Supplemental Information for

Transcription Initiation by *E. coli* RNA Polymerase: Relating Open Complex Stability, Scrunching and Hybrid length for Escape

Kate L. Henderson, Lindsey C. Felth, Cristen M. Molzahn, Irina A. Shkel, Munish Chhabra, Si Wang, Emily F. Ruff, Lauren Bieter, Joseph E. Kraft, M. Thomas Record Jr.

Departments of Biochemistry and Chemistry, University of Wisconsin, Madison, Wisconsin 53706 United States

Materials and Methods

General Procedures. Storage buffer (SB) for RNAP is 10 mM Tris base, 100 mM NaCl, 0.1 mM EDTA, 0.010 mM DTT, pH 7.5 at 4 ^{o}C and 50% v/v glycerol. For filter binding assays, wash buffer (WB) is 10 mM Tris base (pH 8.0), 100 mM NaCl,1 mM EDTA, and binding buffer (BB) is 40 mM Tris base (pH 8.0), 10 mM $MgCl_2$, 1 mM DTT, 0.1 mg/mL BSA, and 120 mM KCl.
Overexpression and Purification of RNAP. Overexpression and purification of σ^{70} subunit was performed as previously described using *E. coli* M5219 transformed with plasmid pMRG8.(1) RNA polymerase holoenzyme (RNAP) was reconstituted in SB by combining core enzyme and σ^{70} at a 1:2 mole ratio, and incubating for 1 hour at 37 ^{o}C. RNAP was stored at -20 ^{o}C overnight before use.
Preparation of Promoter DNAs. Initially, two primers containing a 13 bp overlap were annealed; a forward strand containing the wt promoter upstream region and a reverse strand containing the discriminator element and transcribed region. These annealed strands were filled in using Taq polymerase to generate dsDNA from -70 to +31 for promoters containing the λP_R discriminator, -71 to +31 for promoters containing the T7A1 discriminator, and -72 to +31 for promoters containing the rrnB P1 discriminator. This was mixed with short primers (HTOP and HBOT) to extend each end by an additional 11 bp and was amplified using polymerase chain reaction (PCR) to achieve the full length promoter template. Templates were doubly purified using QIAquick PCR purification kit (Qiagen) and agarose gel purification with a subsequent gel clean up kit (Promega). Sequences were verified using A+G sequencing ladders. A complete list of primers and templates used in this study is given in Table S1.
Designed Promoter Sequences. The promoter fragment designated $\lambda P_R(\lambda P_R)$ has the natural λP_R sequence from -60 to +1, including the UP element, -35 region, spacer region, -10 region, discriminator region (6 bp; nt strand GGTTGC), and +1 start site (T on the template strand). T7A1(T7A1) and rrnBP1(rrnBP1) promoter fragments also have their natural sequences in these regions; because of their longer discriminator regions (T7A1 7 bp, nt strand TACAGCC); rrnBP1 8 bp, nt strand GCGCCACC), their promoter regions end at -61 (T7A1) and -62 (rrnBP1). Hybrid promoter fragments, designated as promoter (discriminator), have the sequence of the parent promoter fragment upstream of the discriminator region, combined with the discriminator sequence of either λP_R, T7A1, or rrnBP1. All promoter fragments have the same, λP_R-based initial transcribed region from +1 to +24. Open regions and ITR of all variant sequences are found in Table S2. In all cases the promoter and ITR sequences are flanked by the same 18 bp sequence (downstream) and 20 bp sequence (upstream) of nonpromoter DNA.

Fragment lengths are: $\lambda P_R(\lambda P_R)$ and variants with the λP_R discriminator, 123 bp (-81 to +42); T7A1(T7A1) and variants with the T7A1 discriminator,124 bp (-82 to +42); and rrnBP1(rrnB P1),125 bp (-83 to +42). For lifetime and footprinting experiments, the ITR is that of λP_R. For transcription initiation experiments, the λP_R ITR sequence was modified to obtain single-round production of long transcripts when 3 NTP are added (ATP, UTP, GTP) by eliminating G bases on the template strand before position +17. G bases at +6 and +25 are changed to C, and the C at +17 is changed to G. The first two G bases on the resulting template strand are at +17 and +32, creating a stop at +16 and an additional stop at +31 when CTP is withheld.

Analysis of Filter Binding Data. The fraction of the initial population of open complexes remaining at each time point (θ_t) was calculated as

$$\theta_t = (cpm_t - cpm_{bkgd})/(cpm_0 - cpm_{bkgd})E$$

where *E* is the filter efficiency (0.75), cpm_0 is the filter-retained initial cpm before addition of competitor DNA (at t = 0) and cpm_{bkgd} is the filter retained background cpm in the absence of RNAP.(2) Dissociation rate constants k_d were obtained by fitting the fraction of open complexes remaining (θ_t) as a function of time to a three parameter (θ_0, θ_∞, k_d) single exponential decay (2):

$$\theta_t = \theta_0 + (\theta_\infty - \theta_0)\exp(-k_d t)$$

Analysis of MnO_4 Footprinting Gels. GE ImageQuant software was used obtain peak heights of all thymines in the open region from line scans. Peak heights were normalized by comparison to the total pixel count of a group of short background bands in the ITR (common to all promoters studied) to correct for loading differences between lanes and gels. For quantitative comparisons of $KMnO_4$ reactivities, relative reactivities were determined by dividing normalized peak heights by that of the least reactive thymine (+1T of T7A1(T7A1) promoter).

