Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection

Irina O. Vvedenskaya[a,b,1], Hanif Vahedian-Movahed[b,c,1], Yuanchao Zhang[a,d], Deanne M. Taylor[a,d,e,f], Richard H. Ebright[b,c,2], and Bryce E. Nickels[a,b,2]

[a]Department of Genetics, Rutgers University, Piscataway, NJ 08854; [b]Waksman Institute, Rutgers University, Piscataway, NJ 08854; [c]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854; [d]Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; [e]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; and [f]Department of Obstetrics, Gynecology and Reproductive Sciences, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901

Edited by Lucia B. Rothman-Denes, The University of Chicago, Chicago, IL, and approved April 7, 2016 (received for review February 26, 2016)

During transcription initiation, RNA polymerase (RNAP) holoenzyme unwinds ∼13 bp of promoter DNA, forming an RNAP-promoter open complex (RPo) containing a single-stranded transcription bubble, and selects a template-strand nucleotide to serve as the transcription start site (TSS). In RPo, RNAP core enzyme makes sequence-specific protein–DNA interactions with the downstream part of the nontemplate strand of the transcription bubble ("core recognition element," CRE). Here, we investigated whether sequence-specific RNAP–CRE interactions affect TSS selection. To do this, we used two next-generation sequencing-based approaches to compare the TSS profile of WT RNAP to that of an RNAP derivative defective in sequence-specific RNAP–CRE interactions. First, using massively systematic transcript end readout, MASTER, we assessed effects of RNAP–CRE interactions on TSS selection in vitro and in vivo for a library of 4^7 (∼16,000) consensus promoters containing different TSS region sequences, and we observed that the TSS profile of the RNAP derivative defective in RNAP–CRE interactions differed from that of WT RNAP, in a manner that correlated with the presence of consensus CRE sequences in the TSS region. Second, using 5′ merodiploid native-elongating-transcript sequencing, 5′ mNET-seq, we assessed effects of RNAP–CRE interactions at natural promoters in *Escherichia coli*, and we identified 39 promoters at which RNAP–CRE interactions determine TSS selection. Our findings establish RNAP–CRE interactions are a functional determinant of TSS selection. We propose that RNAP–CRE interactions modulate the position of the downstream end of the transcription bubble in RPo, and thereby modulate TSS selection, which involves transcription bubble expansion or transcription bubble contraction (scrunching or antiscrunching).

RNA polymerase | transcription start site selection | promoter | transcription bubble | transcription initiation

Transcription initiation consists of a number of biochemical steps leading to formation of a phosphodiester bond between a nucleoside triphosphate (NTP) bound in the RNA polymerase (RNAP) active-center initiating NTP binding site (i site) and an NTP bound in the RNAP active-center extending NTP binding site (i+1 site) (1–3). For bacterial RNAP, promoter-specific initiation requires the RNAP core enzyme (subunit composition $\alpha_2\beta\beta'\omega$) to associate with a σ factor forming the RNAP holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma$). The σ factor contains determinants for sequence-specific protein–DNA interactions with four core promoter elements: the −35 element, the extended −10 element, the −10 element, and the discriminator element (4).

During transcription initiation, RNAP holoenzyme unwinds promoter DNA to form an RNAP-promoter open complex (RPo) containing an unwound, single-stranded "transcription bubble." The process of promoter unwinding begins within the promoter −10 element and propagates downstream, enabling single-stranded nucleotides at the downstream end of the transcription bubble template strand to occupy the RNAP active center i and i+1 sites (Fig. 1*A*) (1–3). In particular, in RPo, the second-most downstream nucleotide of the transcription bubble template strand occupies the active center i site and serves as the transcription start site (TSS), and the downstream-most nucleotide of the transcription bubble template strand occupies the active center i+1 site. We designate the template-strand nucleotide at the TSS position as TSS_T (Fig. 1, base in pink) and the template-strand nucleotide at the next base pair as $TSS{+}1_T$ (Fig. 1, base in red).

The position of the TSS relative to the position of the promoter −10 element is variable (5–11). TSS selection preferentially occurs at the position 7-bp downstream of the promoter −10 element, but can occur over a range of at least five positions, encompassing the positions 6-, 7-, 8-, 9-, or 10-bp downstream of the promoter −10 element. Thus, there must be flexibility in the structure of RPo that enables the position of the TSS to vary relative to the position of the −10 element. We previously have proposed that variability in TSS selection is mediated by variability in the size of the unwound transcription bubble (Fig. S1*A*) (11–13). According to this model, RPo generally contains a 13-bp unwound transcription bubble that places the template-strand nucleotide 7-bp downstream of the −10 element in the i site and places the template-strand nucleotide 8-bp downstream of the −10 element in the i+1 site (Fig. 1*A* and Fig. S1*A*) (TSS = 7). For TSS selection to occur

Significance

For all cellular RNA polymerases, the position of the transcription start site (TSS) relative to core promoter elements is variable. Furthermore, environmental conditions and regulatory factors that affect TSS selection have profound effects on levels of gene expression. Thus, identifying determinants of TSS selection is important for understanding gene expression control. Here we identify a previously undocumented determinant for TSS selection by *Escherichia coli* RNA polymerase. We show that sequence-specific protein–DNA interactions between RNA polymerase core enzyme and a sequence element in unwound promoter DNA, the core recognition element, modulate TSS selection.

Author contributions: I.O.V., H.V.-M., R.H.E., and B.E.N. designed research; I.O.V. and H.V.-M. performed research; I.O.V., H.V.-M., Y.Z., D.M.T., R.H.E., and B.E.N. analyzed data; and R.H.E. and B.E.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the NIH/NCBI Sequence Read Archive (accession no. SRP071742).

[1]I.O.V. and H.V.-M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: ebright@waksman.rutgers.edu or bnickels@waksman.rutgers.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603271113/-/DCSupplemental.

