

# **XACT-seq comprehensively defines the promoter-position and promoter-sequence determinants for initial-transcription pausing**

Jared T. Winkelman<sup>1,2,3,†</sup>, Chirangini Pukhrambam<sup>1,3,†</sup>, Irina O. Vvedenskaya<sup>1,3</sup>, Yuanchao Zhang<sup>4</sup>, Deanne M. Taylor<sup>4</sup>, Premal Shah<sup>1</sup>, Richard H. Ebright<sup>2,3\*</sup>, Bryce E. Nickels<sup>1,3\*</sup>.

<sup>1</sup>Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.

<sup>2</sup>Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA.

<sup>3</sup>Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA.

<sup>4</sup>Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA 19041, USA.

<sup>†</sup>Equal contribution

\*Corresponding authors: [ebright@waksman.rutgers.edu](mailto:ebright@waksman.rutgers.edu), [bnickels@waksman.rutgers.edu](mailto:bnickels@waksman.rutgers.edu)

## Abstract

Pausing by RNA polymerase (RNAP) during transcription elongation, in which a translocating RNAP uses a “stepping” mechanism, has been studied extensively, but pausing by RNAP during initial transcription, in which a promoter-anchored RNAP uses a “scrunching” mechanism, has not. We report a method that directly defines RNAP-active-center position relative to DNA *in vivo* with single-nucleotide resolution (XACT-seq; crosslink-between-active-center-and-template sequencing). We apply this method to detect and quantify pausing in initial transcription at  $4^{11}$  (~4,000,000) promoter sequences *in vivo*, in *Escherichia coli*. The results show initial-transcription pausing can occur in each nucleotide addition during initial transcription, particularly the first 4-5 nucleotide additions. The results further show initial-transcription pausing occurs at sequences that resemble the consensus sequence element for transcription-elongation pausing. Our findings define the positional and sequence determinants for initial-transcription pausing and establish initial-transcription pausing is hard-coded by sequence elements similar to those for transcription-elongation pausing.

## Introduction

RNA polymerase (RNAP) initiates transcription by binding double-stranded promoter DNA, unwinding a turn of promoter DNA to yield an RNAP-promoter open complex (RPo) containing a ~13 base pair single-stranded “transcription bubble” (Figure 1A) and selecting a transcription start site (TSS) (Ruff et al., 2015; Mazumder and Kapanidis, 2019; Winkelman et al., 2019). TSS selection entails placement of the start-site nucleotide (position +1) and the next nucleotide (position +2) of the template DNA strand into the RNAP-active-center product site (“P-site”) and addition site (“A-site”), respectively, and binding an initiating entity in the P-site and an extending NTP in the A-site.

The first ~10 nucleotides (nt) of an RNA product are synthesized as an RNAP-promoter initial transcribing complex (ITC) in which RNAP remains anchored on promoter DNA through sequence-specific protein-DNA interactions (Figure 1A; Mazumder and Kapanidis, 2019; Winkelman et al., 2019). Initial transcription starts with phosphodiester bond formation between the initiating entity and the extending NTP, to yield an initial RNA product (Figure 1A; ITC,2). Each nucleotide-addition cycle after formation of the initial RNA product requires translocation of the RNAP active center relative to DNA and RNA, starting from a “pre-translocated” state, and yielding a “post-translocated” state (Figure 1A,B; Erie et al., 1992; Zhang and Landick, 2009; Larson et al., 2011; Belogurov and Artsimovitch, 2019). Translocation of the RNAP active center repositions the RNA 3' nucleotide from the RNAP-active-center “A-site” to the RNAP-active-center “P-site,” rendering the A-site available to bind the next extending NTP. Initial transcription proceeds until synthesis of an RNA product of a threshold length of ~10 nt (Mazumder and Kapanidis, 2019; Winkelman et al., 2019). Upon synthesis of a threshold-length RNA product, RNAP breaks the sequence-specific protein-DNA interactions that anchor it on

promoter DNA, escapes the promoter, and synthesizes the rest of the RNA product as a transcription elongation complex (TEC; Figure 1B).