Table S1. Primer and Template Sequences

λP_R wt (5'-3') -71 to -12	CCACGAATTCGGATAAATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTG
T7A1 wt (5'-3') -72 to -14	CCACGAATTCAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAG
rrnB P1 wt (5'-3') -73 to -14	CCACGAATTCGTCAGAAAATTATTTTAAATTTCCTCTTGTCAGGCCGGAATAACTCCCT
λP_R wt (3'-5') +31 to -24	TGGAGACCGCCACTATTACCAACGTACATGATTCCTCCAACATACTTCGAAAACA
T7A1 wt (3'-5') +31 to -26	CAGATTGGATATCCTATGAATGTCGGTACATGATTCCTCCAACATACTTCGAAAACA
rrnB P1 wt (3'-5') +31 to -27	GCCTTATTGAGGGATATTACGCGGTGGTACATGATTCCTCCAACATACTTCGAAAACA
λP_R(T7A1) (3'-5') +31 to -25	TGGAGACCGCCACTATTAATGTCGGTACATGATTCCTCCAACATACTTCGAAAACA
λP_R(rrnB P1) (3'-5') +31 to -26	TGGAGACCGCCACTATTACGCGGTGGTACATGATTCCTCCAACATACTTCGAAAACA
T7A1(λP_R) (3'-5') +31 to -25	CAGATTGGATATCCTATGACCAACGTACATGATTCCTCCAACATACTTCGAAAACA
T7A1(rrnB P1) (3'-5') +31 to -27	CAGATTGGATATCCTATGACGCGGTGGTACATGATTCCTCCAACATACTTCGAAAACA
rrnB P1(λP_R) (3'-5') +31 to -25	GCCTTATTGAGGGATATTACCAACGTACATGATTCCTCCAACATACTTCGAAAACA
rrnB P1(T7A1) (3'-5') +31 to -26	GCCTTATTGAGGGATATTAATGTCGGTACATGATTCCTCCAACATACTTCGAAAACA
HTOP	CCAGCATTCCTCCACGAATTC
HBOT	CACCTGCACCGACAAAAGCTT
λP_R wt TXN (3'-5') +31 to -24	TGGAGACCGCCACTATTACCAACGTACAT**C**ATTCCTCCAA**G**ATACTTC**C**AAAACA
T7A1 wt TXN (3'-5') +31 to -26	CAGATTGGATATCCTATGAATGTCGGTACAT**C**ATTCCTCCAA**G**ATACTTC**C**AAAACA
λP_R(T7A1) TXN (3'-5') +31 to -25	TGGAGACCGCCACTATTAATGTCGGTACAT**C**ATTCCTCCAA**G**ATACTTC**C**AAAACA
T7A1(λP_R) TXN (3'-5') +31 to -25	CAGATTGGATATCCTATGACCAACGTACAT**C**ATTCCTCCAA**G**ATACTTC**C**AAAACA
HBOT TXN	CACCTGCACCGACAAAA**C**CTT
TXN sequence (+1 to +42)	AUGUA**G**UAAGGAGGUU**C**UAUGAAG**G**UUUUGUCGGUGCAGGUG
(5'–3') -60 to -13 λP_R wt	CGGATAAATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGG
(5'–3') -61 to -14 T7A1 wt	CAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAG
(5'–3') -62 to -15 rrnB P1 wt	CGTCAGAAAATTATTTTAAATTTCCTCTTGTCAGGCCGGAATAACTCCC

Statistical Calculation of ^{32}P-UTP Incorporation

The probability ($F(n)$) of ^{32}P-UTP incorporation in a transcript length containing n U residues was calculated using the binomial density function:

$$F(n) = \sum_{x=0}^{n} \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x}$$

where p is the probability of α-^{32}P-UTP incorporation at any site, q is the probability of incorporating a non-radiolabeled UTP at that site, n is the number of U sites in a given transcript length, and x is the number of radiolabeled U. In our case, 17.5 nM α-^{32}P-UTP and 10 µM UTP were used in the reaction solution, so the probability for α-^{32}P-UTP incorporation was 0.00175 (and 0.99825 for UTP) per U site. The binomial density function accounts for the statistical chance to incorporate α-^{32}P-UTP at multiple U sites for those transcripts containing more than one U site (i.e. any RNA length greater than 3 nucleotides). Table S1 contains the ITR sequence with modified bases underlined, and the probabilities used in our analysis for each transcript length are found in Table S2.

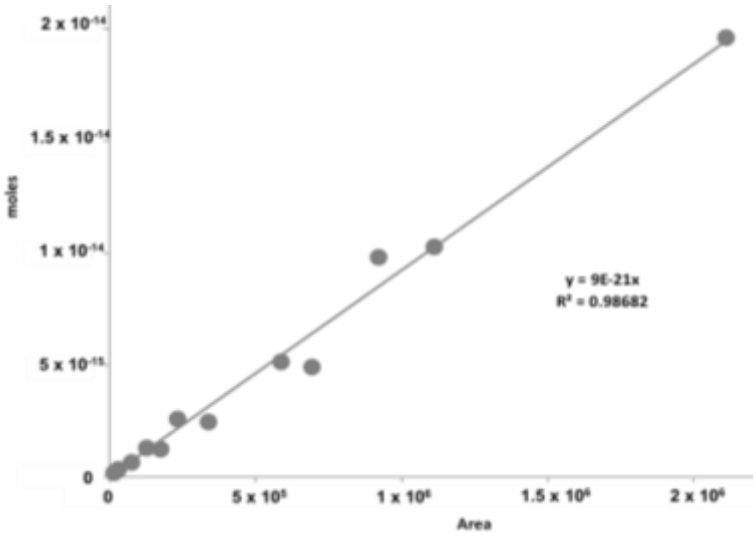


Figure S1. Calibration Curve of the α-^{32}P-UTP Standard Line. Known quantities of α-^{32}P-UTP (Perkin Elmer, Waltham MA: see website for information regarding concentrations and decay) in 5 µL volumes were run on a 40% (19:1) polyacrylamide gel long enough for the leading dye band to migrate approximately 30 cm. Imaging and line scans were obtained using the same procedures as experimental gels (see Materials and Methods). The peak area versus quantity (in moles) of α-^{32}P-UTP is plotted and fit to a linear slope. This slope is used in experiments for the conversion of peak area to the observed number of moles for each gel band.

Table S2. Calculated Probabilities of α-^{32}P-UTP Incorporation.

RNA Length (nt)	*n*	*F(n)*
2 & 3	1	0.00175
4-6	2	0.00350
7-14	3	0.00524
15	4	0.00698
16	5	0.00884
31	12	0.02080

The ITR used in our transcription experiments was a modified ITR of λP_R and contains 12 U in the longest fragment observed (31-mer).

