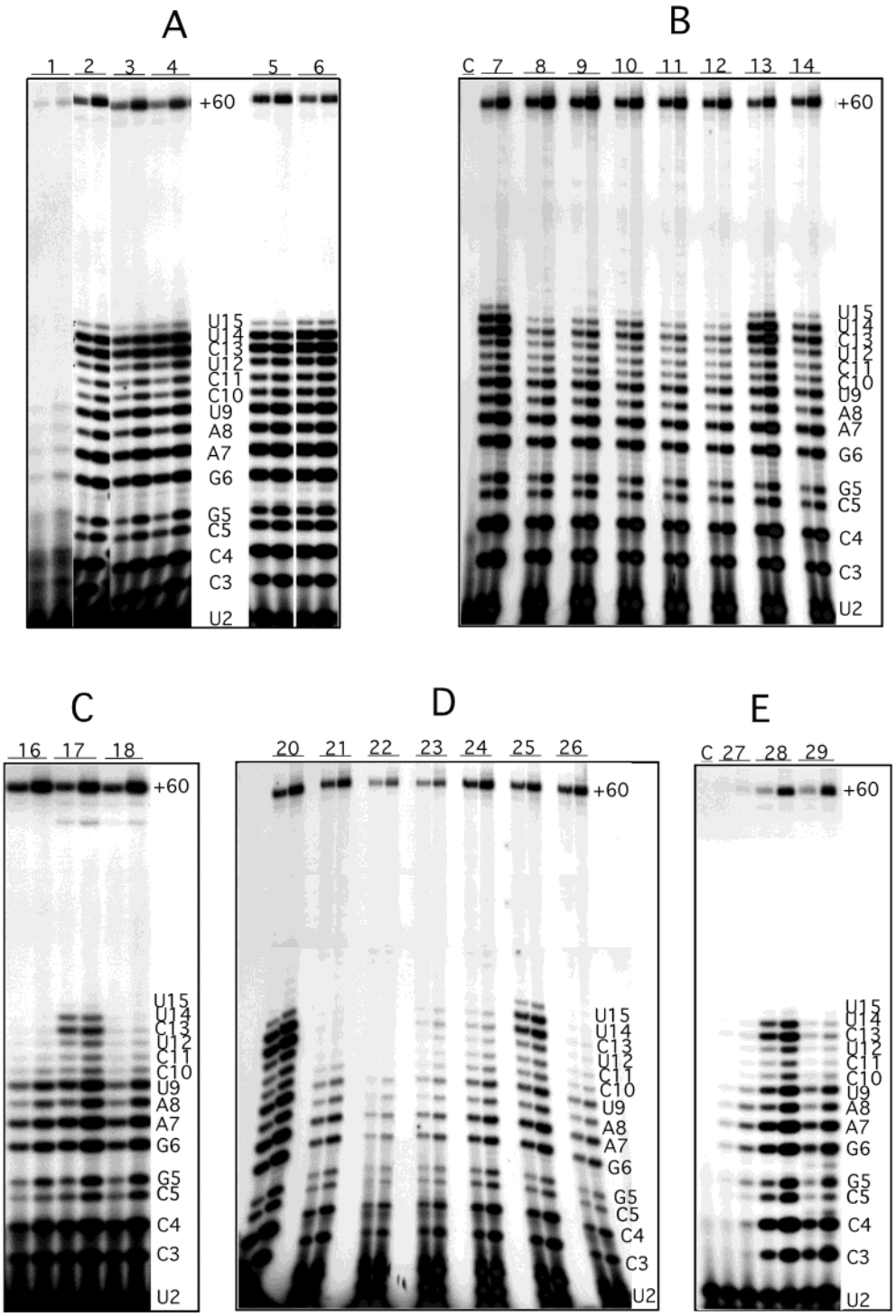


FIGURE 5: Gel profiles of abortive and productive RNA obtained from the PRR variants. For each promoter, two time point aliquots (at 30 and 60 min for A and 10 and 40 min for B−E) are shown. The promoter variants are indicated on top by their assigned construct number and organized into five panels: (A) UPS variants, (B) −35 variants, (C) SPACER variants, (D) −10 variants, and (E) DIS variants. The abortive RNAs are indicated by a letter and number combination (see the legend of Figure 2). The runoff productive transcript is 60. In panels B and E, the left-most lane is a negative control without RNA polymerase to show the background signal from free radioactive ATP.

to 30 is shown as the top line in Figure 4. The choice of the ITS from $N25_{antiDSR}$ was based on the finding that it generates an unusually long but well-characterized abortive RNA ladder and a high abortive yield in the context of the T5 N25 promoter (*6*). With its impressive abortive RNA profile, one can easily detect changes, qualitative and quantitative, that are a result of promoter mutations.

Because of the large number of promoter constructs that have been analyzed, we opt to present our results in a summary format. Sequence variations of the promoter constructs are compiled in Figure 4. Gel profiles of abortive and productive transcripts are shown in Figure 5. The abortive probability profiles for a pair of constructs within each DNA element are compared in Figure 6. Numerical values of the quantitative initiation parameters are given in Table 3.

*The UPS Element Enhances Abortive Initiation and Inhibits Promoter Escape.* The effect of the upstream

