BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Sequence-Dependent Promoter Escape Efficiency Is Strongly Influenced by Bias for the Pretranslocated State during Initial Transcription

Jørgen Skancke,*[,†] Nadav Bar,[†] Martin Kuiper,[‡] and Lilian M. Hsu[§]

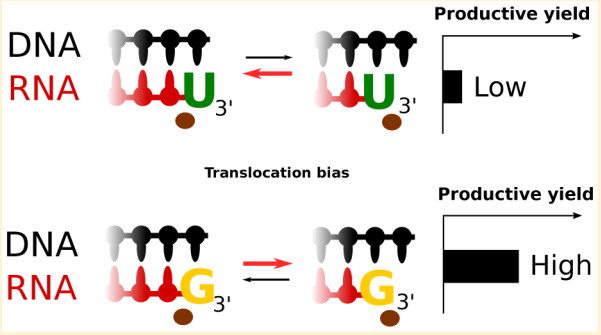
[†]Department of Chemical Engineering, Norwegian University of Science and Technology, Sem Sælandsvei 4, 7491 Trondheim, Norway

[‡]Department of Biology, Norwegian University of Science and Technology, Høgskoleringen 5, 7491 Trondheim, Norway

[§]Program in Biochemistry, Mount Holyoke College, South Hadley, Massachusetts 01075, United States

Supporting Information

ABSTRACT: Abortive transcription initiation can be rate-limiting for promoter escape and therefore represents a barrier to productive gene expression. The mechanism for abortive initiation is unknown, but the amount of abortive transcript is known to vary with the composition of the initial transcribed sequence (ITS). Here, we used a thermodynamic model of translocation combined with experimental validation to investigate the relationship between ITS and promoter escape on a set of phage T5 N25 promoters. We found a strong, negative correlation between RNAP's propensity to occupy the pretranslocated state during initial transcription and the efficiency of promoter escape ($r = -0.67$; $p < 10^{-6}$). This correlation was almost entirely caused by free energy changes due to variation in the RNA 3′ dinucleotide sequence at each step, implying that this sequence element controls the disposition of initial transcribing complexes. We tested our model experimentally by constructing a set of novel N25-ITS promoter variants; quantitative transcription analysis again showed a strong correlation ($r = -0.81$; $p < 10^{-6}$). Our results support a model in which sequence-directed bias for the pretranslocated state during scrunching results in increased backtracking, which limits the efficiency of promoter escape. This provides an answer to the long-standing issue of how sequence composition of the ITS affects promoter escape efficiency.



To begin transcription in bacteria, the RNAP-σ holoenzyme first binds to a promoter and melts DNA between the −10 element and the transcription start site to form an open complex. As RNA synthesis begins from the open complex, RNAP-σ is still bound to the promoter, which presents a challenge for RNAP as it needs to translocate its active site forward by one step on the template DNA to incorporate the incoming NTP. This challenge is overcome by the mechanism of scrunching, in which RNAP translocates by pulling in template DNA after each nucleotide incorporation step, compacting an enlarged DNA bubble and forming a stressed intermediate.[1,2] Translocation during initial transcription via scrunching is thus different from translocation during processive elongation by base stepping, where the DNA bubble maintains a more or less constant size.[3,4]

An initiation attempt can end when the stressed RNAP abortively releases the nascent RNA to fall back to the open complex conformation, which is known as abortive initiation.[5−7] Alternatively, the initiation attempt can result in promoter escape, which can occur if σ loses attachment to the upstream DNA contacts. As illustrated in the model of transcription initiation in Figure 1, the starting point for abortive initiation is a backtracking event from the scrunched complex.[7] During elongation, backtracking is suggested to occur from a prerequisite nonbacktracked "elemental paused state".[8,9] While less is known about the backtracking step during initial transcription, single-molecule experiments indicate that RNAP does not spend a prolonged amount of time in paused states during abortive cycling on the N25 promoter.[1] As RNAP backtracks during initial transcription, the DNA bubble will "unscrunch", and the RNA 3′ end moves into the NTP entry channel, until abortive release occurs (Figure 1). How the nascent RNA is released from the backtracked conformation is not known, but this presumably happens rapidly when the RNA−DNA hybrid has become shortened to ≤ 4−5 bp, because backtracked complexes with RNAs of these lengths cannot be recovered by GreB.[10,11] When the nascent RNA is released, RNAP is free to fall back to the open complex conformation to begin RNA synthesis *de novo*, thus completing one round of abortive cycling.

Received: March 12, 2015
Revised: June 15, 2015
Published: June 17, 2015

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

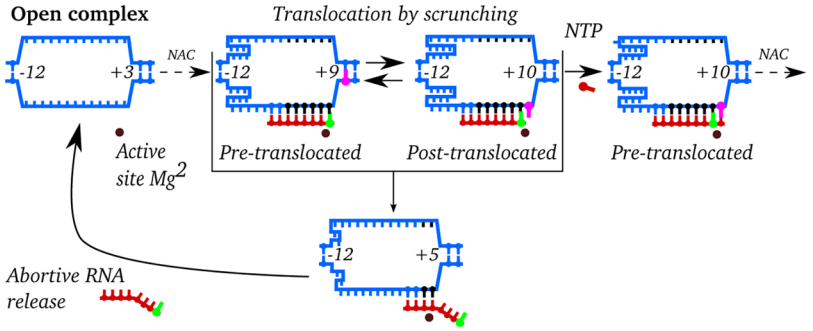


Figure 1. Abortive cycling that includes scrunching and backtracking. A scrunched complex having just incorporated the eighth nucleotide (green) is shown in the pretranslocated state. This complex translocates forward by scrunching 1 bp of DNA (pink), thereby moving the template DNA relative to the active site (indicated by a brown dot), now adopting the post-translocated state. Incorporation of the next NTP may occur from the post-translocated state, thus completing a nucleotide addition cycle (NAC). Like for elongation, translocation during initiation is assumed to be reversible;[14] the scrunched complex may translocate backward by releasing the 1 bp of DNA (pink) and return to the pretranslocated state. Scrunched complexes may also undergo backtracking of more than 1 bp; here, backtracking is shown to result in a partially "unscrunched" complex with a 4 bp RNA−DNA hybrid, with the 3′ end of the RNA extruded into the secondary channel. From the backtracked complex, RNA may be abortively released, returning RNAP to the open complex conformation to restart RNA synthesis, thereby completing one round of abortive cycling. In this diagram, DNA is colored light blue, DNA that has been pulled in by scrunching is colored black, and RNA is colored red.

It has been shown that the amount of abortive transcripts and productive RNA formed at a given promoter depends on the sequence of both the core promoter region and the initial transcribed sequence (ITS).[10,12,13] In an investigation using *Escherichia coli* RNAP Eσ^{70} and 43 randomized ITS variants of T5 phage N25 promoter, Hsu et al.[10] found that with different ITS sequences, promoter escape efficiency indicated by the productive yield (PY) varied more than 20-fold. The variation in PY between the different ITSs remained similar over a wide range of NTP concentrations.[10]

How the sequence in the ITS affects the promoter escape efficiency is not known. Hsu et al.[10] found no association between the GC content of the ITS and the PY or any effect of ≥100 μM NTP. Instead, a positive association was found between the number of nontemplate strand purines in the ITS and the PY. How the number of purines affects the PY is unclear. It is known, however, that the rate-limiting steps during abortive cycling are backtracking and abortive RNA release.[15] This makes it likely that the ITS affects promoter escape efficiency at either the backtracking, the abortive RNA release step, or both.