Peak areas from line scans were first converted to moles of observed product, then divided by the probability of α-^{32}P-UTP incorporation to obtain the actual amount of moles transcript produced during the reaction. The mole number was divided by the volume of solution applied to the wells of the gel (5 µL), and then multiplied by 2 to account for the dilution of the reaction solution when mixed with the quench dye. The final number obtained reflected the molar concentration of RNA transcript produced during the reaction time. A sample calculation for the λP_R (λP_R) promoter 3- and 16-mer is listed below (Table S3).

Table S3. Sample Calculations for the Conversion of Peak Area to nM RNA

RNA	Peak Area	Observed moles[a]	Total moles[b]	nM RNA in electrophoresis sample	nM RNA in reaction mixture
3	9129	8.22E-17	4.69E-14	9.39	18.8
16	8733	7.86E-17	8.89E-15	1.78	3.6

[a] ^{32}P-labeled RNA

[b]Labeled + Unlabeled using the probability of incorporating a labeled UTP (see Table S2).

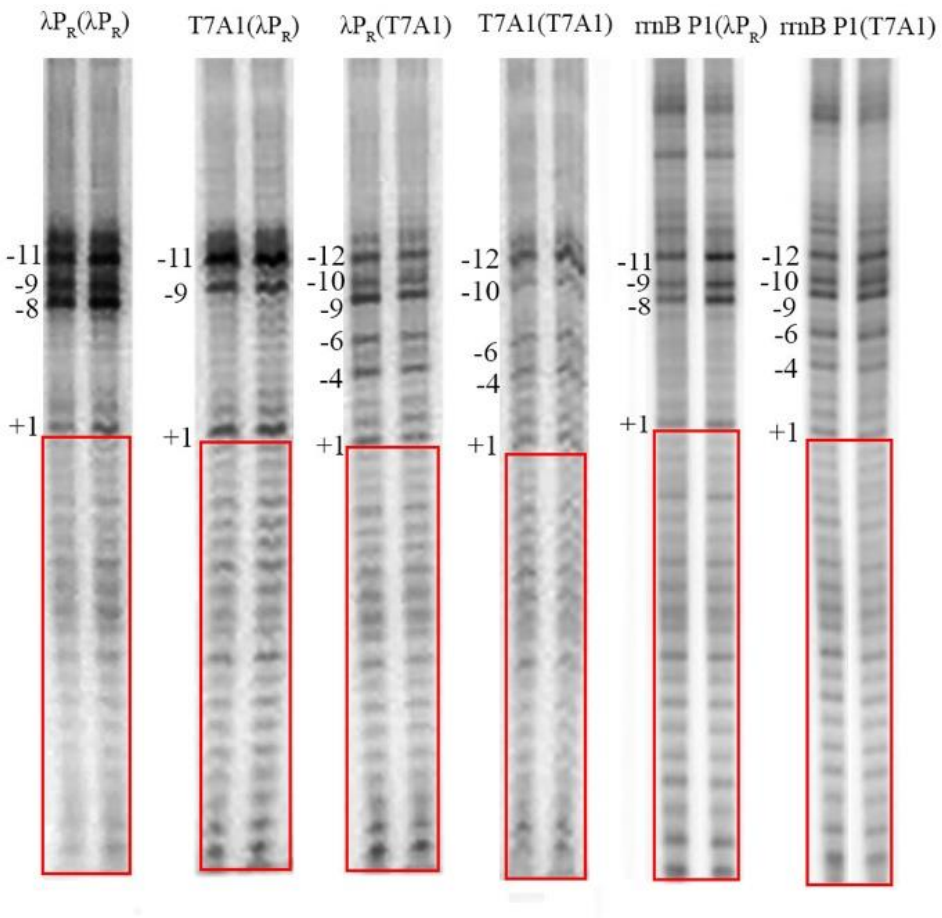
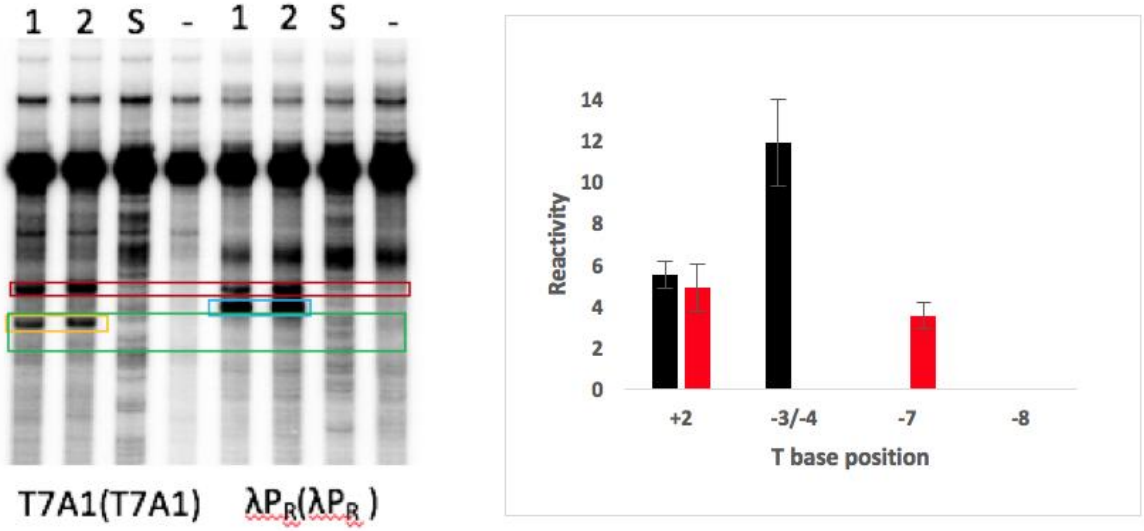


Figure S2. Quantitative Permanganate Footprinting of the Template Strands of Open Promoter Complexes. Promoter variants and positions of MnO_4^--reactive t-strand thymines are indicated on the gels. Experimental replicates for each promoter(discriminator) variant are shown. All experiments were at the same dose of MnO_4^- and intensities of all lanes were adjusted visually so that the overall intensities of the block of background bands in the common initial transcribed region of all six promoters (shown in red outline) are similar. This allows visual comparison of intensities of the reactive T at the TSS (+1) and in the -10 region (from -11 to -7 for promoters with the 6 bp λP_R discriminator and -12 to -8 for promoters with the 7 bp T7A1 discriminator).