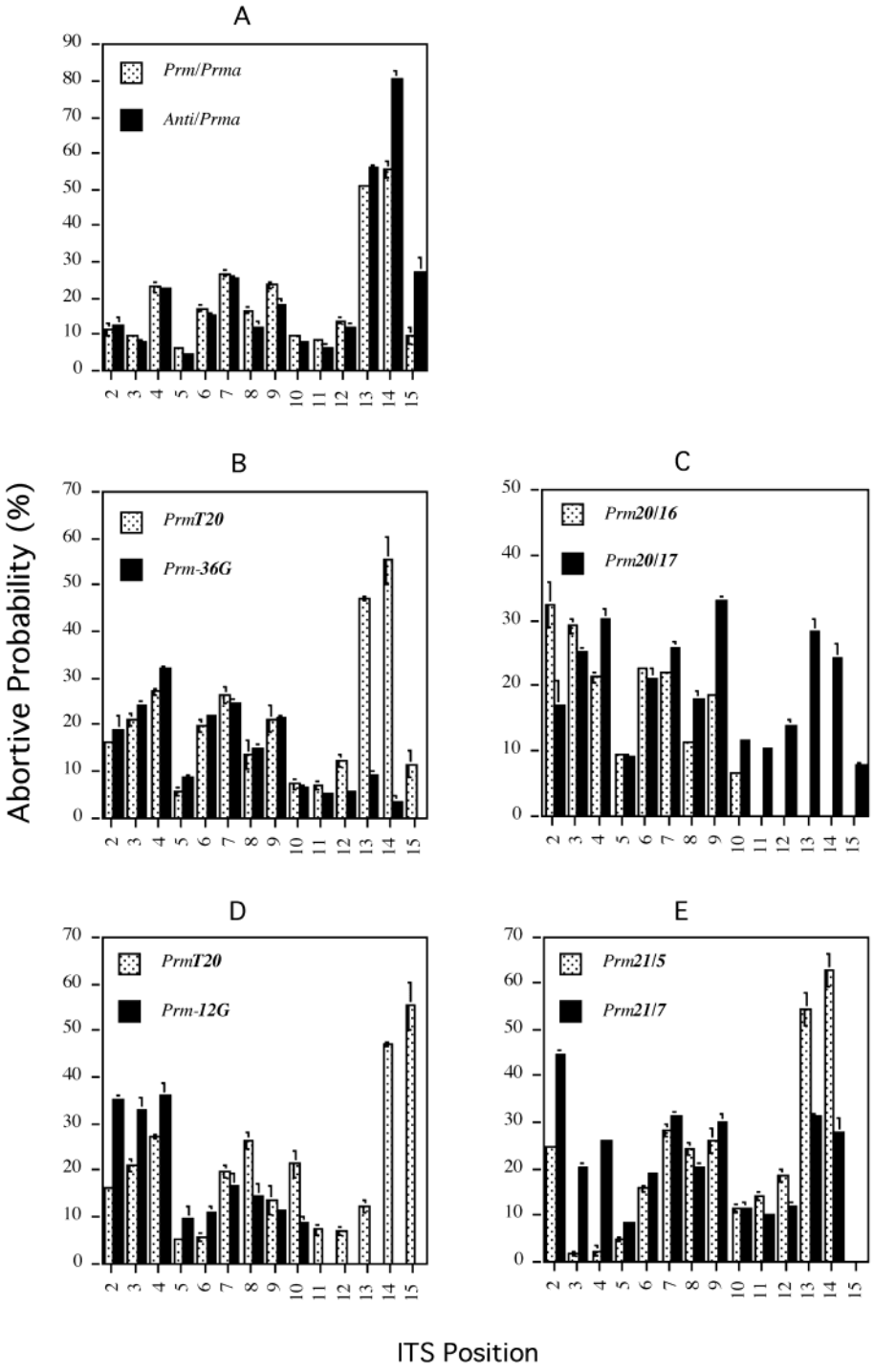


FIGURE 6: Comparison of the abortive probability profiles of two related promoter variants. The abortive probability at each ITS position was calculated as described previously (*15*). Abortive probabilities were plotted as average values from three independent experiments with one standard deviation. The *X*-axis gives ITS positions and the *Y*-axis the abortive probability in percent: (A) UPS variants, (B) −35 variants, (C) SPACER variants, (D) −10 variants, and (E) DIS variants.

sequence was investigated with six promoter variants (constructs 1−6) that differ in the extent and sequence composition of the region immediately upstream of the −35 conserved hexamer. Constructs 1−4 are based on PrmT20 (construct 2) but differ in either lacking an upstream region (−36/PrmT20) or containing the UPS from the rrnBP1 promoter (rrnB/PrmT20) or one with multiple A tracts (An/PrmT20). Constructs 5 and 6 contain a slightly altered core promoter, Prma, joined to the UPS of Prm (designated Prm/Prma) or $N25_{antiDSR}$ (designated Anti/Prma). The upstream sequence from rrnBP1, from $N25_{antiDSR}$ (which is identical to that of N25), or from a multiple A tract was chosen because their AT-rich feature has been correlated with high promoter strength (*44*−*46*).

As shown with −36/PrmT20, the physical absence of the upstream sequence produced a large effect; the total level of initiation from the promoter decreased by 3−4-fold compared to that for PrmT20 (construct 2), and the longest abortive transcript was reduced from 15 to 9 nt. As a result, its abortive probability profile was drastically altered.

Of the five promoters with contiguous upstream sequence, sequence variation did not affect the length of the abortive

Table 3: Quantitative Initiation Parameters for Variants of the PRR Elements[a]