There are two clear differences in the mechanism of initial transcription (performed by ITC) and transcription elongation (performed by TEC): (i) a different mechanism of RNAP active center translocation, and (ii) a different RNAP subunit composition (Figure 1).

The first difference in the mechanisms of initial transcription and transcription elongation is a consequence of the sequence-specific protein-DNA interactions that anchor RNAP on promoter DNA in initial transcription, but not in transcription elongation (Larson et al., 2011; Belogurov and Artsimovitch, 2019; Mazumder and Kapanidis, 2019; Winkelman et al., 2019). These sequence-specific protein-DNA interactions prevent RNAP from moving relative to DNA in initial transcription, but not in transcription elongation. Therefore, RNAP uses a different mechanism of RNAP active center translocation in initial transcription vs transcription elongation (Larson et al., 2011; Belogurov and Artsimovitch, 2019; Mazumder and Kapanidis, 2019; Winkelman et al., 2019). In initial transcription, RNAP uses a “scrunching” mechanism, in which, in each nucleotide-addition cycle, RNAP remains anchored to promoter DNA, unwinds one base pair of DNA downstream of the RNAP active center, pulls the unwound single-stranded DNA (ssDNA) into and past the RNAP active center, and accommodates the additional unwound ssDNA as bulges in the transcription bubble (Figure 1A; Kapanidis et al., 2006; Margeat et al., 2006; Revyakin et al., 2006). In contrast, in transcription elongation, RNAP uses a “stepping” mechanism, in which, in each nucleotide-addition cycle, RNAP steps forward by one base pair relative to the DNA (Figure 1B; Abbondanzieri et al., 2005).

The second difference in the mechanisms of initial transcription and transcription elongation is a consequence of the fact that initial transcription is carried out by a

macromolecular assembly containing the initiation factor  $\sigma$  (Figure 1A), whereas transcription elongation is typically carried out by a macromolecular assembly lacking  $\sigma$  (Figure 1B; Mooney et al., 2005; Larson et al., 2011; Belogurov and Artsimovitch, 2019; Mazumder and Kapanidis, 2019; Winkelman et al., 2019). In the ITC, a module of  $\sigma$  referred to as the “ $\sigma$  finger,” reaches into the RNAP-active-center cleft and interacts with template-strand ssDNA in the transcription bubble close to the RNAP active center (Figure 1A; Severinov et al., 1994; Murakami et al., 2002; Kulbachinskiy and Mustaev, 2006; Zhang et al., 2012; Basu et al., 2014). These interactions pre-organize template-strand ssDNA to engage the RNAP active center, thereby facilitating the binding of initiating and extending NTPs (Murakami et al., 2002; Kulbachinskiy and Mustaev, 2006; Zhang et al., 2012; Pupov et al., 2014). Interactions of the  $\sigma$  finger with template-strand ssDNA places the  $\sigma$  finger in the path that will be occupied by nascent RNA upon RNA extension (Murakami et al., 2002; Kulbachinskiy and Mustaev, 2006; Zhang et al., 2012; Basu et al., 2014; Pupov et al., 2014; Zuo and Steitz, 2015). Thus, interactions of the  $\sigma$  finger with template-strand ssDNA presumably create an obstacle that must be overcome during initial transcription. Specifically, when the RNA reaches a length of  $\sim 4$  to  $\sim 5$  nt (after  $\sim 2$  to  $\sim 3$  RNAP-active-center translocation steps), the RNA 5' end appears to contact, and make favorable interactions with, the  $\sigma$  finger, but, when the RNA reaches a length of  $\sim 5$  to  $\sim 7$  nt (after  $\sim 3$  to  $\sim 5$  RNAP-active-center translocation steps), the RNA 5' end is expected to collide with, and clash with, the  $\sigma$  finger (Figure 1A; Murakami et al., 2002; Kulbachinskiy and Mustaev, 2006; Zhang et al., 2012; Basu et al., 2014; Pupov et al., 2014; Zuo and Steitz, 2015). The occurrence at one point of a favorable interaction with the  $\sigma$  finger (after  $\sim 2$  to  $\sim 3$  RNAP-active-center translocation steps), and subsequently a clash with the  $\sigma$  finger (after  $\sim 3$  to  $\sim 5$  RNAP-active-center translocation steps), is expected to have position-specific effects on RNAP

active center translocation during initial transcription. In contrast, in transcription elongation, where  $\sigma$  typically is absent (Mooney et al., 2005; Belogurov and Artsimovitch, 2019), no such position-specific effects on RNAP-active-center translocation are expected to occur.