Supplemental Figure S3. Nontemplate strand MnO_4^- footprinting for $\lambda P_R(\lambda P_R)$ and T7A1(T7A1) promoters. See Methods for dose, competitor, and analysis. **Left:** Two experiments are shown for each promoter, along with sequencing lanes and negative controls. Highly reactive +2 is boxed in red, -3/-4 of λP_R is boxed in blue. The -10 region is boxed in green. The reactive T at position -7 on T7A1 is shown in yellow. **Right:** nt strand reactivity comparison between λP_R (black) and T7A1 (red) discriminators. There is no signal for -7T and -10T on $\lambda P_R(\lambda P_R)$ and -8T and -11T on T7A1(T7A1).

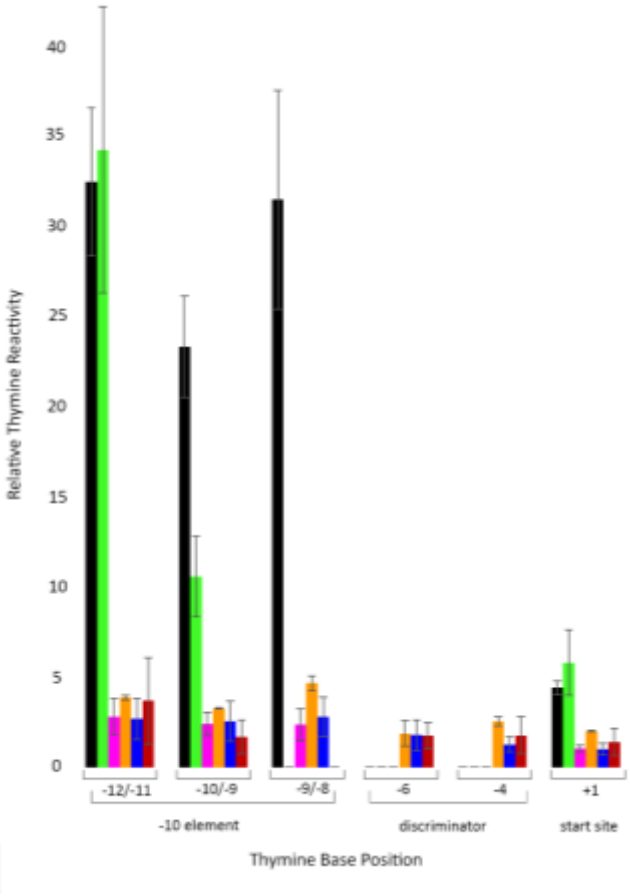


Figure S4. Bar-graph Comparison of MnO_4^- Reactivities of Template Strand -10, Discriminator and Start Site (+1) Thymines of Discriminator Variant Promoters.

MnO_4^- reactivities of t-strand T bases relative to that of T7A1(T7A1) +1 T are shown for all T positions in the -10 element, the discriminator region (only for the T7A1 discriminator), and +1 on the template strand. Details of the calculation of reactivities are provided in Methods.

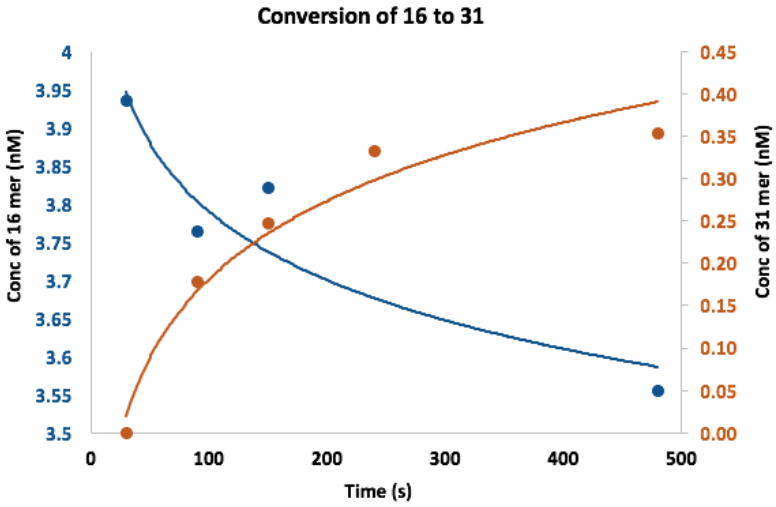
Table S4. Open Complex Dissociation Rate Constants and Lifetimes

Promoter	k_d (s^{-1})	Lifetime
$\lambda P_R(\lambda P_R)$	2.2 (± 0.3) x 10^{-5} [a,b]	12.6 hrs
T7A1(λP_R)	3.5 (± 0.6) x 10^{-5} [c]	7.9 hrs
rrnB P1(λP_R)	4 (± 1) x 10^{-4} [c]	41.7 min
λP_R(T7A1)	6.4 (± 0.9) x 10^{-4} [c]	26 min
rrnB P1(T7A1)	1.3 (± 0.3) x 10^{-3} [c]	12.8 min
T7A1(T7A1)	2.7 (± 0.6) x 10^{-3} [c]	6.2 min

[a] Data from Roe et al (ref).

[b] Data were obtained using 50-100 µg/mL heparin as an inert competitor.

[c] Data were obtained using an excess (10-fold over the concentration of RNAP) of λP_R+UP as an inert competitor.



Supplemental Figure S5. Conversion rates of 16-mer to 31-mer. Plots of the decay of 16-mer (blue line, left axis) and increase in 31-mer (orange line, right axis) during the slower phase of transcription initiation from the gel shown in Figure 3. The sum of 16-mer and 31-mer concentration (average values plotted in Figure 4) does not vary with time in the slower phase. Data points were fit to a single exponential using Origin software which yielded similar first order rate constants for the synthesis of 31-mer ($k = 0.010 \pm 0.003$) and decay of 16-mer ($k = 0.012 \pm 0.003$), respectively. This is consistent with the rate of misincorporation of a NTP when only a subset of NTPs are present (3).