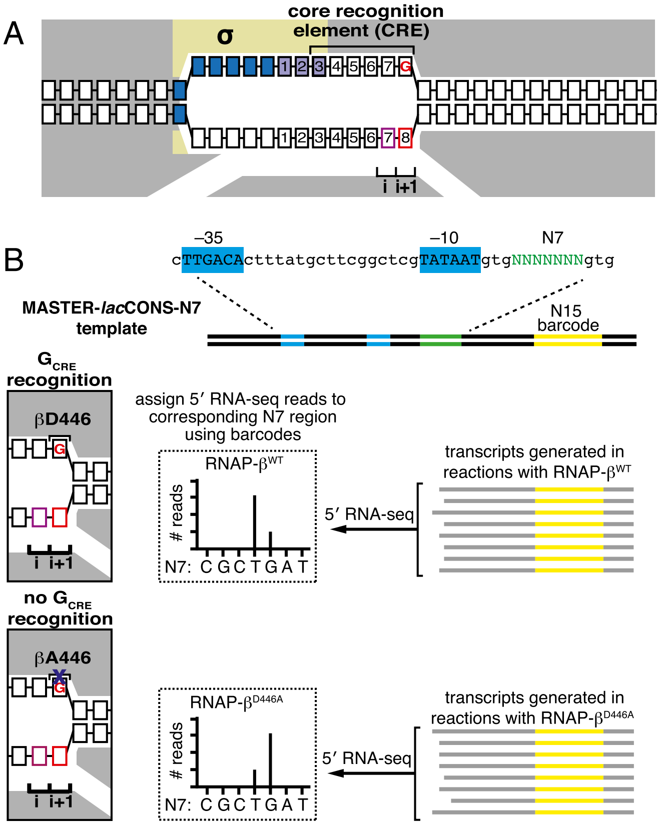

A
σ
core recognition element (CRE)
1 2 3 4 5 6 7 G
1 2 3 4 5 6 7 8
i i+1
B
−35
−10
N7
cTTGACActttatgcttcggctcgTATAATgtgNNNNNNNgtg
MASTER-*lac*CONS-N7 template
N15 barcode
G_{CRE} recognition
βD446
G
i i+1
assign 5′ RNA-seq reads to corresponding N7 region using barcodes
RNAP-β^{WT}
# reads
N7: C G C T G A T
5′ RNA-seq
transcripts generated in reactions with RNAP-β^{WT}
no G_{CRE} recognition
βA446
G
i i+1
RNAP-β^{D446A}
# reads
N7: C G C T G A T
5′ RNA-seq
transcripts generated in reactions with RNAP-β^{D446A}

Fig. 1. Analysis of effects of sequence-specific RNAP–CRE interactions by MASTER (11). (*A*) RPo for TSS at position 7. Gray, RNAP; yellow, σ; blue boxes, −10 element nucleotides; purple boxes, discriminator nucleotides; black boxes, DNA nucleotides (non–template-strand nucleotides above template-strand nucleotides; nucleotides downstream of −10 element numbered); pink box, TSS_T; red box, $TSS+1_T$; i and i+1, RNAP active-center initiating NTP binding site and extending NTP binding site; red “G,” G_{CRE}. (*B*, *Upper*) DNA fragment carrying the MASTER template library *lac*CONS-N7. Promoter −35 and −10 elements are indicated. Randomized nucleotides are green and 15-nt barcode sequence in the transcribed region is yellow. (*Lower Right*) 5′ RNA-seq analysis of RNA products generated from the MASTER-N7 template library in vitro. The sequence of the barcode is used to assign the RNA product to an N7 region and the sequence of the 5′ end is used to define the TSS. (*Lower Left*) Structural organization of downstream end of transcription bubble in RPo for promoter containing G_{CRE} formed with WT RNAP (*Upper*) or RNAP derivative carrying the βD446A substitution (*Lower*). Black “βD446,” RNAP β-subunit residue that makes sequence-specific favorable interaction with G_{CRE}. Black “βA446,” RNAP β-subunit residue in mutant RNAP defective in sequence-specific interaction with G_{CRE}. Other rendering and colors as in *A*.

at positions further downstream, the downstream DNA duplex is unwound, the unwound DNA is pulled into and past the RNAP active center, and the unwound DNA is accommodated as single-stranded DNA bulges within the transcription bubble, yielding a “scrunched” complex (Fig. S1*A*) (TSS = 8 and TSS = 9). For TSS selection to occur at positions further upstream, the opposite occurs: downstream DNA is rewound, downstream DNA is extruded from the RNAP active center, and the extrusion of DNA from the RNAP active center is accommodated by stretching DNA within the transcription bubble, yielding an “antiscrunched” complex (Fig. S1*A*) (TSS = 6). According to this model, any protein–DNA or protein–protein interaction that affects the energy landscape for transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo potentially could modulate TSS selection (13, 14).

In the structure of RPo, the RNAP core makes direct protein–DNA interactions with the non–template-strand DNA segment at the downstream part of the transcription bubble (15); this DNA segment has been designated the “core recognition element” (CRE; Fig. 1*A*) (15). RNAP–CRE interactions with the non–template-strand nucleotide at the extreme downstream end of the transcription bubble (i.e., $TSS+1_{NT}$) are sequence specific, with preference for the base G (G_{CRE}) (Fig. 1, red G) (15).

It has been proposed that sequence-specific RNAP–G_{CRE} interactions facilitate promoter unwinding to form the transcription bubble, stabilize the unwound transcription bubble, and define the downstream end of the transcription bubble (15). According to this proposal, sequence-specific RNAP–G_{CRE} interactions should affect the energy landscape for transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo and therefore potentially could affect TSS selection (Fig. S1*B*). Here we tested the proposal that sequence-specific RNAP–G_{CRE} interactions affect TSS selection. To do this, we used high-throughput sequencing–based approaches to compare TSS selection by WT RNAP to TSS selection by a mutant RNAP defective in sequence-specific RNAP–G_{CRE} interactions. Our results demonstrate that sequence-specific RNAP–CRE interactions are a determinant of TSS selection.

Results

Sequence-Specific RNAP–CRE Interactions Are a Determinant of TSS Selection in Vitro. In crystal structures of RNAP–promoter open complexes, residue D446 of the RNAP β subunit makes direct H-bonded interactions with Watson–Crick H-bond–forming atoms of G at G_{CRE} (15). The interactions by βD446 determine specificity at G_{CRE}. Thus, substitution of βD446 by alanine eliminates the ability of RNAP to distinguish A, G, C, and T at the G_{CRE} position (16). Accordingly, an RNAP derivative carrying the βD446A substitution can serve as a reagent to assess the functional significance of sequence-specific RNAP-G_{CRE} interactions (Fig. 1*B*, *Lower Left*).

To define the contribution of sequence-specific RNAP–G_{CRE} interactions to TSS selection, we used a high-throughput sequencing–based methodology termed massively systematic transcript end readout (MASTER) (11). MASTER entails the construction of a template library that contains up to 4^{10} (~1,000,000) bar-coded sequences, production of RNA transcripts from the template library in vitro or in vivo, and analysis of transcript ends using high-throughput sequencing (11, 13).