In both initial transcription and transcription elongation, in the presence of saturating NTPs, each nucleotide-addition cycle takes, on average,  $\sim 20$  ms (Belogurov and Artsimovitch, 2019). In both initial transcription and transcription elongation, pauses--nucleotide-addition cycles that occur on the second or longer timescales--are off-pathway states that potentially modulate gene-expression levels (Larson et al., 2011; Belogurov and Artsimovitch, 2019; Kang et al., 2019; Mazumder and Kapanidis, 2019). Transcription-elongation pausing has been the subject of extensive analysis, both *in vitro* and *in vivo* (Larson et al., 2011; Belogurov and Artsimovitch, 2019; Kang et al., 2019). The results establish that transcription-elongation pausing is determined by the sequence of the DNA template, rather than TEC position relative to the TSS or the length of RNA in the TEC. For *Escherichia coli* RNAP, both *in vitro* and *in vivo*, transcription-elongation pausing occurs at DNA sequence elements that have the consensus sequence  $G_{-10}N_{-9}N_{-8}N_{-7}N_{-6}N_{-5}N_{-4}N_{-3}N_{-2}Y_{-1}G_{+1}$ , where Y is a pyrimidine and  $Y_{-1}$  corresponds to the position of the RNA 3' end (Herbert et al., 2006; Larson et al., 2014; Vvedenskaya et al., 2014; Imashimizu et al., 2015; see also Kang et al., 2019; Saba et al., 2019). In contrast to transcription-elongation pausing, initial-transcription pausing has been subject to only limited studies of two promoters *in vitro* (Duchi et al., 2016; Lerner et al., 2016; Dulin et al., 2018) and no studies *in vivo*. As a result, it is not known whether initial-transcription pausing occurs *in vivo*, it is not known what fraction of promoter sequences undergo initial-transcription pausing, the promoter-position dependence of initial-transcription pausing has not been defined, and the promoter-sequence determinants for initial-transcription pausing have not been defined.

## RESULTS

### Rationale

During RNA synthesis, the dwell time of the RNAP active center at each transcribed-region position, “RNAP occupancy,” is correlated with the tendency of RNAP to pause at that position. Accordingly, pause sites can be identified using methods that provide a measure of RNAP occupancy (Figure 2). Previously reported sequencing-based methods to monitor RNAP occupancies, such as native elongating transcript sequencing (NET-seq), rely on identifying and quantifying RNA 3' ends (Figure 2A; Churchman and Weissman, 2011; Churchman and Weissman, 2012; Larson et al., 2014; Vvedenskaya et al., 2014; Imashimizu et al., 2015). The identities of RNA 3' ends allow estimation of RNAP-active-center positions--subject to uncertainties, due to ability of RNAP to sample pre-translocated, post-translocated, reverse-translocated, and hyper-translocated states (Figure 2A,B; Larson et al., 2011; Belogurov and Artsimovitch, 2019)--and the quantities of RNA 3' ends allow estimation of RNAP-active-center dwell times. Unfortunately, however, these methods provide only an indirect measure of RNAP-active-center positions relative to DNA (Figure 2A), and because these methods cannot be applied to RNAs less than ~15-nt in length (Figure S1), these methods are not suitable for analysis of initial transcription. Methods combining chromatin immunoprecipitation (ChIP) with sequencing, such as ChIP-seq and ChIP-exo, are alternative sequencing-based methods of mapping RNAP relative to DNA (Barski et al., 2007; Rhee and Pugh, 2012; Srivastava et al., 2013; Latif et al., 2018). However, these methods provide insufficient resolution for most purposes and define overall RNAP boundaries relative to DNA rather than RNAP-active-center A-site positions.