| element altered | promoter | rate of RNA synthesis[b] | | | APR[c] | MSAT (nt) |
|---|---|---|---|---|---|---|
| | | total | productive | abortive | | |
| (A) UPS | 1. −36/PrmT20 | 345 ± 17 | 13 ± 1 | 332 ± 16 | 26 ± 1 | 9 |
| | 2. Prm/PrmT20 | 1125 ± 34 | 30 ± 2 | 1094 ± 33 | 36 ± 2 | 15 |
| | 3. rrnB/PrmT20 | 1091 ± 32 | 30 ± 2 | 1062 ± 32 | 35 ± 2 | 15 |
| | 4. An/PrmT20 | 982 ± 29 | 27 ± 2 | 955 ± 29 | 35 ± 1 | 15 |
| | 5. Prm/Prma | 839 ± 27 | 26 ± 5 | 813 ± 124 | 31 ± 3 | 15 |
| | 6. Anti/Prma | 1056 ± 131 | 14 ± 3 | 1042 ± 129 | 74 ± 10 | 15 |
| (B) −35 | 7. PrmT20 | 1125 ± 34 | 30 ± 2 | 1094 ± 33 | 36 ± 2 | 15 |
| | 8. Prm-36G | 529 ± 8 | 55 ± 3 | 474 ± 7 | 9 ± 1 | 15 |
| | 9. Prm-35C | 752 ± 94 | 78 ± 12 | 674 ± 88 | 9 ± 2 | 15 |
| | 10. Prm-34T | 743 ± 16 | 61 ± 5 | 682 ± 13 | 11 ± 1 | 15 |
| | 11. Prm-33T | 635 ± 53 | 60 ± 4 | 575 ± 51 | 10 ± 1 | 15 |
| | 12. Prm-32G | 638 ± 54 | 58 ± 5 | 580 ± 52 | 10 ± 1 | 15 |
| | 13. Prm-31T | 1431 ± 22 | 39 ± 2 | 1392 ± 21 | 35 ± 2 | 15 |
| | 14. Prm-CTT | 921 ± 37 | 57 ± 6 | 864 ± 34 | 15 ± 1 | 15 |
| (C) SPACER | 16. Prm20/16 | 449 ± 4 | 33 ± 3 | 416 ± 6 | 13 ± 1 | 10 |
| | 17. Prm20/17 | 1125 ± 34 | 30 ± 2 | 1094 ± 33 | 36 ± 2 | 15 |
| | 18. Prm10/18 | 1064 ± 97 | 41 ± 7 | 1023 ± 13 | 25 ± 1 | 11 |
| (D) −10 | 20. PrmT20 | 1125 ± 34 | 30 ± 2 | 1094 ± 33 | 36 ± 2 | 15 |
| | 21. Prm-13G | 421 ± 5 | 40 ± 3 | 381 ± 4 | 9 ± 1 | 10 |
| | 22. Prm-12G | 136 ± 3 | 17 ± 1 | 120 ± 3 | 7 ± 1 | 9 |
| | 23. Prm-11C | 344 ± 44 | 32 ± 4 | 311 ± 42 | 9 ± 1 | 14 |
| | 24. Prm-10T | 519 ± 30 | 61 ± 6 | 458 ± 27 | 7 ± 1 | 14 |
| | 25. Prm-9C | 808 ± 13 | 50 ± 3 | 758 ± 12 | 15 ± 1 | 15 |
| | 26. Prm-8C | 247 ± 3 | 43 ± 1 | 204 ± 3 | 5 ± 2 | 14 |
| (E) DIS | 27. Prm21/5 | 114 ± 8 | 3 ± 1 | 111 ± 6 | 37 ± 3 | 14 |
| | 28. Prm21/6 | 839 ± 125 | 26 ± 5 | 813 ± 124 | 31 ± 3 | 14 |
| | 29. Prm21/7 | 530 ± 65 | 17 ± 4 | 513 ± 63 | 30 ± 3 | 14 |

[a] See the footnotes of Table 1. [b] The rate of RNA synthesis is expressed as the femtomoles of RNA molecules produced per enzyme per hour. The total rate is the sum of the abortive and productive rates. [c] APR represents the molar ratio of abortive vs productive transcripts derived from each promoter.

ladder (Figure 5A) and therefore only slightly affected the abortive probability profiles. Comparing Prm/Prma and Anti/Prma (constructs 5 and 6, respectively), significant abortive probability changes were noted only for the late ITS positions (Figure 6A); in this case, the AT-rich upstream element in Anti/Prma resulted in elevated abortive probabilities of 14 and 15, presumably raising the barrier for promoter escape at these positions.

Changes in the upstream sequence, however, greatly influenced the APR. The ratio of abortive to productive rates ranged from 26 to 74, being lowest in −36/PrmT20 and highest in Anti/Prma. Since the AT-rich UP element of rrnBP1 had been shown to interact with the α subunits of RNA polymerase to enhance promoter binding (*26*), the effect of upstream sequence deletion likely reflects the missing interactions.

Taken together, these results suggest a correlation between the strength of UPS−polymerase interaction and the extent of abortive initiation; the higher the binding stability, the higher the abortive probabilities, particularly at the late ITS positions, and also the higher the abortive yield. Thus, strong UP element−RNA polymerase contacts appear to enhance abortive initiation and inhibit promoter escape. This observation is in agreement with the findings of Carpousis et al. (*47*), Ellinger et al. (*2*), and Strainic et al. (*48*).

*The −35 Element Affects the Abortive Probability Especially at the Late ITS Positions.* Binding of the *E. coli* RNA polymerase holoenzyme to a promoter site is mediated mainly through the specific interactions between the $\sigma^{70}$ subunit and the −10 and −35 conserved hexamers (*23*, *49*, *50*). Repeated compilation of E$\sigma^{70}$ promoters confirmed the **TTG**ACA consensus sequence for the −35 element and the **TA**TAA**T** sequence for the −10 element, with the bold nucleotides being the most conserved in these elements (*51*−*53*). To test the role of the −35 hexamer on abortive initiation and promoter escape, we started with the perfect consensus sequence of TTGACA$_{-31}$ in PrmT20 and introduced single-base pair changes at each of the six positions to a nonconsensus base pair (constructs 7−14, Figure 4B). Several of the sequence alterations correspond to mutations studied previously; for example, a T$_{-36}$G or T$_{-35}$C single substitution was shown to dramatically reduce the extent of open complex formation at the *lac* promoter (*54*). Only one promoter variant, Prm-CTT, contains a 3 bp change that converts the perfect consensus to the −35 box of the T5 N25 promoter (TTGCTT).

As shown in Figure 5B, single-base pair alterations of the −35 element do not significantly affect the length of the abortive RNA ladder, which differs by only 1 nt among the −35 variants. Instead, they differ greatly in the abortive probability of the late ITS positions, from 11 to 15. The abortive probability values at positions 13−15 are 60, 50, and 10%, respectively, for PrmT20, whereas they are 10, 4, and 0% for Prm-36G, respectively (Figure 6B); little or no change in abortive probability was found at positions 2−10. The lowered abortive probability at positions 13−15 in Prm-36G is indicative of the increased stability of the initial complexes at these positions and can be correlated with the enhanced ease of promoter escape to yield more productive RNA.