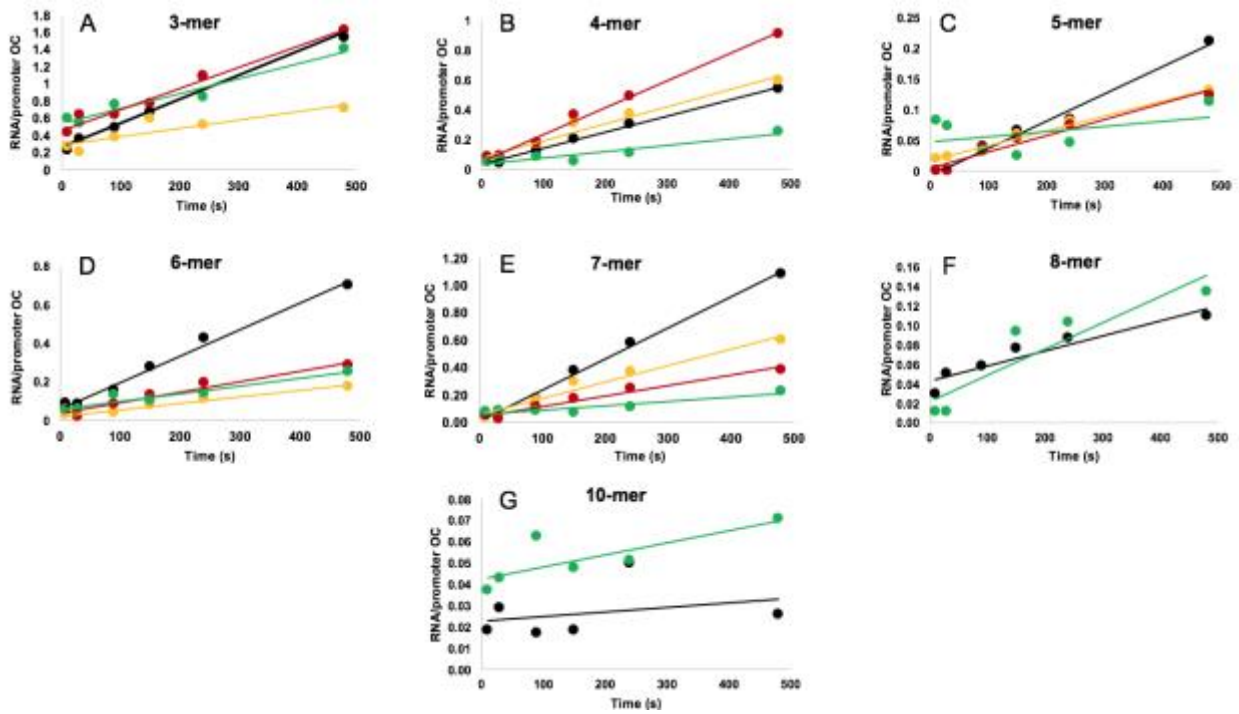


Figure S6. Rates of Synthesis of Short RNA. A-G are plots from the gel of Figure 3 for the time course of the synthesis of each short RNA (3-mer to 10-mer) at each promoter. The average rates for all experiments show similar behavior (see Table S7). Intercepts of linear fits at t=0 predict the number of RNA synthesized per OC in the initial phase (average values reported in Table S6). The slopes of these linear fits are steady-state rates of synthesis of each RNA species (renormalized average values reported in Table S7). For each length, (•) $\lambda P_R(\lambda P_R)$, (•) λP_R(T7A1), (•) T7A1(λP_R), and (•) T7A1(T7A1).

Table S5. Amounts of Short and Long RNA Synthesized per OC in the Initial Phase (t <30 s)

RNA Length	$\lambda P_R(\lambda P_R)$	T7A1(λP_R)	T7A1(T7A1)	λP_R(T7A1)
3	0.25 (±0.01)	0.56 (±0.07)	0.33 (±0.08)	0.25 (±0.02)
4	0.04 (±0.02)	0.02 (±0.03)	0.15 (±0.03)	0.07 (±0.02)
5	0.01 (±0.01)	0.02 (±0.04)	0.01 (±0.01)	0.03 (±0.01)
6	0.05 (±0.01)	0.08 (±0.05)	0.07 (±0.01)	0.02 (±0.01)
7	0.01 (±0.01)	0.05 (±0.02)	0.04 (±0.01)	0.5 (±0.02)
8	0.04 (±0.01)	0.06 (±0.03)		
10	0.02 (±0.06)	0.01 (±0.01)		
long	0.43 (±0.02)	0.45 (±0.04)	0.33 (±0.01)	0.31 (±0.02)

Tabluated values are the averages of extrapolated intercept values (t = 0) for 2-4 experiments like that shown in Figure 3 and analyzed in Figure S6. Long RNA includes all RNA lengths greater than 10 nucleotides.

Table S6. Rates of Short RNA Synthesis in the Slower (Abortive) Phase

(Rates expressed as RNA synthesized per abortively-initiating OC per 1000s)

RNA Length	$\lambda P_R(\lambda P_R)$	T7A1(λP_R)	T7A1(T7A1)	λP_R(T7A1)
3	4.2 ($\pm$1.4)	2.6 ($\pm$0.92)	4.2 ($\pm$0.43)	1.7 ($\pm$0.29)
4	1.6 ($\pm$0.31)	0.68 ($\pm$0.11)	2.2 ($\pm$0.16)	1.6 ($\pm$0.03)
5	0.59 ($\pm$0.16)	0.17 ($\pm$0.03)	0.49 ($\pm$0.07)	0.28 ($\pm$0.11)
6	1.8 ($\pm$0.44)	0.45 ($\pm$0.45)	0.73 ($\pm$0.04)	0.44 ($\pm$0.08)
7	3.5 ($\pm$0.59)	0.39 ($\pm$0.26)	1.2 ($\pm$0.07)	1.3 ($\pm$0.47)
8	0.19 ($\pm$0.09)	0.22 ($\pm$0.29)		
10	0.08($\pm$0.05)	0.05 ($\pm$0.07)		

Tabulated rates are obtained from the averages of slopes of 2-4 experiments like that shown in Figure 3 and analyzed in Figure S6. In all cases, rates are normalized by the fraction of abortively-initiating complexes (0.57 $\lambda P_R(\lambda P_R)$, 0.55 T7A1(λP_R), 0.67 T7A1(T7A1), 0.69 λP_R(T7A1)), determined from intercepts of the full length RNA data in Figure 4B (See Table S5).