A way to investigate the relationship between DNA sequence and promoter escape efficiency is by thermodynamic or kinetic modeling of initial transcription. Previously, Xue et al.[14] published a kinetic model of initial transcription, which includes scrunching, backtracking, and promoter escape reactions. This model could partially reproduce some of the sequence-dependent features of transcription initiation from three phage promoters[12] but was not validated experimentally and was unable to provide additional insight into the mechanisms of promoter escape or abortive initiation. One challenge of using a detailed kinetic model to describe transcription initiation is the lack of available rate constants and knowledge about key reactions in the process, such as backtracking or promoter escape itself.

The model by Xue et al. is similar to previous kinetic and thermodynamic models of transcription elongation.[16−20] In these models, translocation is described as a reversible reaction that reaches equilibrium before NTP binding. The equilibrium constant for translocation has generally been calculated from the free energy change of the DNA bubble and the RNA−DNA hybrid between the pre- and post-translocated states.[21−23] Recently, however, a third sequence-dependent factor that influences the balance between the translocation states has been identified. Hein et al.[24] found that the translocation equilibrium is strongly influenced by the sequence of the RNA 3′ dinucleotide. In particular, the RNA 3′ nucleotide tends to bias the equilibrium toward the pretranslocated state in the following order: U > C > A > G. This order of sequence-directed bias was confirmed by Malinen et al.,[25] who showed that the identity of the 3′ base is the major factor that determines the balance between the pre- and post-translocated states. Further corroborating evidence of this relationship is the consensus pause sequence that is thought to disfavor the post-translocated state,[26,27] $G_{-10}Y_{-1}G_{+1}$, where −1 corresponds to the position of the RNA 3′ end. The reason for the strong effect of the RNA 3′ end on translocation has been suggested to stem from differential interactions between the RNA 3′ base and the folded trigger loop at the RNAP active site.[25]

In this work, we have formulated an improved computational model of translocation by incorporating the effect of the RNA 3′ dinucleotide. Applying this model to initial transcription, we found that the PY of an ITS variant could be explained by the disposition toward the pretranslocated state. We propose a mechanistic explanation whereby ITSs that bias RNAP to the pretranslocated state during initial transcription increase the chance of backtracking and thereby the amount of abortive RNA produced, thus reducing the promoter escape efficiency.

MATERIALS AND METHODS

Rationale and Approach. Our goal in this work is to provide a quantitative underpinning for understanding the sequence-dependent nature of promoter escape efficiency. Because promoter escape depends on RNAP conducting a succession of forward translocation steps, we hypothesized that variations in translocation equilibrium at different steps during initial transcription could be the reason for the strong sequence dependence of the promoter escape efficiency for the N25-ITS variants.[10] To test this hypothesis, we decided to investigate transcription of the ITS through thermodynamic modeling using an equilibrium description of translocation.

Model Description. Previous thermodynamic and kinetic models of transcription elongation have generally assumed that translocation occurs in equilibrium,[19−21] although recent work

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

has found better agreement for a nonequilibrium model.[28] Even if an equilibrium description is valid for elongation, this does not necessarily imply the same for initiation, which differs from the former primarily by the strain from the increasingly large DNA bubble.[1,2] On the other hand, the equilibrium assumption is consistent with the observation that, during abortive initiation, RNAP active center forward translocation is a fast step.[15] Ultimately, however, for the purpose of modeling RNAP movement, it is necessary only that the equilibrium assumption be approximately correct.

Equilibrium Constant of Translocation. To describe translocation via scrunching, we derived the equilibrium constant from $\Delta(\Delta G)$ between the pre- and post-translocated states; thus

$$\Delta(\Delta G) = \Delta G_{\mathrm{Pre}} - \Delta G_{\mathrm{Post}} \quad (1)$$

$\Delta(\Delta G)$ obtained this way gives rise to an equilibrium constant for backward translocation [i.e., post- to pretranslocated (see the explanation below)]. Each free energy variable in eq 1 consists of three terms: ΔG for the scrunched DNA bubble ($\Delta G_{\mathrm{DNA-DNA}}$), the RNA−DNA hybrid ($\Delta G_{\mathrm{RNA-DNA}}$), and the RNA 3′ dinucleotide sequence ($\Delta G_{3\mathrm{N}}$).

We calculate $\Delta(\Delta G)$ according to how a translocation step occurs during initial transcription. (i) the DNA bubble does not close upstream but opens one nucleotide ahead of the active site,[29] and (ii) the RNA−DNA hybrid length remains constant between the pre- and post-translocated states before a full hybrid length of 9 bp is attained,[14,29] resulting in $\Delta(\Delta G_{\mathrm{RNA-DNA}}) = 0$ for transcription up until this point. We used nearest neighbor estimations to calculate $\Delta G_{\mathrm{DNA-DNA}}$ (Table 1 of ref 30) and $\Delta G_{\mathrm{RNA-DNA}}$ (temperature-dependent energies in Table 2 of ref 31).

A novel component in our model is the inclusion of the recently discovered effect of the RNA 3′ dinucleotide sequence on translocation equilibrium. Hein et al.[24] showed that equilibrium constants of pyrophosphorolysis varied depending on the RNA 3′ dinucleotide and attributed this to variation in translocation bias between the pre- and post-translocated states. We quantify the coupling to translocation by assuming a linear relationship between the free energy change of pyrophosphorolysis and the free energy change of backward translocation in these experiments. Thus, we set the free energy change for the RNA 3′ dinucleotide sequence $[\Delta(\Delta G_{3\mathrm{N}})]$ equal to the free energy change of pyrophosphorolysis (see Table 1) multiplied by a constant, or weight factor, which will be estimated from data; thus, $\Delta G_{3\mathrm{N}} = c\Delta G_{\mathrm{pp_i}}$. We also associate $\Delta G_{\mathrm{RNA-DNA}}$ and $\Delta G_{\mathrm{DNA-DNA}}$ with weight factors that will be estimated from data. This is because these free energy variables have been obtained under experimental conditions other than transcription, so it cannot be expected that the values of these variables correspond exactly to the magnitudes of their energetic contribution to translocation.