Here we describe a sequencing-based method to monitor RNAP occupancy that overcomes these limitations, providing a direct, single-nucleotide-resolution readout of RNAP-active-center position relative to DNA (Figure 2B), and that is suitable for analysis of initial transcription as well as transcription elongation. The method entails formation of transcription complexes *in vitro* or *in vivo* using an RNAP derivative that has a photo-activatable crosslinking agent incorporated at a single, defined site in RNAP that, upon photo-activation *in vitro* or *in vivo*, forms covalent crosslinks with DNA at a defined position relative to the RNAP-active-center A-site; photo-activation to initiate covalent crosslinking of RNAP to DNA *in vitro* or *in vivo*; and high-throughput sequencing of primer extension products to define crosslink positions and crosslink yields. We term this method:

“crosslink-between-active-center-and-template sequencing” (XACT-seq; Figure 2B).

XACT-seq takes advantage of an RNAP derivative that has a photo-activatable crosslinking amino acid *p*-benzoyl-L-phenylalanine (Bpa) incorporated at RNAP-β' subunit residue R1148 (RNAP-β' <sup>R1148Bpa</sup>), which, upon photo-activation *in vitro* or *in vivo*, forms covalent crosslinks with DNA at a position exactly 5-nt downstream of the RNAP-active-center A-site (Figure 2B; Yu et al., 2017). In previous work, this reagent was used for structural analysis of static, trapped transcription complexes (Winkelman et al., 2015; Winkelman et al., 2016a; Yu et al., 2017). In this work, we apply the reagent for analysis of actively-transcribing complexes.

**RNAP-active-center A-site positions in initial-transcription pausing at the *lac*CONS promoter *in vitro* and *in vivo***



To demonstrate that the RNAP derivative that underpins XACT-seq, RNAP- $\beta'^{R1148Bpa}$ , enables detection of RNAP-active-center A-site positions in actively-transcribing complexes, and enables detection of initial-transcription pausing, we used RNAP- $\beta'^{R1148Bpa}$  to analyze actively-transcribing complexes engaged in initial-transcription pausing (Figure 3). We analyzed initial-transcription pausing at the *lac*CONS promoter (*plac*CONS; Figure 3A, top), the best-characterized example of initial-transcription pausing (Duchi et al., 2016; Lerner et al., 2016; Mazumder and Kapanidis, 2019). Lerner *et al.* and Duchi *et al.* showed *in vitro* that initial-transcription pausing occurs at *plac*CONS after ~4 to ~6 RNAP-active-center translocation steps (Duchi et al., 2016; Lerner et al., 2016). Duchi *et al.* and Dulin *et al.* further showed *in vitro* that this pause is reduced upon deletion of the  $\sigma$  finger (Duchi et al., 2016) and is increased upon substitution of the RNAP  $\beta$ -subunit residue 446 ( $\beta^{D446A}$ ) (Dulin et al., 2018), a substitution that alters the sequence dependence of RNAP-active-center translocation behavior (Zhang et al., 2012; Vvedenskaya et al., 2014).

We first used RNAP- $\beta'^{R1148Bpa}$  to detect initial-transcription pausing at *plac*CONS *in vitro* (Figure 3A). The results in Figure 3A, lanes 5, 8, and 11, confirm the previously reported ability of the RNAP- $\beta'^{R1148Bpa}$  to detect the RNAP-active-center A-site position with single-nucleotide resolution in static initial-transcribing complexes in the absence of NTP substrates *in vitro* (Winkelman et al., 2015; Winkelman et al., 2016a; Winkelman et al., 2016b; Yu et al., 2017). The results in Figure 3A, lanes 6, 9, and 12, demonstrate that the detected position of the RNAP-active-center A-site position does not change upon addition of a 2-nt RNA (ApA), indicating that the resulting ITC<sub>2</sub> is in a pre-translocated state (ITC<sub>2</sub>-pre). The results in Figure 3A, lane 7, demonstrate the ability of RNAP- $\beta'^{R1148Bpa}$  to detect RNAP-active-center A-site positions--and to detect initial-transcription pausing--in active initial-transcribing

complexes in the presence of all NTP substrates *in vitro*. We observe pausing in the initial-transcribed region (positions +3 to +9; Figure 3) and in the downstream transcribed-region sequence (positions  $\geq +10$ ). For the purpose of this paper we focus on the initial-transcribed region. Results for promoter occupancies during transcription elongation will be reported elsewhere.