The A/P Ratio of Short (Abortive, A) to Long (Productive, P) RNA Increases with Time

Figure 4A in the main text shows that the ratio of short (abortive, A) to long (productive, P) RNA (the A/P ratio) increases with time because long RNA synthesis occurs rapidly and is constrained to be single round, while short RNA synthesis is reiterative. In the initial phase of initiation, this ratio is small, in the range from ~1 ($\lambda P_R(\lambda P_R)$, λP_R(T7A1)) to ~2 (T7A1(λP_R), T7A1(T7A1)). At 480 s, the A/P ratio has increased to ~5 (T7A1(λP_R)), ~7 (λP_R(T7A1)), ~9 ($\lambda P_R(\lambda P_R)$, and ~11 (T7A1(T7A1)). In 1 hr transcription assays in which both short and long RNA synthesis are multi-round, abortive/productive (A/P) ratios for $\lambda P_R(\lambda P_R)$ and T7A1(T7A1) promoters with their natural ITR are 8 and 7 RNA/OC, respectively, highly similar to our values obtained at 480 s (4). Using our production rates, Much larger A/P ratios (44 for $\lambda P_R(\lambda P_R)$; 67

for T7A1(T7A1)) are predicted from 1 hr assays in which long RNA synthesis is single round. A plausible explanation of the very large difference in A/P ratio between the two assays (when compared at the same assay time) is that the time required for a round of long RNA synthesis and release (~10 s) is much less than that required for a round of short RNA synthesis and release of short RNA (>100 s). Even though each reinitiation produces a mixed population of ITC, in which only a minority of RNAP can escape to make another full-length RNA, this process is sufficiently faster than short RNA synthesis/release to increase the full-length RNA population and reduce the A/P ratio greatly.

Free Energy Analysis of Scrunching of the Open Strands of Promoter DNA

We model DNA scrunching as bending of elastic rod (flexible chain) models of the nt and t strands. In the absence of detailed structural information, we assume that deformation (measured by the bend angle θ in radians) is uniform along the bent region of each strand, so that the strand is the arc of a circle of length L and chord length C (see Fig S7, panel A).

From geometrical considerations, θ is related to L and C by

$$\sin\left(\frac{\theta}{2}\right) = C\theta/2L \qquad \text{S1}$$

Value of θ are determined numerically from known values of C and L.

The bending free energy of an elastic rod G_{bend} is proportional to θ^2/L:

$$G_{bend} = \frac{k\theta^2}{L} \qquad \text{S2}$$

The proportionality constant k = RTp/2 where p is the persistence length (in the range 7-30 Å for ss DNA with various base compositions at various ionic conditions (5, 6) and 500 Å for duplex DNA (7), R is the gas constant and T is Kelvin temperature. G_{bend} is calculated for each strand from L, C and numerical values of θ (from Eqs. S1, S2). While Eq S2 specifies the dependences of G_{bend} on θ and L, other considerations are needed to obtain the dependence of G_{bend} for each strand on the size of the RNA-DNA hybrid N_H.

We consider first the small deformation limit to obtain an analytical expression for G_{bend} for each strand as a function of the RNA-DNA hybrid length N_H, and then give results of numerical solution of Eqs. S1-S2 valid at any deformation as long as bending is uniform.

For small deformations (θ/2 << 1), use of the McLaurin series expansion of the sine function yields the approximate analytical results

$$\theta = 4.90\,(1 - C/L)^{0.5} \quad \text{and} \quad G_{bend} = 24(k/L)(1 - C/L). \qquad \text{S3}$$

We analyze the situation where the discriminator length is 6 bases and there is no prescrunching of the open strands. From measured distances between the downstream end of the -10 region (-7) and the +1 residue obtained from the modeled strands in an initiation complex (8), we find that the average axial distance between adjacent residues b_{ss} is approximately 4 Å. The choice of b_{ss} in the range 3.0 Å to 6.0 Å doesn't significantly affect the conclusions of this analysis (see below). For the nt strand (Figure S7, panel B), each step of translocation of downstream DNA into the active site increases the amount of ss DNA by one residue and hence increases the arc length of ss nt DNA by b_{ss} = 4 Å. For the t strand (Figure S7, panel C), the amount of ssDNA of course does not change in translocation. Each step of translocation increases the RNA-DNA hybrid length by b_{ds} = 3.2 Å, the axial distance between residues of mixed-sequence A-form RNA-DNA hybrids (9, 10), and therefore reduces the chord length of the arc formed by the discriminator residues by 3.2 Å. Because dsDNA is much more rigid than ssDNA, we assume that for the t strand all the deformation occurs in the single strand, and not in the hybrid duplex.

In the small deformation limit (corresponding to θ/2<<1 and N_H << 2):

$$G_{bend}^{nt} = \frac{4k}{b_{ss}}\frac{X}{(1+X)^2} \approx \frac{4k}{b_{ss}}X \quad and \quad G_{bend}^{t} = \frac{4k}{b_{ss}^2}b_{ds}X \quad \text{where } X = \frac{N_H}{6}\,,\; k = \frac{RTp}{2} \qquad \text{S4}$$

Hence,
$$G_{scrunch} = G_{bend}^{nt} + G_{bend}^{t} = \frac{4k}{b_{ss}}(1 + b_{ds}/b_{ss})X \qquad \text{S5}$$

Eqs S4-5 predict that, for small deformations, both $G_{scrunch}$ and the individual strand contributions increase linearly with RNA-DNA hybrid length N_H. At 37 °C, for b_{ss} = 4 Å and b_{ds} = 3.2 Å, the small-deformation prediction is $G_{scrunch}/N_H$ = 93 p where p is in Å.