Although the equilibrium constants of pyrophosphorolysis from Hein et al. were obtained with *Thermus thermophilus* RNAP,[24] Malinen et al.[25] found the same effect order of the RNA 3′ nucleotide on the stability of the post-translocated state: G > A > C > U for *E. coli* RNAP, which suggests that the effect order of the 3′ nucleotide is conserved from *T. thermophilus* to *E. coli*. For dinucleotides, Hein et al.[24] found that the effect of 3′ GU and UG on pyrophosphorolysis was conserved from *T. thermophilus* to *E. coli*, yeast, and mammalian RNAPs, even though the value of equilibrium constants for the GU/UG pair was not conserved.[24] This implies that while the exact equilibrium constant value is not conserved, the difference in order for the different dinucleotides is. For our model, it is sufficient that only the difference in effect order of equilibrium constants from *T. thermophilus* to *E. coli* is conserved.

Table 1. Changes in Free Energy for Pyrophosphorolysis[a]

RNA 3′ dinucleotide	$\Delta G_{\mathrm{PP_i}}$ (kcal/mol)
TG	1.85
TA	1.48
GC	0.99
AG	0.74
GG	0.74
CG	0.74
CA	0.56
TC	0.56
GA	0.43
CC	0.43
AA	0.31
AC	0.22
GT	−0.46
AT	−0.59
TT	−0.63
CT	−0.66

[a]The values are converted from the K_{eq} of pyrophosphorolysis from Table 1, column 5, of ref 24 to the change of free energy using the equation $\Delta G = -RT\ln(K_{\mathrm{eq}})$. The RNA 3′ dinucleotides are listed as their corresponding 5′ → 3′ nontemplate DNA sequence. The dinucleotides are arranged from top to bottom in increasing order of propensity for the pretranslocated state. $\Delta G_{3\mathrm{N}}$ is estimated from these values by the equation $\Delta G_{3\mathrm{N}} = c\Delta G_{\mathrm{PP_i}}$.

Having thus discussed the free energy terms, we formulate the equilibrium constant for the backward translocation of a scrunched complex at position i, $K_{\mathrm{bt},i}$, as follows:

$$K_{\mathrm{bt},i} = \mathrm{e}^{-[c_1\Delta(\Delta G_{\mathrm{RNA-DNA},i})+c_2\Delta(\Delta G_{\mathrm{DNA-DNA},i})+c_3\Delta(\Delta G_{3\mathrm{N},i})]/RT} \quad (2)$$

The value of K_{bt} at a given position (e.g., 8) is interpreted as follows: if $K_{\mathrm{bt},8}$ of ITS variant A is larger than $K_{\mathrm{bt},8}$ of ITS variant B, then the initial complex at position 8 for variant A has a stronger propensity for the pretranslocated state than for variant B.

Average Equilibrium Constant of Backward Translocation as a Measure of an RNAP's Propensity To Reside in the Pretranslocated State during Initial Transcription. We compare the different ITS variants by RNAP's average K_{bt} up to promoter escape. This measure allows an ordering of different ITS variants by their overall degree of preference for the pretranslocated state during initial transcription. We assume for each variant that translocation occurs up to the synthesis of the maximal size of abortive transcript (MSAT) as found by Hsu et al.,[10] which corresponds to MSAT-1 translocation steps:

$$\bar{K}_{\mathrm{bt,MSAT}} = \frac{\sum_{i=2}^{\mathrm{MSAT}} K_{\mathrm{bt},i}}{\mathrm{MSAT\text{-}1}} \quad (3)$$

More generally, we refer to the average K_{bt} up to the synthesis of an RNA of length n:

$$\bar{K}_{\mathrm{bt},n} = \frac{\sum_{i=2}^{n} K_{\mathrm{bt},i}}{n-1} \quad (4)$$

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

Model Implementation. All calculations were performed with custom Python scripts using the numerical and statistical libraries Numpy and Scipy. All correlation coefficients and p values were obtained using the Spearman correlation. The source code is available upon request.

Parameter Estimation. The weight factors c_1, c_2, and c_3 in eq 2 are estimated by optimizing for values that achieve maximal correlation between PY and $\bar{K}_{bt,MSAT}$. The values of the weight factors are varied independently from 0 to 1 in steps of 0.05, where 0 can be interpreted as no contribution to translocation and 1 indicates the full contribution to translocation by the free energy variable. For each combination of weight factor values, we calculate $\bar{K}_{bt,MSAT}$ and correlate the resulting values with the variants' PYs. The optimal values that give highest correlation are then used (see Table 2).

Table 2. Weight Factors (mean and standard deviation) Used in eq 2 That Give the Best Correlation between PY and $\bar{K}_{bt,n}$[a]

	c_1, $\Delta G_{RNA-DNA}$	c_2, $\Delta G_{DNA-DNA}$	c_3, ΔG_{3N}
Figures 2, 3C (blue), 4, 5	0.26	0.00	0.58
Figure 3A (blue)	–	–	1.00
Figure 3A (green)	*	–	–
Figure 3A (purple)	–	*	–
Figure 3B (blue)	0.25	–	0.55
Figure 3B (green)	–	0.00	0.75
Figure 3B (purple)	*	*	–
Figure 3C (green)	0.26 ± 0.15	0.04 ± 0.06	0.68 ± 0.22
Figure 3C (purple)	*	*	*

[a]Values shown were used to calculate $K_{bt,i}$ in the indicated figures. An asterisk indicates that no significant correlation was obtained for the given combination of free energy variable(s), and a dash indicates that the free energy variable was excluded for parameter estimation and $K_{bt,i}$ calculation.

Construction of the DG400 Promoter Series. To test the validity of the correlation reported here, we constructed 26 N25 promoters containing novel ITS variations from position +3 to +15, designated as DG427–DG452, that are predicted to show a wide range of $\bar{K}_{bt,15}$. We also included four control promoters (N25, DG115a, DG133, and N25/A1anti) selected from the original set,[10] now with ITS variation from position +3 to +15, to measure their PY concurrently. All 30 promoters analyzed here are 152 bp long spanning residues −85 to +67; their sequences are identical to the N25 sequence (see Figure 1 of ref 10) except for ITS positions +3 to +15. The variant promoters were synthesized via polymerase chain reaction (PCR) amplification of six overlapping oligonucleotide primers A–F: A-primer (NT), residues −85 to −58; B-primer (T), residues −72 to −28; C-primer (NT), residues −49 to +2; D-primer (T), residues −20 to +34; E-primer (NT), residues +16 to +56; and F-primer (T), residues +38 to +67. NT and T denote nontemplate and template strands, respectively. Each reaction differs only in the identity of the D-primer (i.e., Dx), which encodes the desired ITS (+3 to +15) changes. A typical PCR mixture contains, per 100 μL, primers B, C, Dx, and E (0.05 μM each), primers A and F (0.5 μM each), 250 μM dNTP, in 1× HF (Phusion) buffer, and 5 units of Phusion high-fidelity DNA polymerase (NEB) and is amplified through 50 cycles of a thermal program of 15 s each at 95, 50, and 72 °C. Afterward, the promoter DNAs were purified as described previously[13] and redissolved in TE [10 mM Tris (pH 8) and 1 mM Na_2EDTA] buffer to a final concentration of 300 nM, for use in *in vitro* transcription reactions.