We observe strong pausing at a position corresponding to exactly 4 RNAP-active-center translocation steps relative to RPitc,2-pre (i.e., when the RNAP-active-center A-site is at promoter position +6; Figure 3A). We observe that RNAP occupancy at this position decreases upon deletion of the  $\sigma$  finger (Figure 3A; compare lanes 7 and 10) and increases upon substitution of RNAP  $\beta$ -subunit residue 446 (Figure 3A; compare lanes 7 and 13). This pause exhibits the previously reported hallmarks of initial-transcription pausing at *placCONS*: i.e., pausing after  $\sim 4$  to  $\sim 6$  active-center translocation steps (Duchi et al., 2016; Lerner et al., 2016), a decrease in pausing upon deletion of the  $\sigma$  finger (Duchi et al., 2016), and an increase in pausing upon substitution of RNAP  $\beta$ -subunit residue 446 (Dulin et al., 2018). We conclude that the pausing that occurs when the RNAP-active-center A-site is at promoter position +6 represents initial-transcription pausing.

We next used RNAP- $\beta'$ <sup>R1148Bpa</sup> to detect initial-transcription pausing at *placCONS* *in vivo* (Figure 3B). We produced, in *Escherichia coli*, a Bpa-labeled, decahistidine-tagged RNAP- $\beta'$ <sup>R1148Bpa</sup> in the presence of unlabeled, untagged, wild-type RNAP, using a three-plasmid system (Figure 3B, top) comprising (i) a plasmid carrying a gene for RNAP  $\beta'$ -subunit containing a UAG codon at position 1148 and a decahistidine coding sequence; (ii) a plasmid carrying genes for an engineered Bpa-specific UAG-suppressor tRNA and an engineered Bpa-specific aminoacyl-tRNA synthetase; and (iii) a plasmid containing *placCONS*. We then grew cells in

medium containing Bpa, UV irradiated cells, lysed cells, purified crosslinked material using immobilized metal-ion-affinity chromatography (IMAC) targeting the decahistidine tag on RNAP- $\beta'^{R1148Bpa}$ , and mapped crosslinks using primer extension (Figure 3B, middle). In order to trap static initial-transcribing complexes containing 2- to 3-nt RNA products *in vivo*, despite the presence of all NTP substrates, we used a chemical biology approach, exploiting the RNAP inhibitor Rifampin (Rif), which blocks extension of RNA products beyond a length of 2 to 3 nt (McClure and Cech, 1978). The results showed exact correspondence of crosslinking patterns *in vitro* and *in vivo* for the initial-transcription pause at promoter position +6 (Figure 3B).

The results in Figure 3 establish that RNAP- $\beta'^{R1148Bpa}$  enables detection of the RNAP-active-center A-site position in actively-transcribing complexes. The results further establish that initial-transcription pausing at *placCONS* occurs *in vivo* and exhibits the same promoter-position dependence *in vivo* as *in vitro*. The results reveal the exact RNAP-active-center position in initial-transcription pausing at *placCONS*. In particular, the results show that the RNAP-active-center A-site during initial-transcription pausing at *placCONS* is located at promoter position +6 and thus, neglecting possible fractionally-translocated states, show that pausing occurs in either RPitc,5-post or RPitc,6-pre (Figure 3C).