To analyze the effect of disrupting sequence-specific RNAP–G_{CRE} interactions on TSS selection, we used a MASTER template library, *lac*CONS-N7, that contained 4^7 (~16,000) sequence variants at positions 4–10 bp downstream of the −10 element of a consensus *Escherichia coli* σ^{70}-dependent promoter (Fig. 1*B*, *Upper*) (11). We performed in vitro transcription experiments with the *lac*CONS-N7 template library, using, in parallel, WT RNAP (RNAP-β^{WT}) or the RNAP derivative containing the βD446A substitution (RNAP-β^{D446A}). RNA products generated in the transcription reactions were isolated and analyzed using high-throughput sequencing of RNA barcodes and 5′ ends (5′ RNA-seq) to define, for each RNA product, the template that produced the RNA and the TSS position (Fig. 1*B*, *Lower Right*). For each sequence variant, we calculated the percentage of reads starting at each position within the randomized TSS region, $\%TSS_Y = 100 \times$ (no. reads starting at position Y/total no. reads starting at positions 4–10).

To determine the effect of disrupting RNAP–G_{CRE} interactions on TSS selection, we considered TSS positions where $TSS+1_{NT}$ is included within the randomized region of the MASTER template library (i.e., TSS positions 6, 7, 8, and 9). We first calculated %TSS values for each of these positions on the basis of the identity of $TSS+1_{NT}$. Thus, for each TSS position, we averaged the %TSS values for the ~4,000 templates having A at $TSS+1_{NT}$, the ~4,000 templates having C at $TSS+1_{NT}$, the ~4,000 templates having G at $TSS+1_{NT}$, and the ~4,000 templates having T at $TSS+1_{NT}$. Next, we calculated the difference in these %TSS values for reactions performed with RNAP-β^{WT} vs. reactions performed with RNAP-β^{D446A}. We observed that, for all four tested TSS positions (positions 6, 7, 8, and 9), the βD446A substitution decreased the %TSS when TSS+1_{NT} was G (1.3–7.3% decreases; Fig. 2*A*, top row of table). In contrast, for three of the four tested TSS positions (positions 6, 7, and 8), the βD446A substitution did not decrease the %TSS when TSS+1_{NT} was A, C, or T, and, for the fourth position (position 9), the βD446A substitution did not decrease the %TSS, or decreased the %TSS by smaller amounts, when TSS+1_{NT} was A, C, or T (Fig. 2*A*, bottom three rows of table).

We identified 1,230 TSS positions (5.6% of the 21,872 above-threshold TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element) that exhibited large, ≥20%, reductions in %TSS in reactions performed with RNAP-β^{D446A} vs. reactions performed with RNAP-β^{WT}. For these 1,230 TSS positions with large, ≥20%, CRE effects, ~90% contained G at TSS+1_{NT} (Fig. 2*B*, top row, *Right*), whereas, for the total sample of 21,872 TSS positions, there were no detectable sequence preferences at position TSS+1_{NT} (Fig. 2*B*, top row, *Left*). Enrichment of G at TSS+1_{NT} for TSS position with large, ≥20%, CRE effects was observed for TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element (TSS = 6, 7, 8, or 9) (Fig. 2*B*, bottom four rows). In summary, the overwhelming majority of TSS positions that exhibit large, ≥20%, CRE effects have G at TSS+1_{NT}.

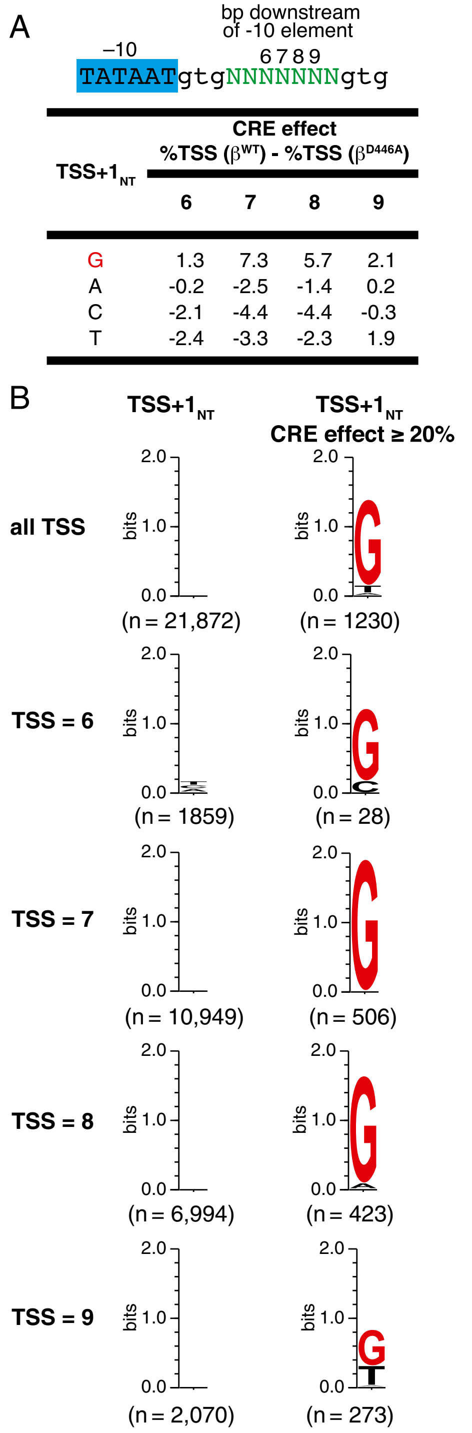
To validate the MASTER results, we performed further analyses of two TSS region sequences that exhibited large, ≥20%, CRE effects, contained a TSS at the most common position (position 7), and contained G at TSS+1_{NT} (position 8) (Fig. 3). For each of these two TSS region sequences, we prepared templates containing G, A, C, or T at position TSS+1_{NT}, performed in vitro transcription experiments with RNAP-β^{WT} or RNAP-β^{D446A}, and analyzed RNA products by primer extension. For each of the two sets of constructs, the primer-extension results matched the MASTER results. A large, ~30%, CRE effect was observed when TSS+1_{NT} was G but not when TSS+1_{NT} was A, C, or T (Fig. 3).

The results in Figs. 2 and 3 establish that disrupting sequence-specific RNAP–G_{CRE} interactions affects TSS selection in vitro in a manner that correlates with the presence and position of G_{CRE} in the TSS region. We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in vitro.

Sequence-Specific RNAP–CRE Interactions Are a Determinant of TSS Selection in Vivo.