Steady-State *in Vitro* Transcription Reaction. Quantitative *in vitro* transcription analysis was performed as described previously[32] with all 30 promoters prepared above. The PY values are calculated as the fraction of full length transcripts relative to the sum of full length and abortive transcripts (Figure S1 of the Supporting Information) and are reported as averages of three independent experiments (Table 3). All

Table 3. DG400 Series Library Designed Using eq 2 and Controls[a]

name	sequence	$\bar{K}_{bt,15}$	PY (%)	purine count
N25	ATAAATTTGA GAGAG	13.6	29	11
DG450	ATTAGAGAAG AAGAA	12.7	26	13
DG452	ATTAGAAAAG AGAAT	13.9	24	12
DG447	ATACAGAGAC AGAGA	11.7	23	12
DG445	ATTAAAGAAC CAGGT	13.8	22	10
DG451	ATACAAGAGA AGAAT	13.7	21	12
DG448	ATTAGAAGAA ACAGG	12.3	21	12
DG444	ATTACAAAGG AGCAA	11.8	20	11
DG442	ATTCAAATAC ACCAA	13.8	16	8
DG443	ATAAGAGGAA GAGCG	10.6	15	13
DG446	ATAGAGGAAG AGGAG	11.3	15	14
DG449	ATAGACAGAA AGGAT	13.5	13	12
DG439	ATAGGAACCG ATAGG	12.7	12	11
DG440	ATAACAAGAC ATCCT	14.4	11	8
DG428	ATCTATGGTG TGTGC	13.7	10	7
DG133	ATATCGAATT ACTCA	16.9	10	7
DG435	ATTACAGCAG CTACT	15.1	9	7
DG438	ATTCATAAAC ATTGG	16.1	8	8
DG115a	ATCCCGCTCA AGAGC	13.0	8	7
DG441	ATAGAAGCAG ATTCC	15.2	7	9
DG436	ATAGAACGTA CTGAT	15.9	7	9
DG430	ATGTGATGAC TTCCT	19.3	6	6
DG437	ATCCAGAAGC CTAAT	15.9	6	8
DG432	ATACACTTCC ATATT	20.0	5	5
DG431	ATGATTGTCC GTCAC	15.8	5	6
DG429	ATCTTCTCTA TTGCA	18.0	5	4
DG427	ATCTCTCTTC TGTTA	19.8	4	3
N25/A1anti	ATATCTCTTT CACAT	19.2	3	4
DG433	ATGTTGCCTT TGAGC	16.8	2	6
DG434	ATCGACTTGC TTTAC	19.3	1	5

[a]The sequences shown are from +1 to +15. For each variant are shown the $\bar{K}_{bt,15}$, the experimentally obtained PY, and the purine count of the sequence. The table is sorted with respect to PY.

recently obtained PY values are elevated ∼5-fold, as compared to the reported values for the four control promoters.[10] We are not sure of the exact reasons for this discrepancy but suspect the different PCR schemes and enzymes used account for the differences. The elevated PY values, however, do not detract from the comparative analysis and correlation reported here.

RESULTS

The Promoter Escape Efficiency of N25-ITS Variants Is Strongly Anticorrelated with a Bias for the Pretranslocated State during Initial Transcription. To investigate the association between translocation bias during initial transcription and PY, we optimized the weight constants in eq 2 (given in Table 2) to calculate the correlation between PY and

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267–4275

$\bar{K}_{bt,MSAT}$ for all 43 N25-ITS variants in the DG100 library for initial transcription up to the promoter escape step (see Materials and Methods). The correlation we found was high (r = −0.67; $p < 10^{-6}$); as can be seen in Figure 2A, there is a marked tendency that ITS variants with a lower average $\bar{K}_{bt,MSAT}$ have a higher PY and, conversely, variants with a higher average $\bar{K}_{bt,MSAT}$ have a lower PY. This shows a clear negative association between sequence-dependent bias for the pretranslocated state during initial transcription and promoter escape efficiency.

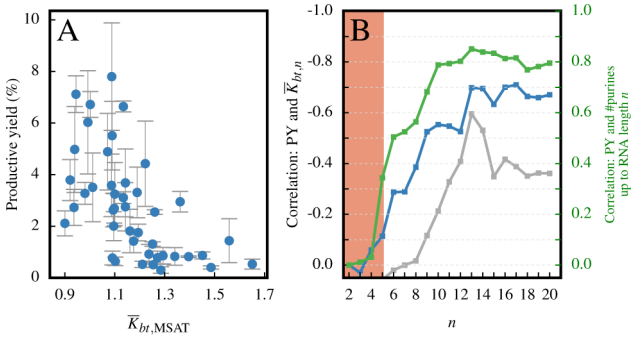


Figure 2. (A) PY plotted vs $\bar{K}_{bt,MSAT}$ using optimal weight constants. PY is anticorrelated with RNAP's sequence-directed bias for the pretranslocated state during initial transcription. (B) Blue: correlation coefficients for PY and $\bar{K}_{bt,n}$ for values of n given on the x-axis, using optimal weight constants. Gray: correlation coefficients obtained with all weight constants set equal to 1. Green: correlation coefficients (values on the right-hand vertical axis) between PY and the number of purines in nontemplate strand DNA/RNA up to the synthesis of a length n RNA. The red area indicates values of n up to the first significant correlation ($p > 0.05$) between $\bar{K}_{bt,n}$ and PY at +6 when using optimal weight constants. The left vertical axis is inverted to show increasing negative correlation.

The Correlation between Promoter Escape Efficiency and Translocation Bias Increases with the Number of Translocation Steps. Figure 2A shows that there exists an association between PY and backward translocation equilibria for initial transcription up to promoter escape. However, it does not give information about which translocation steps contribute to this correlation. Therefore, to investigate the contribution of each additional translocation step, we used the optimal parameters for PY and $\bar{K}_{bt,MSAT}$ to correlate PY with average $\bar{K}_{bt,n}$ for successively longer subsequences of the ITS up to the MSAT for each variant, where the maximal MSAT is 20 for some variants.[10] We found that correlation between PY and $\bar{K}_{bt,n}$ starts low before +5 and after that increases cumulatively with each translocation step, reaching significance at +6 and a maximum at +17 (r = −0.71; $p < 10^{-7}$) (Figure 2B, blue curve). As shown in Figure 2B, most of the correlation is explained by the subsequence of the ITS up to +13 (r = −0.70; $p < 10^{-6}$).

The RNA 3′ Dinucleotide Is the Main Contributing Factor for the Correlation between PY and Translocation bias. The weight factors that were used to calculate the correlation between $\bar{K}_{bt,MSAT}$ and PY (Figure 2A) can be interpreted to represent the contribution of each variable in eq 2 to the correlation. The weight factor for the RNA 3′ dinucleotide (c_3) was resolved to 0.58, while the weights for the DNA bubble (c_2) and the RNA–DNA hybrid (c_1) were 0 and 0.26, respectively (Table 2, first row). This shows that the free energy of the RNA 3′ dinucleotide sequence is the dominating term in eq 2 for the correlation between translocation bias and PY.