### **Promoter-position dependence of initial-transcription pausing for a library of 4<sup>11</sup> (~4,000,000) promoters *in vivo***

Having validated RNAP- $\beta'^{R1148Bpa}$  as a reagent for detection of the RNAP-active-center A-site position in actively-transcribing complexes *in vitro* and *in vivo*, we next combined RNAP- $\beta'^{R1148Bpa}$  with a library of *placCONS* derivatives (*placCONS*-N11; Figure 4A)

containing all possible sequences from promoter position +3 to promoter position +13 ( $4^{11}$ ; ~4,000,000 sequences), in order to enable multiplexed analysis of initial-transcription pausing on all possible initial-transcribed region sequences *in vivo* (Figures 4, 5, S2, S3). We used a three-plasmid system analogous to that used in the previous section but having representatives of the *placCONS*-N11 library of plasmids instead of the plasmid carrying *placCONS*. We UV irradiated cells to initiate RNAP-DNA crosslinking, lysed cells, isolated crosslinked material, and mapped crosslinks using primer extension as in the previous section. We used the same chemical-biology approach, exploiting the fact that the RNAP inhibitor Rif prevents extension of RNA products beyond the length of 2 to 3 nt to trap static initial-transcribing complexes corresponding to RPitc,2 and RPitc,3 (Figure 4B).

In order to read out position-specific RNAP occupancies across the entire sampled initial-transcribed-region sequence space, we performed urea-PAGE of radiolabeled primer extension products (Figure 4B). In the presence of Rif, we see two major bands and a minor band across the sampled initial-transcribed-region sequence space at positions corresponding to RPitc,2; RPitc,3; and RPitc,4--i.e. when the RNAP-active-center A-site is at promoter positions +2, +3, or +4 (see Figure 1A)--exactly as above for the *placCONS* initial-transcribed-region sequence (compare Figure 4B, lane 5 with Figure 3B, lane 7). In the absence of Rif, we observe a series of additional bands corresponding to positions of initial-transcription pausing (Figure 4B, lane 6). The prominence of the additional bands indicates that initial-transcription pausing is a prominent feature of transcription across the entire sampled initial-transcribed-region sequence space. As compared to the experiments in the presence of Rif, in the absence of Rif, we observe higher levels of RNAP occupancy at positions corresponding to exactly 1, 2, 3, 4, 5, 6, and 7 RNAP-active-center translocation steps relative to RPitc,2-pre (i.e., when the

RNAP-active-center A-site is at promoter positions +3, +4, +5, +6, +7, +8, and +9; Figure 4B,C). RNAP occupancy levels are highest at the positions corresponding to 1 to 4 RNAP-active-center translocation steps (i.e., when the RNAP-active-center A-site is at promoter positions +3, +4, +5, and +6; Figure 4B,C) and are lower for the positions corresponding to 5, 6, or 7 RNAP-active-center translocation steps (i.e., when the RNAP-active-center A-site is at promoter positions +7, +8, and +9; Figure 4B,C). Similar results are observed in parallel experiments using a promoter library that randomizes 20 bp of the initial-transcribed region, from position +2 through +21 (Figure S2). We conclude that, across the sampled initial-transcribed-region sequence space, initial-transcription pausing occurs at a large fraction of promoters. We further conclude that, across the sampled initial-transcribed-region sequence space, initial-transcription pausing can occur at any promoter initial-transcribed-region position from position +3 through position +9, with highest levels for promoter positions +3 through +6.

### **Promoter-sequence determinants for initial-transcription pausing in a library of $4^{11}$ (~4,000,000) promoters *in vivo***

In order to quantify the promoter position dependence of initial-transcription pausing, and in order to define the promoter sequence determinants for initial-transcription pausing, for the library of  $4^{11}$  (~4,000,000) initial-transcribed-region sequences, we performed the full XACT-seq protocol (Figure S3). We used the same three-plasmid system and the same procedure of UV irradiation, cell lysis, purification of crosslinked material, and primer extension as in the previous sections. We then used high-throughput sequencing to identify and quantify primer extension products. We identified RNAP-active-center A-site positions based on the identities of 3' ends of primer extension products, we quantified RNAP occupancies at these

positions based on numbers of sequence reads, and we assigned sites of initial-transcription pausing as sites with above-threshold RNAP occupancies.