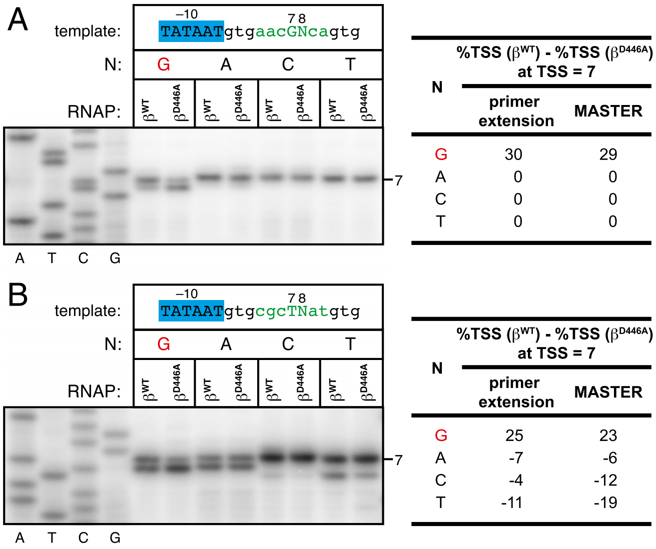
Analysis of 4^7 (~16,000) consensus promoter derivatives. To define the contribution of sequence-specific RNAP–G_{CRE} interactions to TSS selection in vivo, we used merodiploid native-elongating transcript sequencing (mNET-seq) (16). mNET-seq involves selective analysis of transcripts associated with an epitope-tagged RNAP in the presence of a mixed population of epitope-tagged RNAP and untagged RNAP (Fig. 4*A*). In prior work, we used mNET-seq to determine the effect of sequence-specific RNAP–G_{CRE} interactions on pausing during elongation (16). In this work, we used a variant of mNET-seq, 5′ mNET-seq, to determine the effect of sequence-specific RNAP–G_{CRE} interactions on TSS selection (Fig. 4*A*). To do this, we introduced into cells a plasmid encoding 3xFLAG-tagged β^{WT} or 3× FLAG-tagged β^{D446A}, isolated RNA products associated with RNAP-β^{WT} or RNAP-β^{D446A} by immunoprecipitation, converted RNA 5′ ends to cDNAs, and performed high-throughput sequencing (Fig. 4*A*).

To enable direct comparison of in vivo and in vitro results, we performed 5′ mNET-seq using the same MASTER template library of 4^7 (~16,000) consensus core promoter derivatives that we used for in vitro analysis. The results of MASTER in vivo (Fig. 4 *B* and *C*) matched the results of MASTER in vitro (Fig. 2). For all four tested TSS positions (positions 6, 7, 8, and 9), the βD446A substitution decreased the %TSS when TSS+1_{NT} was G (0.6–7.3% decreases) (Fig. 4*B*, top row of table). In contrast, for three of the four tested TSS positions (positions 6, 7, and 8), the βD446A substitution did not decrease the %TSS when TSS+1_{NT} was A, C, or T, and, for the fourth position (position 9), the



TSS+1_{NT}	CRE effect %TSS (β^{WT}) - %TSS (β^{D446A})			
	6	7	8	9
G	1.3	7.3	5.7	2.1
A	-0.2	-2.5	-1.4	0.2
C	-2.1	-4.4	-4.4	-0.3
T	-2.4	-3.3	-2.3	1.9

Fig. 2. Effects of disrupting RNAP–G_{CRE} interactions in vitro: analysis by MASTER. (*A*) Effect of sequence at TSS+1_{NT} on %TSS for RNAP-β^{WT} vs. RNAP-β^{D446A}. Table lists the difference in %TSS (%TSS for RNAP-β^{WT} – %TSS for RNAP-β^{D446A}) at positions 6, 7, 8, or 9 for TSS-regions carrying G, A, C, or T at TSS+1_{NT}. (*B*) Sequence preferences for TSS+1_{NT}. Sequence logo (33) for TSS+1_{NT} of above-threshold TSS (*Left*) and TSS that exhibited a large, ≥20%, reduction in %TSS in reactions performed with RNAP-β^{D446A} vs. reactions performed with RNAP-β^{WT} (*Right*).



N	%TSS (β^{WT}) - %TSS (β^{D446A}) at TSS = 7	
	primer extension	MASTER
G	30	29
A	0	0
C	0	0
T	0	0

N	%TSS (β^{WT}) - %TSS (β^{D446A}) at TSS = 7	
	primer extension	MASTER
G	25	23
A	-7	-6
C	-4	-12
T	-11	-19

Fig. 3. Effects of disrupting RNAP–G_{CRE} interactions in vitro: analysis by primer extension. (*Left*) Primer-extension results. RNA products were generated in reactions performed with RNAP-β^{WT} or RNAP-β^{D446A} and p*lac*CONS templates carrying TSS region sequences (in green) of AACGNCA (*A*) or CGCTNAT (*B*), where N is G, A, C, or T. Bands corresponding to a TSS at position 7 are indicated. (*Right*) Table lists the difference in %TSS (%TSS for reactions with RNAP-β^{WT} - %TSS for reactions with RNAP-β^{D446A}) at position 7 for templates carrying a G, A, C, or T at position 8 calculated by primer extension or calculated by MASTER.

βD446A substitution decreased the %TSS by smaller amounts when TSS+1_{NT} was A, C, or T (Fig. 4*B*, bottom three rows of table). Furthermore, we identified 860 TSS positions (4.3% of the 20,217 above-threshold TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element) with large, ≥20%, CRE effects. For these 860 TSS positions with large, ≥20%, CRE effects, ∼80% contained G at TSS+1_{NT} (Fig. 4*C*, *Right*), whereas, for the total sample of 20,217 TSS positions, there were no detectable sequence preferences at position TSS+1_{NT} (Fig. 4*C*, *Left*).

The results establish that disrupting sequence-specific RNAP–G_{CRE} interactions affects TSS selection in vivo in a manner that correlates with the presence and position of G_{CRE} in the TSS region. We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in vivo.

Analysis of **E. coli** ***transcriptome.*** Having shown by MASTER that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in the context of a consensus core promoter in vivo, we next assessed the contribution of sequence-specific RNAP–CRE interactions to TSS selection in the context of natural promoters in vivo in *E. coli*. (The primers used in the in vivo MASTER analysis by 5′ mNET-seq shown in Fig. 4 provided information only about transcripts from the synthetic consensus promoter derivatives. This is because the primers used for synthesis of the first cDNA strand annealed only to transcripts produced from the synthetic consensus promoter derivatives. A separate experiment, with primers that enable generation of cDNAs from transcripts produced from natural *E. coli* promoters, was necessary to provide information about transcripts from natural *E. coli* promoters. Therefore, to analyze transcripts from natural *E. coli* promoters, the primers used for synthesis of the first cDNA strand carried nine randomized nucleotides at the 3′ end.)