Interestingly, Prm-31T and PrmT20 shared an almost identical abortive probability profile, suggesting that the A$_{-31}$ position (the 3′-most nucleotide of the −35 element) is neutral in mediating the initiation functions. In agreement

with this observation, all four base substitutions at this position in the *lac* promoter were previously shown to have little effect on promoter binding (*54*).

The rates of abortive and productive RNA synthesis are strongly influenced by the −35 sequence (see Table 3). With the exception of Prm-31T, the other −35 variants all exhibited a reduced abortive rate and an increased productive rate, giving rise to an APR lower than that of PrmT20 by almost 4-fold. The productive rate of all seven −35 variants was greater than that of PrmT20. This was a surprise since these same nucleotide substitutions in the context of the *ant* or *lac* promoter resulted in reduced promoter strength (*54*, *55*).

How can these seemingly contradictory results be reconciled? The answer appears to depend on the location of the rate-limiting step during transcript initiation at each promoter. For promoters such as *lac* and *ant*, the rate-limiting step resides with promoter binding. Mutations in either the −35 or −10 hexamer that now deviate from the consensus further compromise promoter binding and consequently promoter strength. On the other hand, for a consensus promoter such as PrmT20, the promoter-binding step is already maximized; instead, the rate-limiting step occurs during promoter escape (*18*). A single-base substitution in the −35 element in PrmT20 probably reduced the level of promoter binding only slightly, but significantly facilitated promoter escape, resulting in a higher productive rate. Thus, the magnitude of the effect of a mutation on promoter strength depends on whether the mutation impacts the rate-limiting step of a given promoter (*56*−*58*).

*Quantitative Initiation Parameters Are Exquisitely Sensitive to Changes in Spacer Length.* Optimal promoter recognition and binding by *E. coli* RNA polymerase requires the −10 and −35 hexamers to be separated by a spacer length of 17 bp. This distance in the helical DNA is thought to present the −10 and −35 elements in the proper orientation for binding by the $\sigma^{70}$ subunit and has been confirmed in the recent structural model of an open complex (*25*). Variation in the spacer length from 15 to 19 bp was shown to have a substantial effect on promoter binding and productive RNA synthesis (*20*, *24*, *59*). To test the effect of spacer length on abortive initiation and promoter escape, we made five variants with spacer lengths from 15 to 19 bp in the PrmT20 promoter context (constructs 15−19, Figure 4C).

Steady state transcription revealed that variation in spacer length nearly eliminated promoter activity in Prm20/15 and Prm20/19 (gel results not shown). The other variants, Prm20/16, Prm20/17, and Prm20/18, exhibited distinct differences in the length of the abortive ladder (Figure 5C), in turn causing large changes in the abortive probability profile (Figure 6C). Compared to Prm20/17 whose longest abortive RNA is 14 nt, Prm20/16 and Prm20/18 show a drastically reduced level of the long abortive RNAs from 11 to 14 nt, suggesting that a significant fraction of RNA polymerase has undergone the promoter escape transition at these positions.

The facility at promoter escape from Prm20/16 and Prm20/18 is substantiated by their lower APR; the reduction can be as much as 3-fold when compared to Prm20/17 (Table 3). In the case of Prm20/16, the lower APR is achieved without changing the productive rate but by greatly diminishing the abortive rate and for Prm20/18 through an elevated productive rate but an unchanging abortive rate.

Interestingly, of the three variants, Prm20/17 with the optimal spacer length displays the highest APR, the longest abortive RNA ladder, and the highest abortive probability at late ITS positions. These observations reinforce the results reported above, that promoter features closest to the consensus presumably confer the strongest promoter binding, thereby causing a high level of abortive initiation by preventing promoter escape.

*The −10 Element Appears To Control the Productive Efficiency of a Promoter.* The −10 hexamer is directly involved in the steps of promoter binding and melting isomerization to form the open complex (*28*, *49*, *50*). To study the role of the −10 element, we made six variants (constructs 21−26) each deviating by a single base from the −10 consensus $TATAAT_{-8}$ in PrmT20 (construct 20, Figure 4D). The choice of nucleotide substitution at each position was based on mutations previously investigated in the *ant* promoter (*55*).

Results of steady state transcription revealed that with the exception of Prm-9C (construct 25), single-base substitution in the −10 element drastically reduced the length of the abortive ladder (Figure 5D). The longest abortive transcript in Prm-13G and Prm-12G (constructs 21 and 22, respectively) was 9 or 10 nt in length compared to 15 nt for PrmT20. Among this set of promoter variants, the abortive probability values varied greatly at positions 11−15 and only moderately at positions 2−10. Abortive probabilities at positions 13 and 14 are 50 and 60%, respectively, for PrmT20, 3 and 2% for Prm-8C, respectively, and 0% for Prm-12G (Figure 6D).

Sequence variation in the −10 element also led to a large reduction in the APR, by as much as 7-fold (see Table 3). Except for that of Prm-12G, the drop in the APR is invariably the result of an elevated productive rate and a greatly diminished abortive rate. Prm-12G (construct 22) achieved a 5-fold reduction in the APR by halving the productive rate but reducing the abortive rate by almost 10-fold. On the basis of the APR values, all −10 variants are more facile at promoter escape than the consensus promoter. Furthermore, base substitutions at the three most conserved positions ($\underline{TA}TAA\underline{T}_{-8}$) resulted in the greatest reduction in the abortive rate. Therefore, the bases directly involved in promoter binding also play important functional roles in dictating the extent of abortive initiation and promoter escape.

These same nonconsensus single-nucleotide substitutions in the −10 element, like the −35 variants described above, were shown to reduce the strength of the *ant* or *lac* promoter (*54*, *55*) but increase the productive rate from our promoter variants. This discrepancy can be reconciled by the same molecular explanation proposed above; that is, the effect of a −10 or −35 mutation is context-dependent and may increase or decrease promoter strength depending on whether the rate-limiting step of initiation is impacted.