To further investigate the role of the ΔG_{3N} variable, we repeated the correlation analysis using an equilibrium constant containing only ΔG_{3N}: $K_{bt,i} = e^{-c_3\Delta(\Delta G_{3N,i})/RT}$. We found that this variable on its own led to a correlation stronger than −0.6 (Figure 3A). On the other hand, when using only $\Delta G_{RNA-DNA}$ or $\Delta G_{DNA-DNA}$ to calculate the equilibrium constant, no significant correlation between PY and $\bar{K}_{bt,n}$ was found (Figure 3A). This result is in line with the values of the optimal weight constants in Table 2. We then evaluated $\bar{K}_{bt,n}$ using different combinations of ΔG_{3N}, $\Delta G_{RNA-DNA}$, and $\Delta G_{DNA-DNA}$; as shown in Figure 3B, only $\bar{K}_{bt,n}$ values that included ΔG_{3N} contributed to significant correlation with PY. This suggests that the

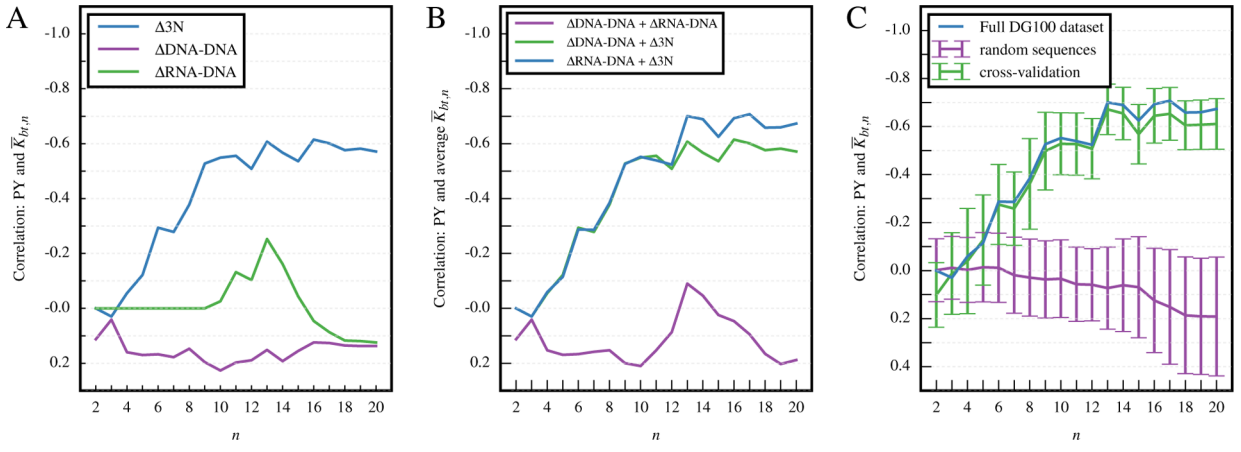


Figure 3. (A) Delineation of the contribution of free energy variables to the correlation with promoter escape efficiency. The blue line shows the correlation between $\bar{K}_{bt,n}$ and PY when only $\Delta(\Delta G_{3N})$ is used in eq 2 to calculate the equilibrium constants (in effect, setting c_1 and c_2 to 0). The correlations obtained when c_2 and c_3 are set to 0 [using only $\Delta(\Delta G_{RNA-DNA})$] and c_1 and c_3 are set to 0 [using only $\Delta(\Delta G_{DNA-DNA})$] are colored green and purple, respectively. (B) Optimal correlation between $\bar{K}_{bt,n}$ and PY for combinations of the energy terms in eq 2. (C) Robustness of the correlation between PY and $\bar{K}_{bt,n}$. The blue line shows the correlation for the full set of 43 N25-ITS variants. The green line shows the correlation obtained from the cross-validation approach involving random sampling of 100 subsets of 22 ITSs out of the original 43. The purple line shows the same for 100 runs of 43 randomly generated ITSs. Bars indicate the standard deviation of correlation for all the samples. The left axes are inverted to show the increasing negative correlation.

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267–4275

correlation between the promoter escape efficiency and average equilibrium constant of backward translocation is due primarily to variation in ΔG_{3N} among the ITS variants.

Having established the role of ΔG_{3N} for the correlation with optimal weight constants, we compared it to the special case in which all weight constants are equal to one ($c_1 = c_2 = c_3 = 1$), which corresponds to calculating $\overline{K}_{bt,n}$ using

$$K_{bt,i} = e^{-[\Delta(\Delta G_{RNA-DNA,i})+\Delta(\Delta G_{DNA-DNA,i})+\Delta(\Delta G_{3N,i})]/RT} \quad (5)$$

This comparison is of interest because eq 5 is the form that the equilibrium constant has taken in previous models,[14,18] except for the term $\Delta(\Delta G_{3N,i})$ introduced in this study. This showed a lower correlation between PY and $\overline{K}_{bt,MSAT}$ ($r = -0.36$; $p < 0.02$) compared to the optimal case, with a peak of correlation at +13 between PY and $\overline{K}_{bt,13}$ ($r = -0.60$; $p < 10^{-4}$) (Figure 2B, gray line).

To compare with the results of Hsu et al., we calculated how the correlation between the number of purines in the ITS correlates with PY for increasingly longer subsequences of the ITS. As shown in Figure 2B (green line), the number of purines in the ITS correlates more strongly with PY than with $\overline{K}_{bt,n}$ and the correlation also increases cumulatively with the length of the DNA sequence (see also Figure S2 of the Supporting Information).

The Correlation between PY and $\overline{K}_{bt,n}$ Is Robust with Respect to Random Sampling and Cross Validation. Because the correlation between average translocation equilibria and PY was obtained through weight factor estimation, we wanted to rule out the possibility that the correlation between PY and $\overline{K}_{bt,n}$ was due to overfitting in the weight factor estimation procedure itself. Two signs that a model is overfitting data are that it produces significant results even with noisy data and that model output is not robust with respect to cross validation.

To test the response of the model to noisy data, we replaced the sequence of each variant in the DG100 library with random DNA and obtained an optimal correlation between $\overline{K}_{bt,n}$ (now from random DNA) and PY (from DG100). We repeated this process 100 times and averaged the resulting correlations (Figure 3C). The average correlation coefficients for the random sets were close to 0 for each subsequence up to +15 but showed a moderate bias toward positive correlation between +16 and +20 (Figure 3C, purple curve). This shows that the model is not expected to produce statistically significant results with random data, indicating that the model is not overfitting the data.

We then used a cross-validation approach in which we randomly selected half (22) of the 43 ITS variants in the DG100 library, obtained optimal weight factors, and used these weight factors to calculate correlation between PY and $\overline{K}_{bt,n}$ for the remaining 21 variants in library. By repeating this process 100 times and taking the average of the resulting correlation coefficients, we found that the averages were highly similar to the values obtained for the full data set (Figure 3C, comparing green and blue curves). The robustness of the model is also evident in that the weight factors calculated from the cross-validation approach were similar to those obtained with the full data set (Table 2).