Using data from experiments performed with RNAP-β^{WT}, we identified 1,500 above-threshold TSS positions associated with natural promoters in *E. coli*. Of these 1,500 TSS positions, we identified 44 TSS positions that exhibited large, ≥20%, CRE effects (Table S1); 39 of these 44 (∼90%) contained G at TSS+1_{NT} (Fig. 5*B*, *Right*, and Table S1), whereas for the total sample of 1,500 above-threshold TSS, there were no detectable sequence preferences at TSS+1_{NT} (Fig. 5*B*, *Left*).

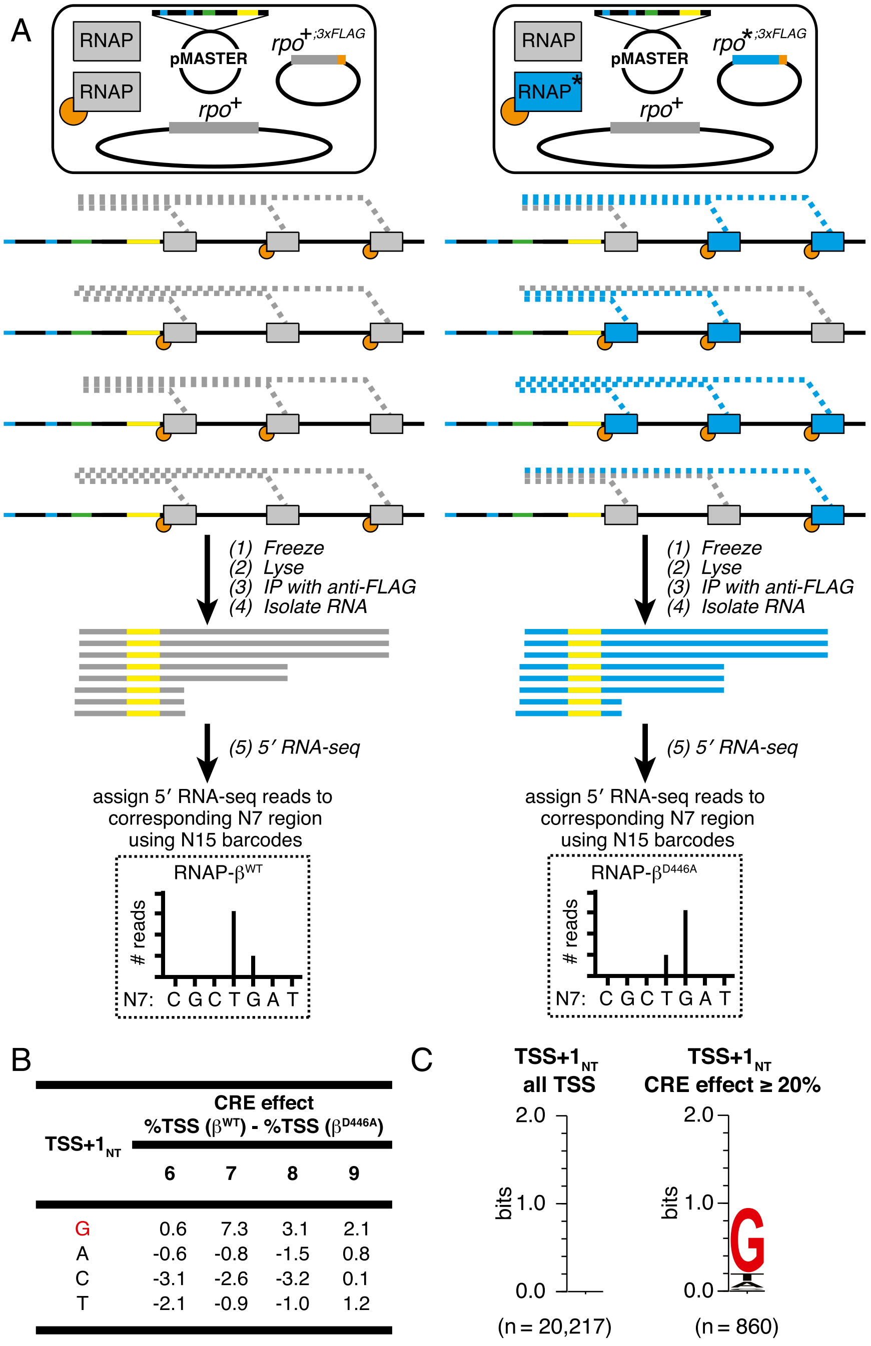
To validate the 5′ mNET-seq results, we performed primer-extension experiments with two *E. coli* promoters that contained a TSS that exhibited a large, ≥20%, CRE effect and contained G at TSS+1_{NT}: P*secE* and P*hemC* (Table S1). We generated linear templates carrying P*secE* or P*hemC*, performed in vitro transcription assays using RNAP-β^{WT} or RNAP-β^{D446A}, and analyzed TSS selection by primer extension (Fig. 5*C*). For each promoter, two prominent start sites were observed in reactions with RNAP-β^{WT}. In the case of P*secE*, ∼60% of the transcripts started at an A located 7-bp downstream of the predicted −10 element (A7) and ∼40% of the transcripts started at a G located 8-bp downstream (G8) (Fig. 5*C*, *Left*). In the case of P*hemC*, ∼30% of the transcripts started at an A located 6-bp downstream of the predicted −10 element (A6) and ∼70% of the transcripts started at a G located 8-bp downstream (G8) (Fig. 5*C*, *Right*). For each promoter, the percentage of transcripts starting at the position that contained G at TSS+1_{NT} (A7 for P*secE* and G8 for P*hemC*) was reduced by ∼30% when reactions were performed with RNAP-β^{D446A} (Fig. 5*C*), consistent with results of 5′ mNET-seq (Table S1). We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in natural promoters in the *E. coli* genome.

Discussion

Sequence-Specific RNAP–CRE Interactions in TSS Selection. Here we show that sequence-specific interactions between RNAP and the downstream segment of the nontemplate strand of the transcription bubble (CRE) are a determinant of TSS selection. In particular, using high-throughput sequencing–based approaches, we define a role of sequence-specific recognition of a G at the most downstream position of the CRE (G_{CRE}) during TSS selection in the context of a library of 4^7 (∼16,000) TSS region sequences of a consensus core promoter in vitro and in vivo (Figs. 2–4) and in the context of natural promoters in *E. coli in vivo* (Fig. 5 and Table S1).