*The Length of the Discriminator Region Greatly Alters the Overall Initiation Frequency.* The region between the −10 element and the +1 transcription start site is melted during the formation of open complexes. The sequence composition of this region is directly involved in mediating the stringent response (*60*, *61*) and, also, in the transcriptional control of certain pyrimidine biosynthetic operons (*62*, *63*).

Dubbed the discriminator sequence (DIS) by Travers (*29*), the length of this region can influence the choice of transcriptional start site. Compilation of promoter sequences revealed the length of the DIS to be 6−8 bp, with 7 bp being the most common length for transcript start site selection (*51*, *53*). To test the effect of varying the length of DIS, we used as the starting promoter Prm21/6 containing a 6 bp DIS region, and made a 5 or 7 bp variant, Prm21/5 or Prm21/7, respectively. Because the transcripts are labeled with [$\gamma$-$^{32}$P]-ATP, our analysis compares the effect of DIS length on transcription from the same start site and does not address transcriptional start from alternate sites.

Results of steady state transcription showed that the overall transcriptional activity from the same start site was drastically reduced for Prm21/5, by 5−8-fold as compared to the other two variants (Figure 5E). Despite the low initiation frequency in Prm21/5, the longest abortive transcript for all three promoters was 14 nt. Interestingly, the relative abundance of abortive transcripts from Prm21/5 differed from that of Prm21/6 and Prm21/7 at two regions, at positions 2−4 near the +1 start site and also at positions 13 and 14. Unlike all other DNA elements in the PRR that influence abortive initiation mostly at the late ITS positions (11−15), Prm21/5 exhibited dramatically lowered abortive probability at positions 3 and 4, greatly stabilizing the early initial complexes at those positions, but somewhat higher abortive probabilities at positions 13 and 14 (Figure 6E).

Changes in the length of DIS, however, had little effect on the APR; this was due to the proportionate rise and fall in the abortive and productive rates in all three promoters (Table 3). Thus, length variation in DIS appears to affect the overall level of initiation without affecting the balance of abortive versus productive synthesis. Of the three DIS variants that have been examined, Prm21/6 is the strongest promoter, yielding the highest level of productive RNA by increasing the total level of initiation; it also prescribes the highest level of abortive RNAs (Figure 5E).

## DISCUSSION

In this study, we analyzed the roles of promoter−polymerase interaction in the transcript initiation reaction by systematically examining the contribution of individual PRR elements to overall promoter activity, specifically focusing on abortive initiation and promoter escape. A recurring theme emerged from our analysis: promoter sequences that confer stronger RNA polymerase−promoter interactions invariably lead to enhanced abortive initiation and inhibited productive synthesis. Our results confirm that a consensus promoter recruits RNA polymerase to form a most stable open complex, but the enhanced stability interferes with the functional mobility of the polymerase necessary for promoter escape. The inhibition of promoter escape by strong PRR−RNA polymerase binding is invariably accompanied by enhanced abortive initiation. Thus, a causal relationship exists where the rate of promoter escape determines the extent of abortive initiation.

The dichotomy of structural stability versus functional mobility in an initiation complex has been noted in earlier studies. *In vivo* footprinting analysis of three strong synthetic promoters showed that a consensus version caused excessive stalling of RNA polymerase at the promoter to lower the extent of productive synthesis (*18*). In *Bacillus subtilis* phage $\phi$29, transcription of the late gene A2c promoter was found to stall at the stage of promoter escape by the upstream binding of protein p4 which strengthened the interaction between RNA polymerase and the −35 element (*64*). Strainic et al. (*48*) found that UAS-containing promoters undergo promoter escape at reduced rates. Since promoter escape requires the translocation of RNA polymerase downstream (*5*, *7*), strong promoter−enzyme binding must interfere directly with the ability of RNA polymerase to relinquish its multiple interactions with the PRR elements.

In this study, we used the quantitative initiation parameters (*14*, *15*) to monitor the changes in the abortive initiation and promoter escape reactions. Of the parameters, the abortive probability is most informative of the step-by-step progress of a transcription complex through the initial template region. By changing the binding strength of individual PRR elements through promoter mutations, we were able to modulate the abortive probability in defined regions of the overall profile. Thus, the DIS length influences the abortive probability at positions 2−5; the −10 sequence, positions 6−10 and 11−15; and the spacer length, the −35 sequence, and the UP element, positions 13−15. Our results suggest that transcription initiation and promoter escape from the PrmT20 promoter involve three high abortive barriers, encountered when RNA polymerase has transcribed to positions 2−4, 6−9, and 13−15 (see the aggregate of all abortive probability profiles shown in Figure 6). Furthermore, these barriers are imposed by the strong binding interactions between RNA polymerase and the various promoter DNA elements.

What is the structural context for the existence of these abortive barriers? To begin transcription, we start with the catalytically active open complex structure (*25*, *65*) where the upstream portion of the promoter DNA (from position −60 to −12) is surface-exposed and anchored to the RNA polymerase through sequence specific contacts of the UP element, the −35 hexamer, and the spacer DNA (at position -14/-15) to the $\alpha$ subunits, $\sigma_4$ and $\sigma_3$ domains, respectively. The $\sigma_2$ domain nucleates the melting at position −12 and further binds the nontemplate strand of the −10 element (*67*) to maintain the melted conformation at the upstream edge of the transcription bubble. The single strands of the transcription bubble that descend into the enzyme are bound in separate protein tunnels; after traversing past the active site, the single strands rewind downstream of position 3 where an additional 9 bp of downstream DNA is held securely in the downstream DNA binding site before it emerges from the enzyme. In addition, the $\sigma$ region 3.2 polypeptide that connects the $\sigma_2$−$\sigma_3$ and $\sigma_4$ domains does so by occupying the part of the active site crevice that forms the RNA exit channel which a nascent transcript of 8−14 nt would traverse. As situated, $\sigma$ 3.2 presents a physical block to the advancing RNA and has been suggested to be the cause of abortive initiation (*68*, *69*). This structural impediment model, if operative, can only account for a basal level of abortive initiation that probably occurs in an ITS sequence-independent but transcript length-dependent manner. It cannot account for the widely different patterns of abortive initiation we have shown, nor can it account for the existence of the three high barriers for promoter escape, the latest one (at positions 13−15) arising when the RNA exit channel is nearly filled with the nascent RNA.