$\overline{K}_{bt,n}$ Predicts Promoter Escape Efficiency for Novel N25-ITS Variants. Most of the correlation between PY and $\overline{K}_{bt,n}$ is reached when averaging over the ITS sequence up to +13 (Figure 2B), and most ITSs in the DG100 library achieve promoter escape at the +15/+16 position.[10] Therefore, to identify novel ITSs that would span the total variation in PY, we reasoned that it was sufficient to vary the ITS up to +15. We evaluated *in silico* a library of all possible DNA 15-mers starting with AT to obtain 26 novel ITS variants in such a way as to uniformly span their $\overline{K}_{bt,15}$ values from highest to lowest. Next we constructed by PCR a corresponding set of N25-ITS variant promoter fragments (152 bp, positions −85 to +67), designated as the DG400 series library, and subjected them to quantitative *in vitro* transcription analysis.[13,32] We included four control promoters from the original 43 N25-ITS variants in the transcription analysis. The empirically measured PY for these 30 promoters (Table 3) again showed a strong, negative correlation with $\overline{K}_{bt,15}$ (for the DG400-series, $r = -0.81$ and $p < 10^{-6}$; for the DG400 series and the four controls, $r = -0.78$ and $p < 10^{-6}$), as shown in Figure 4A. To also see how correlation

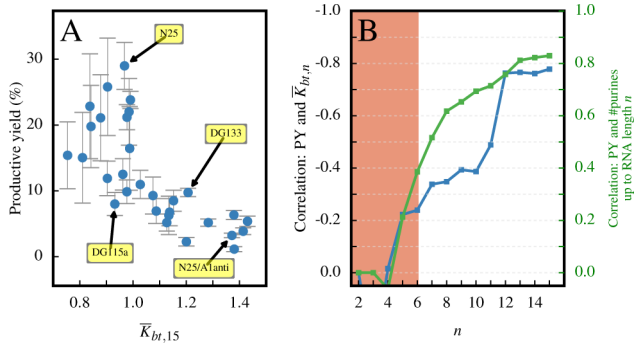


Figure 4. (A) PY and $\overline{K}_{bt,15}$ of the 26 novel N25-ITS variants and four controls. The control promoters are indicated with yellow boxes. (B) Blue: correlation coefficients for PY of the DG 400 series and $\overline{K}_{bt,n}$ for values of n given on the x-axis, using optimal weight constants. Green: correlation coefficients (shown on the right-hand vertical axis) between PY and the number of purines in DNA/RNA up to a length n RNA. The red area indicates values of n up to the first significant correlation ($p > 0.05$) between $\overline{K}_{bt,n}$ and PY at +6 when using optimal weight constants (see Table 2, line 1). The left vertical axis is inverted to show the increasing negative correlation.

changes with sequence length, we evaluated the correlation between PY and $\overline{K}_{bt,n}$ for different values of n. As for the DG100 library, the correlation increases cumulatively with sequence length, reaching a maximum at +12 (Figure 4B). For comparison, we also calculated the correlation with the number of purines in the sequence up to $+n$. This showed the same trend as for the DG100 library, but here the correlation with the number of purines was near identical to the correlation with $\overline{K}_{bt,15}$ after +12.

We have hereby shown that by modulating the equilibrium constant of translocation for the first 15 bases on the N25 promoter that it is possible to construct ITS sequences with a wide range of promoter escape efficiencies. This shows that the model presented here, the $\overline{K}_{bt,15}$ of the ITS, is predictive for the PY of novel ITS variants.

Sequence-Directed Bias for Translocation Is Associated with the Probability of Aborting Initial Transcription. Having established the negative correlation between $\overline{K}_{bt,n}$ and PY values (i.e., high $\overline{K}_{bt,n}$, low PY, and vice versa), we were interested in exploring the causality of this link. There are several possible explanations for the $\overline{K}_{bt,n}$−PY correlation for the N25-ITS variants (see Discussion). For the link with abortive transcript production, we reasoned that initiation

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

complexes that spend a prolonged amount of time in the pretranslocated state are more likely to backtrack and, from the backtracked state, abort the initiation attempt. If this hypothesis is true, one would expect an increased amount of abortive product from positions where the pretranslocated state is favored. As a consequence, one would also expect that ITS variants with a high average preference for the pretranslocated state would have a comparatively large probability of aborting initiation before promoter escape.

To test these hypotheses, we calculated the abortive probability (AP) at each ITS position for the DG100 and DG400 libraries (see ref 13 for the calculation of AP) and the correlation with the K_{bt} values at each position. We found a significant positive correlation ($p < 0.01$) between K_{bt} and AP at +3, +5, +7, and +8 (Figure 5B, bars with asterisk); no significant correlation between K_{bt} and AP was found for the other positions.

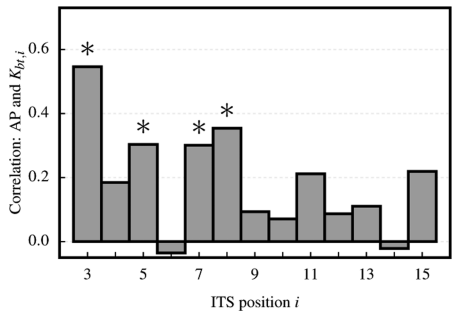


Figure 5. Correlation between K_{bt} and AP values for the DG100 and DG400 libraries at individual ITS positions. The height of the bar corresponds to the correlation coefficient between the AP values and the K_{bt} values at the indicated position. Columns with an asterisk correspond to correlation coefficients that remained significant after adjusting for multiple testing using a false discovery rate of $p = 0.05$ according to the Benjamini–Hochberg procedure.

DISCUSSION

In this study, we have shed light on the long-standing issue of the relationship between the sequence in the +1 to +20 region of a promoter and the efficiency of promoter escape.[12,33] By applying an improved model of translocation, we showed that for N25-ITS promoters there is a negative correlation between the occupancy of the pretranslocated state during initial transcription and the efficiency of promoter escape (Figures 2A and 4A), mainly due to translocation bias directed by the RNA 3′ dinucleotide sequence (Table 2 and Figure 3). The model we formulated was used predictively to obtain a correlation with experimentally measured PY higher than that obtained in the training set (Figure 4A), which also shows that for the ITS it is sufficient to vary the composition of the first 15 positions to span a large variation in PYs. At several ITS positions, we found that the bias for the pretranslocated state was positively correlated with the probability to abort transcription (Figure 5).