As discussed above, variability in TSS selection is believed to involve transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo (Fig. S1*A*) (11–14). We propose that the observed effects of sequence-specific RNAP–CRE interactions on TSS selection occur by influencing transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo (Fig. S1*B*). Specifically, we propose that sequence-specific RNAP–CRE interactions favor TSS selection at sequences that contain G at TSS+1_{NT}. According to this proposal, the role of sequence-specific RNAP–CRE interactions in defining the downstream edge of the transcription bubble concurrently defines the extent of transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo and therefore modulates TSS selection (Fig. S1*B*).

The results of this work, together with results of previous work, establish that TSS selection involves at least four promoter sequence determinants: (*i*) position relative to the −10 element (preference for the position 7-bp downstream of the −10 element) (5–11); (*ii*) sequence of TSS_T and TSS-1_T (strong preference for pyrimidine at TSS_T and preference for purine at TSS-1_T, which enable initiation with a purine NTP and maximize stacking between DNA bases and the initiating purine NTP) (11, 17–20); (*iii*) sequence of the discriminator element (preference for TSS selection at upstream positions for discriminator sequences that disfavor scrunching and preference for TSS selection at downstream positions for discriminator sequences that favor scrunching) (13, 14); and (*iv*) sequence of the CRE (preference for G at TSS+1_{NT}). In addition to these sequence determinants, DNA topology and NTP concentrations also



$TSS+1_{NT}$	CRE effect %TSS (β^{WT}) - %TSS (β^{D446A})			
	6	7	8	9
G	0.6	7.3	3.1	2.1
A	-0.6	-0.8	-1.5	0.8
C	-3.1	-2.6	-3.2	0.1
T	-2.1	-0.9	-1.0	1.2

Fig. 4. Effects of disrupting RNAP–G_{CRE} interactions in vivo: 5′ mNET-seq analysis of 4^7 (~16,000) consensus promoter derivatives. (*A*) Steps in 5′ mNET-seq analysis of TSS selection from plasmid-borne MASTER template library: (*Top*) RNAP derivatives in cells (the blue RNAP derivative with asterisk is RNAP-β^{D446A}); (*Middle*) RNAPs on the same MASTER template in four cells (RNA products in blue are associated with RNAP-β^{D446A}); and (*Bottom*) isolation of RNA products after immunoprecipitation with anti-FLAG affinity gel and sequencing analysis of RNA 5′ ends. In this example, TSS selection at the T in the middle of the randomized TSS region is decreased with the mutant RNAP derivative. (*B*) Effect of sequence at $TSS+1_{NT}$ on %TSS for RNAP-β^{WT} vs. RNAP-β^{D446A}. Table lists the difference in %TSS (%TSS for RNAP-β^{WT} – %TSS for RNAP-β^{D446A}) at positions 6, 7, 8, or 9 for TSS-regions carrying G, A, C, or T at $TSS+1_{NT}$. (*C*) Sequence preferences for $TSS+1_{NT}$. Sequence logo (33) for $TSS+1_{NT}$ of above-threshold TSS positions located 6–9 bp downstream of the –10 element (*Left*) and TSS positions located 6–9 bp downstream of the –10 element that exhibited a large, ≥20%, reduction in %TSS in 5′ mNET-seq analysis of RNAP-β^{D446A} vs. 5′ mNET-seq analysis of RNAP-β^{WT} (*Right*).

influence TSS selection (6, 8, 9, 11, 21–26). Thus, TSS selection is a multifactorial process, in which the ultimate outcome for a given promoter reflects the contributions of multiple promoter sequence determinants and multiple reaction conditions. Because sequence-specific RNAP–CRE interactions are only one of several determinants of TSS selection, their quantitative significance at different promoters differs. At some promoters, such as P*secE* and P*hemC*, sequence-specific RNAP–CRE interactions have quantitatively large, ≥20%, effects on TSS selection (Fig. 5*C* and Table S1), whereas at other promoters, the quantitative effects of RNAP–CRE interactions are smaller.