Rather, a recently proposed model of abortive initiation and promoter escape based on a scrunching−translocation mechanism (*12*) allows us to comprehend the specific interactions that pose the various promoter escape barriers. By this model, as transcription occurs, downstream template DNA would be unwound and increasingly scrunched into the enzyme interior (*66*). Scrunching, however, does not lead to rewinding of the transcription bubble upstream until approximately position 10/11 where promoter escape occurs (*70*, *71*). Thus, during early transcription, cumulative scrunching would increase the size of the bubble incrementally, inducing stress in the initial transcribing complexes. The cumulative strain would distort the bubble, eliciting movement of the sugar−phosphate backbone in either the upstream or the downstream direction, in an attempt to release the strain. Movement of the transcription bubble downstream past the active site leads to backtracking, disengaging the short RNA from the template strand, resulting in abortive release. On the other hand, when enough strain has accumulated, upstream movement of the strands can force the rewinding of the bubble, allowing DNA to translocate out of the tunnel in the upstream direction, leading to promoter escape.

On most promoters, the first abortive barrier is encountered at positions 2−4. The high level of abortive initiation at these positions can be attributed to the high instability of an ITC bearing a short RNA−DNA heteroduplex. Interestingly, one DIS variant, Prm21/5, showed reduced abortive probability at positions 2−4 (see Figure 6E), suggesting that the ITCs at positions 2−4 on this promoter are actually stabilized; no other mutations analyzed in this study exhibited such an effect. The results of the DIS Prm21/5 mutation can be explained by the scrunching−translocation mechanism (*12*) as follows. Because of the shorter DIS sequence, Prm21/5 forms an open complex with a melted region that is 2 bp shorter than the average, resulting in a highly constrained transcription bubble. As RNA synthesis proceeds, downstream DNA scrunched into the enzyme would increase the bubble size to the average, thus stabilizing the early ITCs.

The second high abortive barrier occurs when the nascent RNA has grown to positions 6−9. These positions immediately precede the point of promoter escape at position 10/11 (*70*, *71*). At these lengths, the nascent RNA starts to be peeled away from the template DNA and becomes directed into the RNA exit channel (*72*−*75*). These lengths also represent the juncture when the cumulative strain from DNA scrunching may be sufficiently large to bring about the rewinding of the upstream edge of the transcription bubble (*12*, *70*, *71*). Successful rewinding of the transcription bubble disrupts $\sigma_2$ binding to the −10 element, allowing the template DNA to translocate in the upstream direction; as a result, a fraction of the ITC may undergo promoter escape at this point. However, unsuccessful rewinding of the upstream edge means a failure to displace the $\sigma_2$ binding and may lead to abortive release of the nascent RNA instead. In the −10 promoter variant Prm-12G, the strength of the binding between $\sigma_2$ and the −10 DNA has been lowered such that the rewinding of the upstream edge of the transcription bubble can be more easily achieved, leading to earlier promoter escape as evidenced by the shortening of the abortive ladder to position 9 (see the gel profile of construct 22 in Figure 5D). Similar shortening of the abortive ladder was observed with $\sigma^{70}$ region 2.2 mutants on an $N25_{anti}$ template (*43*), confirming that the strength of the interaction between $\sigma_2$ and −10 controls the ease of promoter escape and, therefore, the height of the abortive barrier of positions 6−9.

For PrmT20 and many of its UP element-containing variants, a third high abortive barrier exists at positions 13−15 (Figure 6A,B). The existence of the third barrier suggests that, for a significant fraction of the polymerase, overcoming the first and second high barriers does not result in promoter escape. Besides the $\sigma_2$ interaction with the consensus −10 DNA, these promoters all have tight binding between the consensus −35 element and the $\sigma_4$ domain and, additionally, a binding interaction between the UP element and the α subunits (*26*, *46*). Apparently, all three binding elements (at $\sigma_2$ and −10, $\sigma_4$ and −35, and α and UP) are required in a favorable spatial configuration (i.e., 17 bp spacer; see Figure 6C) to yield a highly stable open complex structure resistant to promoter escape. Weakening any one of the three interactions—whether by a complete removal of the UP element (i.e., in −36/PrmT20) or by single-base substitution of the consensus hexameric sequences, or even by straining their spatial relationship by changing the spacer length—eases the height of the third abortive barrier. Presumably, to achieve promoter escape, a larger scrunching-induced cumulative strain is necessary to bring about the displacement of $\sigma_2$−, $\sigma_4$−, and α subunit−DNA interactions all at once.

A survey of the abortive pattern of *E. coli* $E\sigma^{70}$ promoters shows that most promoters yield abortive products from 2 to 10 nucleotides in length, suggesting that they contain only the first two high abortive barriers to promoter escape. The existence of two high-barrier steps to promoter escape, at the position 3−position 4 and position 7−position 8 junctures, has been identified for T7 RNA polymerase on a T7 promoter by comparing the crystal structures of initiation and elongation complexes (*66*, *74*, *75*). At the position 3−position 4 juncture, conformational changes are required to remove a protein blockade to the advancing RNA, a situation analogous to the physical impediment posed by the *Taq* σ region 3.2 polypeptide binding in the RNA exit channel (*68*, *69*). At the position 7−position 8 juncture, dramatic conformational changes have occurred to break the promoter contacts and, also, to peel the nascent transcript off the template strand and bind it in a newly formed RNA exit tunnel. Thus, the crystal structure of the T7 RNA polymerase elongation complex has greatly clarified the molecular juggernauts an enzyme−DNA complex must overcome during the initiation−elongation transition (*74*, *75*).