The relationship between sequence-dependent translocation bias and promoter escape efficiency provides a mechanistic explanation for results of Hsu et al.,[10] who initially reported that promoter escape efficiency was positively correlated with the number of purines in the nontemplate strand of the ITS (or the number of purines in the RNA). As shown in Table 3, the $\overline{K}_{bt,15}$ is lower for the ITS variants with a high GA content and higher for variants with a high TC content. The strong role of the RNA 3′ dinucleotide in determining the translocation bias would therefore underlie this original observation. When we calculated the correlation between the number of purines and PY for different subsequences of the ITS, we found that the number of purines correlates even more strongly with PY than $\overline{K}_{bt,n}$ (Figures 2B and 4B). One explanation may be that the ΔG_{3N} term has been obtained for the specific RNA–DNA hybrid used by Hein et al. While corroborating evidence supports the overall validity of the effect of the nature of the 3′ dinucleotide on translocation,[25−27] it has been shown that sequence-dependent interactions between the entire RNA–DNA hybrid and RNAP affect translocation in ways unrelated to RNA–DNA thermodynamics.[34] This suggests that while sequence information at the 3′ end may explain most of the effect of the hybrid, this information is not sufficient to capture more subtle differences. Additionally, there may be processes beyond translocation during initial transcription that are affected by the number of purines or pyrimidines. One such effect may be that purines and pyrimidines in 3′ RNA interact differently with residues in the NTP entry channel during backtracking.[35]

The Propensity for the Pretranslocated State during Scrunching Can Be Linked to Low Promoter Escape Efficiency through Backtracking. Having established the correlation between the occupancy of the pretranslocated state and low promoter escape efficiency, we find the question becomes what the causal link for this correlation is. We believe the explanation is at least twofold. One reason is that initiation complexes that spend more time in the pretranslocated state are more likely to backtrack (assuming that backtracking occurs more readily from the pretranslocated state[36,37]), undergo abortive initiation, and thereby experience a reduced promoter escape efficiency. When we tested this hypothesis, we found that the correlation was positive between translocation bias and abortive probability at all but two ITS positions, although statistical significance was reached for only positions 3, 5, 7, and 8 (Figure 5). The marked lack of correlation at +6 compared to neighboring positions could be caused by the steric collision between the six-nucleotide RNA as it must displace σ region 3.2 during initial transcription.[38,39] The steric collision may itself lead to abortive RNA release, or it may represent a sequence-independent barrier for forward translocation, thus causing a strong bias for the pretranslocated state at this position. Interestingly, there is a significant correlation between AP and K_{bt} at +3 and +5, while the correlation between PY and $\overline{K}_{bt,n}$ does not reach significance before +6. This indicates that while translocation bias may contribute to AP before +6, it is only after this position that it contributes to a promoter's PY.

The other factor that likely contributes to the PY–$\overline{K}_{bt,15}$ correlation is more straightforward: ITSs that bias the RNAPs to dwell longer in the pretranslocated state can reduce the rate of reaching the promoter escape stage, thereby lowering their PY.

Does Scrunching Provide the Driving Force for Abortive Initiation and Promoter Escape? It must be considered that translocation during initial transcription (via scrunching) is a reaction energetically different from translocation during elongation (via base stepping). It has been suggested previously that the free energy of the scrunched DNA bubble is a contributing factor in abortive initiation (i.e., backtracking and abortive RNA release) and promoter escape.[1,7] In this study, we have accommodated the effect of

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

the scrunched DNA bubble by adding to K_{bt} the free energy of each base pair opened during scrunching [the $\Delta(\Delta G_{DNA-DNA})$ term in eq 2]. However, with this approach, we found no significant correlation between the difference in the free energy of the transcription bubble during each scrunching step and promoter escape efficiency (Table 2 and Figure 3).

Recently, the issue of scrunching, and scrunching-induced stress energy, as the driving force for abortive initiation and promoter escape, has been challenged.[40,41] By alleviating stress in scrunched DNA bubbles for both the T7 and *E. coli* RNAPs, Martin and co-workers attributed the abortive release of RNA with no more than eight nucleotides to "hybrid push".[40,41] However, on the Qλ-arrested complex where the "hybrid push" phenomenon does not exist, Strobel and Roberts ascribe the driving force for abortive initiation and escape by the QAC complex entirely to DNA scrunching.[42]

Our finding indicates that the free energy of the scrunched DNA bubble either does not play a role or is irrelevant in determining the sequence-dependent promoter escape efficiency. If the latter is true, it implies that the accumulated energy in the scrunched DNA bubble cannot be calculated from thermodynamic parameters of free duplex DNA. Of the three free energy terms considered ($\Delta G_{DNA-DNA}$, $\Delta G_{RNA-DNA}$, and ΔG_{3N}), the only term that we identified to correlate with productive yield was ΔG_{3N}. For transcription elongation, all three of these terms are associated with the consensus pause sequence through the pretranslocated state.[26,27] The lack of correlation associated with $\Delta G_{DNA-DNA}$ and $\Delta G_{RNA-DNA}$ in our study may suggest that interactions among RNAP, the RNA−DNA hybrid, and the scrunched DNA bubble can change the free energy landscape to such an extent that $\Delta(\Delta G_{DNA-DNA})$ and $\Delta(\Delta G_{RNA-DNA})$ do not adequately describe the energetic effects of these nucleic acid duplexes during initial transcription. This isolation of the effect of ΔG_{3N} was made possible in our study by estimating weights for the free energy terms. By extension, this weight estimation approach may be applied to identify the relative influence of these terms for studies of transcription elongation.

Another aspect in which scrunching may differ from translocation during elongation is that backtracking may occur not only from the pretranslocated state but also directly from the post-translocated state. Such a reaction can account for ITS positions that favor the post-translocated state (i.e., low K_{bt}) that nevertheless show a high AP.

Role of the RNA 3′ Dinucleotide for Scrunching. The scrunching mechanism was put forth as the answer to the paradox that during transcription initiation RNAP remains bound at the promoter but at the same time translocates its active site relative to template DNA.[1,2] While it has been shown that scrunching involves pulling DNA into RNAP, it is not clear how this pulling occurs. The results of Malinen et al. indicate that the RNA 3′ nucleotide has a sequence-specific affinity for the folded trigger loop present in the pretranslocated state, but a sequence-nonspecific affinity for the post-translocated state when the trigger loop unfolds.[25] Thus, if the affinity of the RNA 3′ end for the nucleophilic site is strong enough to pull in template DNA as part of the RNA−DNA hybrid and unwind the DNA duplex one base pair downstream, but not strong enough to break the RNAP−promoter bonds upstream, these forces, in combination, would lead to a scrunching of the DNA bubble.

Conclusion. In conclusion, we propose that promoter escape efficiency from the N25 promoter is largely determined by the translocation bias during transcription initiation, which, in turn, is determined by the RNA 3′ dinucleotide. RNAPs that are comparably more biased toward the pretranslocated state by the ITS during scrunching have a greater probability of becoming backtracked, spend more time in abortive cycling, and therefore have a lower efficiency of promoter escape.

In this work, we provide a model that explains most of the variation previously observed between the number of purines in the ITS and the efficiency of promoter escape. While this study was performed with the N25 promoters, we expect these results to be applicable to all intrinsically strong promoters that are transcribed by RNAP to undergo backtracking-induced abortive initiation and limited promoter escape.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Transcription gel for the DG400 library (Figure S1) and correlation between the number of purines from +1 to +15 and PY (Figure S2). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00272.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jorgsk@nt.ntnu.no.

Funding

The work of J.S., N.B., and M.K. was supported by NTNU. L.M.H.'s work on behalf of this collaborative project was supported by an NSF-RUI grant (MCB0841452).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are thankful to Rahmi Lale and Svein Valla for their contribution in initiating this work and for helpful discussions.