The results confirm and quantify the finding in the preceding section that, across the sampled initial-transcribed-region sequence space, a substantial fraction of promoter initial-transcribed region sequences--~20% of promoter sequences--show initial-transcription pausing (Table S1). The results further confirm and quantify the finding in the preceding section that, across the sampled initial-transcribed-region sequence space, initial-transcription pausing can occur at each promoter initial-transcribed-region position, from position +3 through position +9, with highest levels occurring when the RNAP-active-center A-site is at positions +3 through +7, and lower levels when the RNAP-active-center A-site is at positions +8 and +9 (Table S1).

A global alignment of all initial-transcribed-region sequences showing initial-transcription pausing--aligning by the RNAP-active-center A-site position--yields a clear consensus sequence:  $T_{P-1}N_P\underline{Y}_AG_{A+1}$ , where Y is pyrimidine and  $\underline{Y}_A$  is the position of the RNAP-active-center A-site (Figure 5A,B). A plot of RNAP occupancy as a function of initial-transcribed-region tetranucleotide sequence shows graphically that RNAP occupancy is strongly correlated with match to the consensus sequence (Figure 5B). Separate alignments of initial-transcribed-region sequences showing pausing for each promoter position from +3 through +9, yield, in all cases, a consensus sequence resembling the global alignment (Figures 5C, S4). Plots of RNAP occupancy as a function of initial-transcribed-region sequence shows graphically that RNAP occupancy strongly correlates with match to the consensus sequence for each position (Figure S4). We note that *placCONS*, the promoter previously shown to exhibit initial-transcription pausing at a position in the range +5 to +7 *in vitro* (Duchi et al., 2016; Lerner et al., 2016; Dulin et al., 2018), and shown here to exhibit initial transcription pausing at position

+6 *in vitro* and *in vivo* (Figure 3), contains a 3-of-3 match to the consensus sequence in the register that would yield initial transcription pausing at position +6 (compare Figure 3A and Figure 5A).

Representative individual initial-transcribed-region sequences showing initial-transcription pausing were analyzed for pausing individually, both *in vitro* and *in vivo*, and showed clear pausing at the expected positions *in vitro* and *in vivo* (Figures 5D-F, S5). The staircase pattern in Figure 5E (*in vitro*) and Figure S5B (*in vivo*) shows graphically that initial-transcription pausing can occur at each promoter position in the tested range.

We performed analogous experiments with individual initial-transcribed-region sequences using an RNAP holoenzyme derivative containing a deletion of the  $\sigma$  finger (Figure 5E,F). Qualitatively, the results for wild-type RNAP holoenzyme and the RNAP holoenzyme derivative containing the  $\sigma$  finger deletion were identical for individual initial-transcribed-region sequences (Figure 5E,F), indicating that sequence, rather than position-dependent collision with the  $\sigma$  finger, is the crucial determinant for initial-transcription pausing. Quantitatively, however, we observe reductions in RNAP occupancies for the RNAP holoenzyme containing the  $\sigma$  finger, relative to wild-type RNAP holoenzyme, for promoter positions +5 and +6, but not for promoter positions +7, +8 and +9 (Figure 5E,F). We infer that the  $\sigma$  finger contributes quantitatively to pausing at promoter positions +5 and +6, which are positions where collision of the RNA 5' end and the  $\sigma$  finger potentially may occur (Murakami et al., 2002; Kulbachinskiy and Mustaev, 2006; Zhang et al., 2012; Basu et al., 2014; Pupov et al., 2014; Zuo and Steitz, 2015). Considering the results in Figure 5, as a whole, we conclude that sequence is the crucial determinant of initial-transcription pausing. We further conclude that sequence is the crucial determinant of initial-transcription pausing at each promoter position from position +3 to

position +9, and that the consensus sequences for initial-transcription pausing is similar or identical at each position.

## **Relationship between sequence determinants for initial-transcription pausing and sequence determinants for transcription-elongation pausing**

Our results define the consensus sequence for initial-transcription pausing as:  $T_{P-1}N_P\underline{Y}_A G_{A+1}$  (Figure 6A, top). This is both the global consensus sequence for initial-transcription pausing irrespective of position and, in most, and possibly all, cases the position-specific sequence for initial-transcription pausing at each initial-transcribed region position (Figure 5A-C).