Our survey further shows that it is an occasional strong promoter, for example, a consensus Prm promoter, T5 $N25_{anti}$, or λ $P_{R'}$, that prescribes a long abortive program to position 15. For PrmT20, the long abortive program is as much dependent on the PRR signals (as shown by all the results presented here) as on the anti ITS sequence. As shown in Figure 3B, a consensus Prm promoter outfitted with other ITS sequences (such as PrmP20, PrmA20, and PrmN20) produces an abortive ladder of the normal range. What is the role of the ITS in determining the extent of abortive initiation and promoter escape? For PrmT20, cassette mutagenesis of the anti ITS has not yielded clear-cut answers (*41*). For λ $P_{R'}$, the long abortive ladder precedes the polymerase pausing at position 16/17. This pause is required

for the loading of Q protein to the transcription complex to carry out Q-mediated antitermination at the late gene terminator (*76*). It has been shown that the pause at position 16 is dependent on the sequence at positions 2−6 of the ITS (*76*). During early transcription, $\sigma^{70}$ rebinds to this region of the nontemplate strand to bring about pausing (*77*). It is not clear at present whether abortive initiation and pausing in $\lambda$ $P_{R'}$ can be modulated by mutations in the PRR elements.

In this study, we have subjected each promoter variant to steady state transcription analysis. The reaction conditions have been optimized to allow each RNA polymerase to go through multiple rounds of transcription to synthesize abortive and productive RNA products in large amounts (*37*). The abundant synthesis of RNA products is necessary for the accurate determination of the quantitative initiation parameters. The steady state approach, however, presents a major caveat in that a fraction of the ITCs is "unproductive" under normal *in vitro* transcription conditions and carries out continuous abortive synthesis without undergoing promoter escape (*78*, *79*, *81*). Additionally, the unproductive fraction varies with different promoters (*81*), further generating promoter-dependent bias in the abortive RNA level and greatly complicating the interpretation of the quantitative initiation parameters.

Because of this complication, it might be argued that the quantitative initiation parameters should be determined under single-cycle transcription conditions. However, the repetitive nature of abortive cycling, even under single-cycle conditions, not only does not circumvent the formation of the unproductive ITC but also leads to an underestimation of the productive synthesis. To completely avoid this problem, one must either work under conditions in which all of the ITCs are capable of escaping the promoter, or study transcription of a single RNA polymerase molecule; neither of these conditions is easily attained.

In our study, this problem was solved by way of a compromise. We have shown that the fraction of ITC that fails to escape the promoter is relatively small (3−10%) for most promoters we have examined as long as two conditions are met. First, the NTP concentrations must be kept equal or above 100 $\mu$M, and second, ATP instead of ApU dinucleotide should be used to initiate RNA synthesis (*81*). Therefore, we have used 100 $\mu$M NTP and ATP as the initiating substrate in all of our steady state transcription assays. Under these conditions, most of the promoters gave steady state productive RNA synthesis for as long as 30 min, indicating that the fraction of unproductive ITC must be fairly small. Such a measure aside, we wish to note that the residual small fraction of unproductive ITC may not be inconsequential, and therefore, the quantitative initiation parameters determined by the steady state transcription approach must be interpreted with these cautionary constraints in mind.