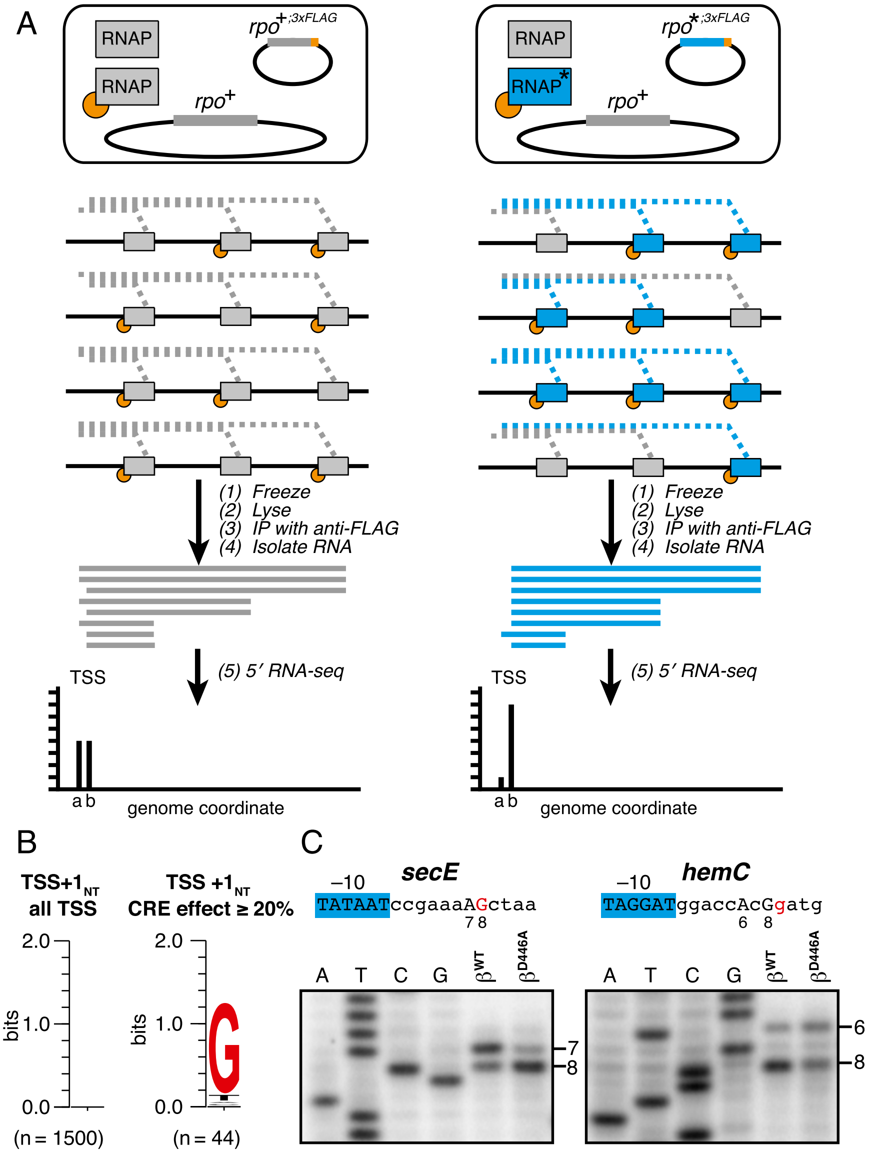


Fig. 5. Effects of disrupting RNAP-G_{CRE} interactions in vivo: 5′ mNET-seq analysis of *E. coli* transcriptome. (*A*) Steps in 5′ mNET-seq analysis of natural promoters: (*Top*) RNAP derivatives in cells (the blue RNAP derivative with asterisk is RNAP-β^{D446A}); (*Middle*) RNAPs on the same transcription unit in four cells (RNA products in blue are associated with RNAP-β^{D446A}); and (*Bottom*) isolation of RNA products after immunoprecipitation with anti-FLAG affinity gel and sequencing analysis of RNA 5′ ends. In this example, TSS selection at genome coordinate labeled "a" is decreased with the mutant RNAP derivative. (*B*) Sequence preferences for TSS+1_{NT}. Sequence logo (33) for TSS+1_{NT} of above-threshold TSS associated with natural promoters (*Left*) and TSS associated with natural promoters that exhibited a large, ≥20%, reduction in %TSS in 5′ mNET-seq analysis of RNAP-β^{D446A} vs. RNAP-β^{WT} (Table S1). (*C*) Primer-extension analysis of TSS selection in vitro from natural promoters. RNA products were generated in reactions performed with RNAP-β^{WT} or RNAP-β^{D446A} and templates carrying P*secE* (*Left*) or P*hemC* (*Right*). The sequence of each promoter, including the −10 element and 12 downstream bp, is provided. In the case of P*secE*, bands corresponding to a TSS at A7 or G8 are indicated. In the case of P*hemC*, bands corresponding to a TSS at A6 or G8 are indicated. Base in red is G_{CRE} associated with the TSS at A7 of P*secE* or with the TSS at G8 of P*hemC*.

Prospect. In prior work, we showed that sequence-specific RNAP–CRE interactions affect RPo formation during transcription initiation, RPo stability during transcription initiation, translocational bias during transcription elongation, and sequence-specific pausing during transcription elongation (15, 16). Accordingly, our findings that sequence-specific RNAP–CRE interactions are a determinant of TSS selection add to an emerging view that sequence-specific RNAP–CRE interactions play functionally important roles during all stages of transcription that involve an unwound transcription bubble. A priority for future work will be to assess the roles of sequence-specific RNAP–CRE interactions in other steps of transcription that involve an unwound transcription bubble (e.g., transcriptional slippage, initial transcription, promoter escape, factor-dependent pausing, and termination). Another priority for future work will be to assess possible roles of sequence-specific RNAP–CRE interactions in eukaryotic transcription, noting that RNAP residues involved in sequence-specific RNAP–CRE interactions are conserved in bacteria and eukaryotes.

Materials and Methods

Details for all procedures are in the *SI Materials and Methods.*

Plasmids and Oligonucleotides. Plasmids are listed in Table S2. Oligonucleotides are listed in Table S3.

Proteins. RNAP-β^{WT} holoenzyme and RNAP-β^{D446} holoenzyme were prepared from *E. coli* strain XE54 (27) transformed with plasmids pRL706 or pRL706-βD446A, respectively, using procedures described in ref. 28.

In Vitro Transcription Assays. For MASTER experiments shown in Fig. 2, single round in vitro transcription assays were performed essentially as described in ref. 11 using a linear DNA template containing the p*lacCONS-N7* library (Fig. 1*B*, *Upper*). RNA products were purified and TSS selection was analyzed by 5′ RNA-seq as described in ref. 11 (see Table S4 for list of samples). In vitro transcription assays shown in Figs. 3 and 5*C* were performed essentially as described in ref. 29. RNA products generated in these reactions were analyzed by primer extension as described in ref. 29.

5′ mNET-seq. For the in vivo MASTER experiments shown in Fig. 4, E. *coli* DH10B-T1^{R} cells (Life Technologies) containing plasmids pRL706-$\beta^{WT;3xFLAG}$ or pRL706-$\beta^{D446A;3xFLAG}$ were transformed with ~50 ng pMASTER-*lacCONS-N7* library to obtain a 25-mL overnight culture representing cells derived from at least 20 million unique transformants; 0.5 mL of the overnight cell culture was used to inoculate 50 mL LB media containing 100 μg/μL carbenicillin and 25 μg/μL chloramphenicol. When the cell density reached an OD_{600} ~0.3, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added, and cells were grown for an additional 2 h. RNA associated with RNAP was isolated using procedures described in ref. 16.

For the experiments shown in Fig. 5, MG1655 cells containing plasmids pRL706-$\beta^{WT;3xFLAG}$ or pRL706-$\beta^{D446A;3xFLAG}$ were shaken at 220 rpm at 37 °C in 100 mL 4× LB (40 g Bacto tryptone, 20 g Bacto yeast extract, and 10 g NaCl per liter) containing 200 μg/μL carbenicillin in 500-mL DeLong flasks (Bellco). When cell density reached an OD_{600} ~0.6, 1 mM IPTG was added, and cells were grown for an additional 4 h. RNA associated with RNAP was isolated using procedures described in ref. 16.

RNA products associated with RNAP were analyzed by 5′ RNA-seq using procedures described in ref. 30 (see Table S4 for list of samples).

In Vitro and in Vivo MASTER Data Analysis. Analysis of 5′ RNA-seq data obtained from MASTER experiments was performed essentially as described in ref. 11. Sequencing of template DNA was used to associate the 7-bp randomized TSS region sequence with a corresponding second 15-bp randomized sequence that serves as its barcode. Reads that contained a perfect match to the DNA template from which they were derived were used for the analysis of TSS selection. The percentage of reads starting at a given TSS position (%TSS) was calculated using the following formula: $\%TSS_Y = 100 \times$ (no. reads starting at position Y/total no. reads starting at positions 4–10). Above-threshold TSS positions were those for which the %TSS value was ≥20%.

5′ mNET-seq Analysis of Natural Promoters in Vivo in *E. coli*. Identification of TSS positions and TSS regions for natural promoters in *E. coli* was done essentially as described in ref. 31. The first six bases of each read were trimmed (to remove sequences introduced during the cDNA library construction procedure), and the next 30 bases were aligned to the *E. coli* reference genome (NC_000913.3) using Bowtie (32). Among these reads, we used those that aligned to a unique position in the genome with zero mismatches for the analysis of TSS selection.