Numerical calculations of G_{bend}^{nt} , G_{bend}^{t} and $G_{scrunch}$ are given in Figure S8 for the models discussed above. For b_{ss} = 4 Å, calculations were performed for N_H < 6, above which a more complicated model is needed for t strand bending. Figure S8 shows that the dependence of G_{bend}^{nt} on N_H is approximately hyperbolic, while the dependence of G_{bend}^{t} on N_H is approximately parabolic. The sum G_{bend}^{nt} + G_{bend}^{t} = $G_{scrunch}$ is very well described as linear over a wide range of deformations. The slope is $G_{scrunch}/N_H$ = 73 p kcal/mol, similar to the initial slope (93p) as shown in Fig S8. From comparison of the linear slope with the experimental value of approximately 1.0 kcal/mol, the effective persistence length of the DNA strands in the RNAP cleft is approximately 14 Å, which is in the experimental range. Since the value of b_{ss} is an informed estimate, we examined the possible range from 3.0 Å to 6.3 Å. In all cases the plot of $G_{scrunch}$ vs. N_H is linear. As b_{ss} increases from 3.0 Å to 6.3 Å, the range of N_H which can be analyzed with the uniform bending model increases from 5 to 11 bp and the calculated persistence length p increases from 8 Å to 34 Å.

Figure S8 shows that contributions of t and nt strands to the scrunching free energy increment $\Delta G_{scrunch}/\Delta N_H$ are similar in the small deformation limit (θ/2 << 1, corresponding here to N_H < 2). However at larger hybrid lengths the slope $dG_{scrunch}/dN_H$ becomes completely determined by scrunching of the t strand, as previously predicted (8).

A scrunching free energy change of 1.0 kcal/mol per translocation step of 3.2 Å indicates that the force that scrunches the two strands is approximately 20 pN, comparable to that exerted in single molecule force experiments to melt a nucleic acid hairpin (12 to 18 pN (11-13)).

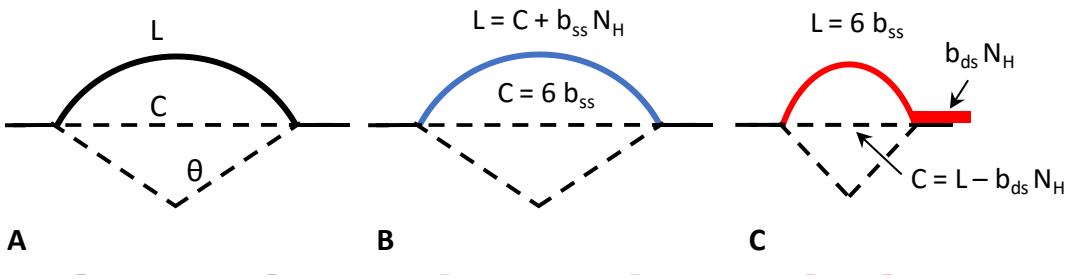


Figure S7. Modeling Scrunching as Smooth Bending of Strands of Open Region for a 6 base discriminator (no pre-scrunching). Panel A) Geometrical parameters of smooth bending: arc length L, chord length C, deformation angle θ. **Panel B)** Bending nt Strand: With each translocation step of initiation, L increases by the single-strand residue distance b_{ss} (4 Å), while C = 6 b_{ss} = 24 Å is constant. **Panel C)** Bending of t Strand: With each translocation step, C decreases by b_{ds} =3.2 Å while L = 6 b_{ss} = 24 Å is constant.

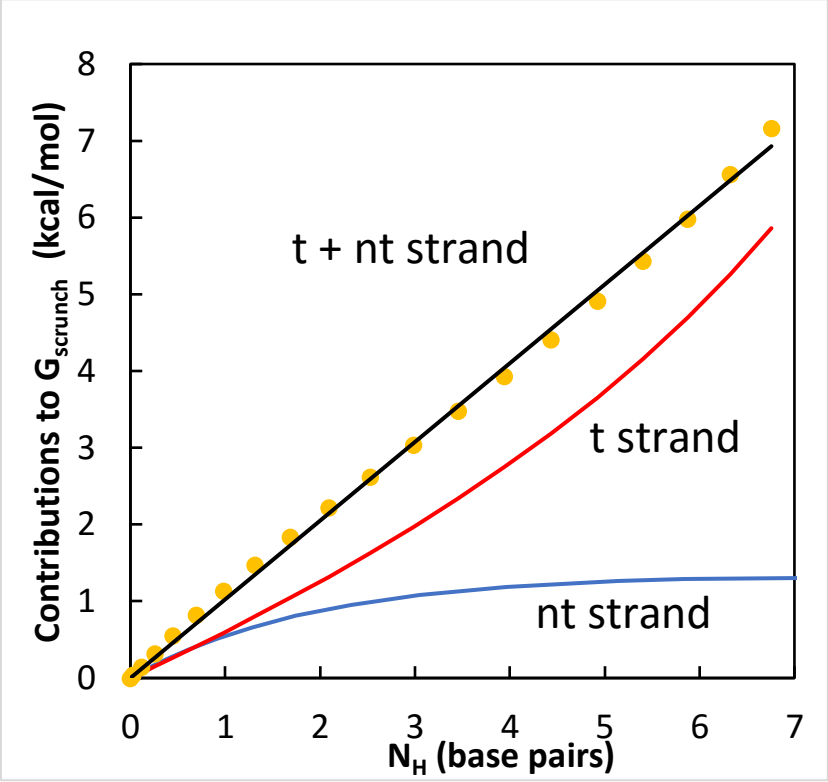


Figure S8. Dependence of scrunching free energy $G_{scrunch}$ on N_H, the number of RNA-DNA base pairs in the hybrid. Blue curve: contribution G_{bend}^{nt} from nt strand. Red curve: contribution G_{bend}^{t} from t strand (blue curve). Yellow points and black line: $G_{scrunch}$ calculated as the sum G_{bend}^{nt} + G_{bend}^{t}. Calculated for p = 14 Å, b_{ss} =4.0 Å, b_{ds} =3.2 Å, T=37 °C.