## REFERENCES


1. Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986) *EMBO J. 5*, 2987.
2. Ellinger, T., Behnke, D., Knaus, R., Bujard, H., and Gralla, J. D. (1994) *J. Mol. Biol. 239*, 466.
3. Carpousis, A. J., and Gralla, J. D. (1980) *Biochemistry 19*, 3245.
4. Munson, L. M., and Reznikoff, W. S. (1981) *Biochemistry 20*, 2081.
5. Krummel, B., and Chamberlin, M. J. (1989) *Biochemistry 28*, 7829.
6. Hsu, L. M., Vo, N. V., and Chamberlin, M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 11588.
7. Carpousis, A. J., and Gralla, J. D. (1985) *J. Mol. Biol. 183*, 165.
8. Straney, D. C., and Crothers, D. M. (1987) *J. Mol. Biol. 193*, 267.
9. Metzger, W., Schickor, P., Meier, T., Werel, W., and Heumann, H. (1993) *J. Mol. Biol. 232*, 789.
10. Bar-Nahum, G., and Nudler, E. (2001) *Cell 106*, 443.
11. Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W., and Ebright, R. H. (2001) *Cell 106*, 453.
12. Hsu, L. M. (2002) *Biochim. Biophys. Acta 1577*, 191.
13. Hsu, L. M. (2002) *Nat. Struct. Biol. 9*, 502.
14. 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, R. G. Landes Co., Austin, TX.
15. Hsu, L. M. (1996) *Methods Enzymol. 273*, 59.
16. Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986) *EMBO J. 5*, 2995.
17. Levin, J. R., Krummel, B., and Chamberlin, M. J. (1987) *J. Mol. Biol. 196*, 85.
18. Ellinger, T., Behnke, D., Bujard, H., and Gralla, J. D. (1994) *J. Mol. Biol. 239*, 455.
19. Feng, G. H., Lee, D. N., Wang, D., Chan, C. L., and Landick, R. (1994) *J. Biol. Chem. 269*, 22282.
20. Mulligan, M. E., Brosius, J., and McClure, W. R. (1985) *J. Biol. Chem. 260*, 3529.
21. Kovacic, R. T. (1987) *J. Biol. Chem. 262*, 13654.
22. Helmann, J. D., and Chamberlin, M. J. (1988) *Annu. Rev. Biochem. 57*, 839.
23. Gardella, T., Moyle, H., and Susskind, M. M. (1989) *J. Mol. Biol. 206*, 579.
24. Warne, S. E., and deHaseth, P. L. (1993) *Biochemistry 32*, 6134.
25. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) *Science 296*, 1285.
26. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993) *Science 262*, 1407.
27. Rao, L., Ross, W., Appleman, J. A., Gaal, T., Leirmo, S., Schlax, P. J., Record, M. T., Jr., and Gourse, R. L. (1994) *J. Mol. Biol. 235*, 1421.
28. Siebenlist, U. (1979) *Nature 279*, 651.
29. Travers, A. A. (1980) *J. Bacteriol. 141*, 973.
30. Gonzalez, N., Wiggs, J., and Chamberlin, M. J. (1977) *Arch. Biochem. Biophys. 182*, 404.
31. Chamberlin, M., Nierman, W., Wiggs, J., and Neff, N. (1979) *J. Biol. Chem. 254*, 10061.
32. Uptain, S. M. (1997) Ph.D. Thesis, University of California, Berkeley, CA.
33. Brosius, J., Ullrich, A., Raker, M. A., Gray, A., Dull, T. J., Gutell, R. R., and Noller, H. F. (1981) *Plasmid 6*, 112.
34. Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) *BioTechniques 5*, 376.
35. Mytelka, D. S., and Chamberlin, M. J. (1996) *Nucleic Acids Res. 24*, 2774.
36. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res. 16*, 7351.
37. Chamberlin, M. J., Kingston, R., Gilman, M., Wiggs, J., and deVera, A. (1983) *Methods Enzymol. 101*, 540.
38. Arndt, K. M., and Chamberlin, M. J. (1988) *J. Mol. Biol. 213*, 79.
39. Sheen, J. Y., and Seed, B. (1988) *BioTechniques 6*, 942.
40. Josse, J., Kaiser, A. D., and Kornberg, A. (l961) *J. Biol. Chem. 236*, 864.
41. Vo, N. V. (1998) Ph.D. Thesis, University of California, Berkeley, CA.
42. Giladi, H., Igarashi, K., Ishihama, A., and Oppenheim, A. B. (1992) *J. Mol. Biol. 227*, 985.
43. Chan, C. L., and Gross, C. A. (2001) *J. Biol. Chem. 276*, 38201.
44. Gentz, R., and Bujard, R. (1985) *J. Bacteriol. 164*, 70.
45. Koo, H. S., and Crothers, D. M. (1987) *Biochemistry 26*, 3745.
46. Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W., Niu, W., Ebright, R. H., and Gourse, R. L. (1999) *Genes Dev. 13*, 2134.
47. Carpousis, A. J., Stefano, J. E., and Gralla, J. D. (1982) *J. Mol. Biol. 157*, 619.
48. Strainic, M. G., Jr., Sullivan, J. J., Velevis, A., and deHaseth, P. L. (1999) *Biochemistry 37*, 18074.
49. Siegele, D. A., Hu, J. C., Walter, W. A., and Gross, C. A. (1989) *J. Mol. Biol. 206*, 591.
50. Zuber, P., Healy, J., Carter, H. L., III, Cutting, S., Moran, C. P., Jr., and Losick, R. (1989) *J. Mol. Biol. 206*, 605.

51. Hawley, D. K., and McClure, W. R. (1983) *Nucleic Acids Res. 11*, 2237.
52. Harley, C. B., and Reynolds, R. P. (1987) *Nucleic Acids Res. 15*, 2343.
53. Lisser, S., and Margalit, H. (1993) *Eur. J. Biochem. 223*, 823.
54. Kobayashi, M., Nagata, K., and Ishihama, A. (1990) *Nucleic Acids Res. 18*, 7367.
55. Youderian, P., Bouvier, S., and Susskind, M. M. (1982) *Cell 30*, 843.
56. Record, M. T., Jr., Reznikoff, W. S., Craig, M. L., McQuade, K. L., and Schlax, P. J. (1996) in *Escherichia coli and Salmonella typhimurium*: *Cellular and Molecular Biology* (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.
57. Rojo, F. (1999) *J. Bacteriol. 181*, 2987.
58. deHaseth, P. L., Zupancic, M. L., and Record, M. T., Jr. (1998) *J. Bacteriol. 180*, 3019.
59. Stefano, J. E., and Gralla, J. D. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1069.
60. Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001) *J. Mol. Biol. 305*, 673.
61. Barker, M. M., Gaal, T., and Gourse, R. L. (2001) *J. Mol. Biol. 305*, 689.
62. Qi, F., and Turnbough, C. L., Jr. (1995) *J. Mol. Biol. 254*, 552.
63. Zhou, Y. N., and Jin, D. J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 2908.
64. Monsalve, M., Calles, B., Mencia, M., Salas, M., and Rojo, F. (1997) *Mol. Cell 1*, 99.
65. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R. H. (2000) *Cell 101*, 601.
66. Cheetham, G. M., and Steitz, T. A. (2000) *Curr. Opin. Struct. Biol. 10*, 117.
67. Marr, M. T., and Roberts, J. W. (1997) *Science 276*, 1258.
68. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) *Nature 417*, 712.
69. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) *Science 296*, 1280.
70. Holstege, F. C. P., Fiedler, U., and Timmers, H. T. M. (1997) *EMBO J. 16*, 7468.
71. Liu, C., and Martin, C. T. (2002) *J. Biol. Chem. 277*, 2725.
72. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000) *Science 289*, 619.
73. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001) *Science 292*, 1876.
74. Yin, Y. W., and Steitz, T. A. (2002) *Science 298*, 1387.
75. Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., and Yokoyama, S. (2002) *Nature 420*, 43.
76. Ring, B. Z., and Roberts, J. W. (1994) *Cell 78*, 317.
77. Ring, B. Z., Yarnell, W. S., and Roberts, J. W. (1996) *Cell 86*, 485.
78. Kubori, T., and Shimamoto, N. (1996) *J. Mol. Biol. 256*, 449.
79. Susa, M., Sen, R., and Shimamoto, N. (2002) *J. Biol. Chem. 277*, 15407.
80. Hsu, L. M., Vo, N. V., Kane, C. M., and Chamberlin, M. J. (2003) *Biochemistry 42*, 3777−3786.
81. Vo, N. V., Hsu, L. M., Kane, C. M., and Chamberlin, M. J. (2003) *Biochemistry 42*, 3787−3797.

BI026962V