Using data derived from the analysis of RNA products associated with RNAP-β^{WT}, we defined a list of primary TSS positions that met the following two criteria: (*i*) the read count at the coordinate was above a threshold value (≥50 reads) and (*ii*) the read count at the coordinate represented a local maximum in an 11-bp window centered on the coordinate. For each primary TSS position, we designated the positions spanning 5-bp upstream to 5-bp downstream as a TSS region. Next, for each TSS region, we calculated the percentage of reads starting at each of the 11 positions: $\%TSS_Y = 100 \times$ (no. reads starting at position Y/total no. reads starting within the TSS region). We identified 1,500 TSS positions within TSS regions with an above-threshold value of %TSS (≥20%). For each of these 1,500 TSS positions, we calculated the difference between the average %TSS observed in experiments performed with RNAP-β^{WT} and that observed in experiments performed with RNAP-β^{D446A}. TSS positions for which this difference was ≥20% are listed in Table S1.

ACKNOWLEDGMENTS. We thank Jared Knoblauch for assistance with data analysis. This work was supported by National Institutes of Health Grants GM041376 (to R.H.E.), GM088343 (to B.E.N.), GM096454 (to B.E.N.), and GM115910 (to B.E.N.).

1. Saecker RM, Record MT, Jr, Dehaseth PL (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* 412(5):754–771.
2. Decker KB, Hinton DM (2013) Transcription regulation at the core: Similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* 67:113–139.
3. Ruff EF, Record MT, Jr, Artsimovitch I (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5(2):1035–1062.
4. Feklístov A, Sharon BD, Darst SA, Gross CA (2014) Bacterial sigma factors: A historical, structural, and genomic perspective. *Annu Rev Microbiol* 68:357–376.
5. Aoyama T, Takanami M (1985) Essential structure of *E. coli* promoter II. Effect of the sequences around the RNA start point on promoter function. *Nucleic Acids Res* 13(11): 4085–4096.
6. Sørensen KI, Baker KE, Kelln RA, Neuhard J (1993) Nucleotide pool-sensitive selection of the transcriptional start site *in vivo* at the *Salmonella typhimurium pyrC* and *pyrD* promoters. *J Bacteriol* 175(13):4137–4144.
7. Jeong W, Kang C (1994) Start site selection at *lacUV5* promoter affected by the sequence context around the initiation sites. *Nucleic Acids Res* 22(22):4667–4672.
8. Liu J, Turnbough CL, Jr (1994) Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the *pyrC* promoter in *Escherichia coli*. *J Bacteriol* 176(10):2938–2945.
9. Walker KA, Osuna R (2002) Factors affecting start site selection at the *Escherichia coli fis* promoter. *J Bacteriol* 184(17):4783–4791.
10. Lewis DE, Adhya S (2004) Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol Microbiol* 54(3):692–701.
11. Vvedenskaya IO, et al. (2015) Massively systematic transcript end readout, "MASTER": Transcription start site selection, transcriptional slippage, and transcript yields. *Mol Cell* 60(6):953–965.
12. Robb NC, et al. (2013) The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: Implications for transcription start-site selection. *J Mol Biol* 425(5):875–885.
13. Winkelman JT, et al. (2016) Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* 351(6277):1090–1093.
14. Winkelman JT, Chandrangsu P, Ross W, Gourse RL (2016) Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proc Natl Acad Sci USA* 113(13):E1787–E1795.
15. Zhang Y, et al. (2012) Structural basis of transcription initiation. *Science* 338(6110): 1076–1080.
16. Vvedenskaya IO, et al. (2014) Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* 344(6189):1285–1289.
17. Maitra U, Hurwitz H (1965) The role of DNA in RNA synthesis, IX. Nucleoside triphosphate termini in RNA polymerase products. *Proc Natl Acad Sci USA* 54(3): 815–822.
18. Jorgensen SE, Buch LB, Nierlich DP (1969) Nucleoside triphosphate termini from RNA synthesized in vivo by *Escherichia coli*. *Science* 164(3883):1067–1070.
19. Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* 11(8):2237–2255.
20. Shultzaberger RK, Chen Z, Lewis KA, Schneider TD (2007) Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res* 35(3):771–788.
21. Wilson HR, Archer CD, Liu JK, Turnbough CL, Jr (1992) Translational control of *pyrC* expression mediated by nucleotide-sensitive selection of transcriptional start sites in *Escherichia coli*. *J Bacteriol* 174(2):514–524.
22. Qi F, Turnbough CL, Jr (1995) Regulation of *codBA* operon expression in *Escherichia coli* by UTP-dependent reiterative transcription and UTP-sensitive transcriptional start site switching. *J Mol Biol* 254(4):552–565.
23. Tu AH, Turnbough CL, Jr (1997) Regulation of *upp* expression in *Escherichia coli* by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription. *J Bacteriol* 179(21):6665–6673.
24. Walker KA, Mallik P, Pratt TS, Osuna R (2004) The *Escherichia coli* Fis promoter is regulated by changes in the levels of its transcription initiation nucleotide CTP. *J Biol Chem* 279(49):50818–50828.
25. Turnbough CL, Jr (2008) Regulation of bacterial gene expression by the NTP substrates of transcription initiation. *Mol Microbiol* 69(1):10–14.
26. Turnbough CL, Jr, Switzer RL (2008) Regulation of pyrimidine biosynthetic gene expression in bacteria: Repression without repressors. *Microbiol Mol Biol Rev* 72(2):266–300.
27. Tang H, et al. (1994) Location, structure, and function of the target of a transcriptional activator protein. *Genes Dev* 8(24):3058–3067.
28. Mukhopadhyay J, et al. (2003) Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. *Methods Enzymol* 371: 144–159.
29. Goldman SR, et al. (2011) NanoRNAs prime transcription initiation in vivo. *Mol Cell* 42(6):817–825.
30. Vvedenskaya IO, Goldman SR, Nickels BE (2015) Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5′ ends. *Methods Mol Biol* 1276:211–228.
31. Druzhinin SY, et al. (2015) A conserved pattern of primer-dependent transcription initiation in *Escherichia coli* and *Vibrio cholerae* revealed by 5′ RNA-seq. *PLoS Genet* 11(7):e1005348.
32. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
33. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6):1188–1190.
34. Severinov K, Mooney R, Darst SA, Landick R (1997) Tethering of the large subunits of *Escherichia coli* RNA polymerase. *J Biol Chem* 272(39):24137–24140.
35. Vvedenskaya IO, et al. (2012) Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes Dev* 26(13):1498–1507.