Table S7. Relationship between Open Complex (OC) Stability and Hybrid Length for Promoter Escape

Promoter	OC Formation Rate Constant (k_a; $M^{-1}s^{-1}$)	OC Dissociation Rate Constant (k_d; s^{-1})	OC Stability[a] (kcal/mol)	Hybrid Length at Escape $N_{H, ESC}$
λP_R	3.8×10^6 [(b)]	$2.2 (\pm 0.3) \times 10^{-5}$	-16.2 [b]	11
T7A1	5.4×10^6 [(c)]	$2.7 (\pm 0.6) \times 10^{-3}$	-13.3 [c]	8
lacUV5	2.0×10^6 [(d)]	1.0×10^{-4}	-14.7 [e]	9-10
rrnB P1	3.9×10^6 [(f)]	≤1.4[g]	-8.2 ± 1[g]	3 ± 1[g]

Dissociation rate constants were determined here or taken from the noted references
[a]Binary open complex stability is $\Delta G^o_{R+P \rightarrow OC} = - RT \ln (k_a/k_d)$
[b]Ref. (14), [c]Ref. (15).
[d]Ref (16)
[e]ΔG^o_{oc} was calculated using 30 °C values of k_a and k_d (see reference (16)). Observed $N_{H,ESC}$ from ref (17-19)
[f]Ref (20)
[g]A minimum estimate of k_d for the unstable rrnB P1 OC (1.4 s^{-1}) is obtained from the geometric mean of the corresponding rate constants for dissociation of unstable λP_R and T7A1 OC (21). This predicts an OC stability of -9.2 kcal/mol. However, since the rrnB P1 OC is destabilized by prescrunching 2 bp vs 1 bp for T7A1 OC and none for λP_R OC, the lifetime and stability of the rrnB P1 OC may be reduced by up to 2 kcal/mol (see main text). This gives an estimated stability range of -8.2 ± 1 kcal/mol.

References

1. Gribskov M & Burgess RR (1983) Overexpression and purification of the sigma subunit of *E. coli* RNA polymerase. *Gene* 26:109-118.
2. Drennan A*, et al.* (2012) Key roles of the downstream mobile jaw of *E. coli* RNA polymerase in transcription initiation. *Biochemistry* 51:9447-9459.
3. Erie DA, Yager TD, & von Hippel PH (1992) The single-nucleotide addtion cycle in transcription: a biophysical and biochemical perspective. *Annual Reviews in Biophysics and Biomolecular Structure* 21:379-415.
4. Vo NV, Hsu LM, Kane CM, & Chamberlin MJ (2003) *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. *Biochemistry* 42:3798-3811.
5. Tinland B, Pluen A, Sturm J, & Weill G (1997) Persistence length of single stranded DNA. *Macromolecules* 30:5763-5765.
6. Murphy MC, Rasnik I, Cheng W, Lohman TM, & Ha T (2004) Probing single-stranded DNA conformational flexibiility using fluorescence spectroscopy. *Biophysical Journal* 86:2530-2537.
7. Baumann CG, Smith SB, Bloomfield VA, & Bustamante C (1997) Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences* 94:6185-6190.
8. Zuo Y & Steitz TA (2015) Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Molecular Cell* 58:534-540.
9. Fedoroff OY, Ge Y, & Reid BR (1997) Solution structure of r(gaggacug):d(CAGTCCTC) hybrid: Implications for the initiation of HIV-1 (+)-strand synthesis. *Journal of Molecular Biology* 269:225-239.
10. Lane AN, Ebel S, & Brown T (1993) NMR assignments and solution conformation of the DNA RNA hybrid duplex d(GTGAACTT) r(AAGUUCAC). *European Journal of Biochemistry* 215:297-306.
11. Liphardt J, Dumont S, Smith SB, Tinoco IJ, & Bustamante C (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzinski's equality. *Science* 296:1832-1835.
12. Huguet JM*, et al.* (2010) Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences* 107:15431-15436.
13. Bosco A, Camunas-Soler J, & Ritort F (2013) Elastic properties and secondary structure formation fo single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Research* 42:2064-2074.
14. Saecker RM*, et al.* (2002) Kinetic studies and structural models of the association of *E. coli* σ^{70} RNA polymerase with the λP_R promoter: large scale conformational changes in forming the kinetically significant intermediates. *Journal of Molecular Biology* 319:649-671.

15. Strauss HS, Burgess RR, & Record MTJ (1980) Binding of escherichia coli ribonucleic acid polymerase holoenzyme to a bacteriophage T7 promoter-containing fragment: selectivity exists over a wide range of solution conditions. *Biochemistry* 19:3496-3502.
16. Ross W & Gourse RL (2005) Sequence-independent upstream DNA-aCTD interactions strongly stimulate *E. coli* RNA polymerase-*lac*UV5 promoter association. *Proceedings of the National Academy of Sciences* 102:291-296.
17. Straney DC & Crothers DM (1987) A stressed intermediate in the formation of stably initiated RNA chains at the *E. coli lac* UV5 promoter. *Journal of Molecular Biology* 193:267-278.
18. Gralla JD, Carpousis AJ, & Stefano JE (1980) Productive and abortive initiation of transcription in vitro at the *lac* UV5 promoter. *Biochemistry* 19:5864-5869.
19. Carpousis AJ & Gralla JD (1980) Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter. *Biochemistry* 19:3245-3253.
20. Haugen SP*, et al.* (2006) rRNA promoter regulation by nonoptimal binding of σ^{70} region 1.2: an additional recognition element for RNA polymerase. *Cell* 125:1069-1082.
21. Ruff EF*, et al.* (2015) *E. coli* polymerase determinants of open complex lifetime and structure. *Journal of Molecular Biology* 427:2435-2450.