■ ABBREVIATIONS

ITS, initial transcribed sequence; PY, productive yield; AP, abortive probability; MSAT, maximal size of abortive transcript.

■ REFERENCES

(1) Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science 314*, 1139−1143.

(2) Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., and Ebright, R. H. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science 314*, 1144−1147.

(3) Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R., and Block, S. M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature 438*, 460−465.

(4) Borukhov, S., and Nudler, E. (2008) RNA polymerase: The vehicle of transcription. *Trends Microbiol. 16*, 126−134.

(5) Munson, L. M., and Reznikoff, W. S. (1981) Abortive initiation and long RNA synthesis. *Biochemistry 20*, 2081−2085.

(6) Gralla, J. D., Carpousis, A. J., and Stefano, J. E. (1980) Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry 19*, 5864−5869.

(7) Hsu, L. M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta 1577*, 191−207.

(8) Weixlbaumer, A., Leon, K., Landick, R., and Darst, S. A. (2013) Structural basis of transcriptional pausing in bacteria. *Cell 152*, 431−441.

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275

(9) Herbert, K. M., la Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R., and Block, S. M. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell 125*, 1083−1094.
(10) Hsu, L. M., Cobb, I. M., Ozmore, J. R., Khoo, M., Nahm, G., Xia, L., Bao, Y., and Ahn, C. (2006) Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry 45*, 8841−8854.
(11) Chander, M., Austin, K. M., Aye-Han, N.-N., Sircar, P., and Hsu, L. M. (2007) An alternate mechanism of abortive release marked by the formation of very long abortive transcripts. *Biochemistry 46*, 12687−12699.
(12) Hsu, L. M., Vo, N. V., Kane, C. M., and Chamberlin, M. J. (2003) *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry 42*, 3777−3786.
(13) Hsu, L. M. (1996) Quantitative parameters for promoter clearance. *Methods Enzymol. 273*, 59−71.
(14) Xue, X., Liu, F., and Ou-Yang, Z. (2008) A kinetic model of transcription initiation by RNA polymerase. *J. Mol. Biol. 378*, 520−529.
(15) Margeat, E., Kapanidis, A. N., Tinnefeld, P., Wang, Y., Mukhopadhyay, J., Ebright, R. H., and Weiss, S. (2006) Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys. J. 90*, 1419−1431.
(16) Yager, T. D., and von Hippel, P. H. (1991) A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry 30*, 1097−1118.
(17) von Hippel, P. H., and Yager, T. D. (1991) Transcript elongation and termination are competitive kinetic processes. *Proc. Natl. Acad. Sci. U.S.A. 88*, 2307−2311.
(18) Tadigotla, V. R., Maoiléidigh, D. Ó., Sengupta, A. M., Epshtein, V., Ebright, R. H., Nudler, E., and Ruckenstein, A. E. (2006) Thermodynamic and kinetic modeling of transcriptional pausing. *Proc. Natl. Acad. Sci. U.S.A. 103*, 4439−4444.
(19) Guajardo, R., and Sousa, R. (1997) A model for the mechanism of polymerase translocation. *J. Mol. Biol. 265*, 8−19.
(20) Bai, L., Shundrovsky, A., and Wang, M. D. (2004) Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol. 344*, 335−349.
(21) von Hippel, P. H. (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science 281*, 660−665.
(22) Greive, S. J., and von Hippel, P. H. (2005) Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol. 6*, 221−232.
(23) Bai, L., Shundrovsky, A., Wang, M. D. (2009) in *RNA polymerases as molecular motors* (Buc, H., and Strick, T., Eds.) pp 263−280, Royal Society of Chemistry, Cambridge, U.K.
(24) Hein, P. P., Palangat, M., and Landick, R. (2011) RNA transcript 3′-proximal sequence affects translocation bias of RNA polymerase. *Biochemistry 50*, 7002−7014.
(25) Malinen, A. M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M. S., and Belogurov, G. A. (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res. 40*, 7442−7451.
(26) Larson, M. H., Mooney, R. A., Peters, J. M., Windgassen, T., Nayak, D., Gross, C. A., Block, S. M., Greenleaf, W. J., Landick, R., and Weissman, J. S. (2014) A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science 344*, 1042−1047.
(27) Vvedenskaya, I. O., Vahedian-Movahed, H., Bird, J. G., Knoblauch, J. G., Goldman, S. R., Zhang, Y., Ebright, R. H., and Nickels, B. E. (2014) Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science 344*, 1285−1289.
(28) O Maoileidigh, D., Tadigotla, V. R., Nudler, E., and Ruckenstein, A. E. (2011) A unified model of transcription elongation: What have we learned from single-molecule experiments? *Biophys. J. 100*, 1157−1166.
(29) Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature 448*, 163−168.
(30) SantaLucia, J., and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct. 33*, 415−440.
(31) Wu, P., Nakano, S., and Sugimoto, N. (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem. 269*, 2821−2830.
(32) Hsu, L. M. (2009) Monitoring abortive initiation. *Methods 47*, 25−36.
(33) Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986) Functional dissection of *Escherichia coli* promoters: Information in the transcribed region is involved in late steps of the overall process. *EMBO J. 5*, 2995−3000.
(34) Bochkareva, A., Yuzenkova, Y., Tadigotla, V. R., and Zenkin, N. (2012) Factor-independent transcription pausing caused by recognition of the RNA−DNA hybrid sequence. *EMBO J. 31*, 630−639.
(35) Cheung, A. C. M., and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature 471*, 249−253.
(36) Wang, D., Bushnell, D. A., Huang, X., Westover, K. D., Levitt, M., and Kornberg, R. D. (2009) Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution. *Science 324*, 1203−1206.
(37) Larson, M. H., Landick, R., and Block, S. M. (2011) Single-molecule studies of RNA polymerase: One singular sensation, every little step it takes. *Mol. Cell 41*, 249−262.
(38) Kulbachinskiy, A., and Mustaev, A. (2006) Region 3.2 of the σ subunit contributes to the binding of the 3′-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem. 281*, 18273−18276.
(39) Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esyunina, D., Fernández-Tornero, C., Kulbachinskiy, A., and Murakami, K. S. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem. 289*, 24549−24559.
(40) Vahia, A. V., and Martin, C. T. (2011) Direct tests of the energetic basis of abortive cycling in transcription. *Biochemistry 50*, 7015−7022.
(41) Samanta, S., and Martin, C. T. (2013) Insights into the mechanism of initial transcription in *Escherichia coli* RNA polymerase. *J. Biol. Chem. 288*, 31993−32003.
(42) Strobel, E. J., and Roberts, J. W. (2014) Regulation of promoter-proximal transcription elongation: Enhanced DNA scrunching drives λQ antiterminator-dependent escape from a σ70-dependent pause. *Nucleic Acids Res. 42*, 5097−5108.

DOI: 10.1021/acs.biochem.5b00272
Biochemistry 2015, 54, 4267−4275