Three groups previously have used NET-seq to define the consensus sequence for transcription-elongation pausing:  $G_{-10}N_{-9}N_{-8}N_{-7}N_{-6}N_{-5}N_{-4}N_{-3}N_{-2}Y_{-1}G_{+1}$ , where  $Y_{-1}$  is the position of the RNA 3' end (Figure 6A, middle; Larson et al., 2014; Vvedenskaya et al., 2014; Imashimizu et al., 2015). Mapping of TEC translational state by analysis of pyrophosphorolysis kinetics (Vvedenskaya et al., 2014) and by cryo-EM structure determination (Guo et al., 2018; Kang et al., 2018) indicates that TECs paused at the consensus sequence element for transcription-elongation pausing assume a pre-translocated or fractionally-translocated state. Accordingly, the consensus sequence for transcription-elongation pausing can be expressed as:  $G_{P-8}N_{P-7}N_{P-6}N_{P-5}N_{P-4}N_{P-3}N_{P-2}N_{P-1}N_P\underline{Y}_A G_{A+1}$ , where  $\underline{Y}_A$  is the position of the RNAP-active-center A-site (Figure 6A, bottom). The consensus sequence for transcription elongation pausing,  $G_{P-8}N_{P-7}N_{P-6}N_{P-5}N_{P-4}N_{P-3}N_{P-2}N_{P-1}N_P\underline{Y}_A G_{A+1}$ , comprises a highly-conserved downstream segment,  $\underline{Y}_A G_{A+1}$  (Larson et al., 2014; Vvedenskaya et al., 2014; Imashimizu et al., 2015). In the paused TEC, the  $G_{A+1}:C_{A+1}$  nucleotide pair is located at the



downstream end of the transcription bubble and must be broken for RNAP forward translocation, and the DNA-template-strand  $\underline{R}_A$  interacts with the RNAP active center (Figure 6B,C; Vvedenskaya et al., 2014; see also Saba et al., 2019); the identities of nucleotides at these positions affects RNAP-translocation behavior through sequence-dependent effects on DNA duplex thermal stability and sequence-dependent effects on interactions of the DNA template strand with the RNAP active center (Vvedenskaya et al., 2014). The consensus sequence element for transcription-elongation pausing also comprises a less-highly-conserved upstream segment,  $G_{P-8}$  (Larson et al., 2014; Vvedenskaya et al., 2014; Imashimizu et al., 2015). In the paused TEC, the  $G_{P-8}:C_{P-8}$  RNA:DNA base pair is located at the upstream end of the RNA-DNA hybrid and must be broken for RNAP translocation (Figure 6B; Vvedenskaya et al., 2014); sequence at these positions affects RNAP-translocation behavior through effects on RNA-DNA duplex thermal stability (Vvedenskaya et al., 2014).

The consensus sequence for initial-transcription pausing defined in this work:  $T_{P-1}N_P\underline{Y}_AG_{A+1}$  (Figure 6A, top), exhibits a striking resemblance to the downstream, most-highly-conserved portion of the consensus sequence for transcription-elongation pausing:  $\underline{Y}_AG_{A+1}$  (Figure 6A, bottom). This resemblance extends both to the sequence and the position of the RNAP-active-center A-site relative to sequence. Consistent with the conclusion that the consensus sequence for initial-transcription pausing:  $T_{P-1}N_P\underline{Y}_AG_{A+1}$ , may be functionally related to the downstream, most-highly-conserved portion of the consensus sequence for transcription-elongation pausing,  $\underline{Y}_AG_{A+1}$ , the consensus sequence for initial-transcription pausing contains T at position  $P-2$  (Figure 6A) and two groups defining the consensus sequence for transcription-elongation pausing identified T as the optimal nucleotide at the corresponding position (Larson et al., 2014; Imashimizu et al., 2015). The observation that the consensus

sequence for initial-transcription pausing does not contain a sequence corresponding to the upstream, less-highly-conserved portion of the consensus sequence for elongation,  $G_{P-8}$ , is explained by the fact that the analyzed ITCs contain short RNA products and short RNA:DNA hybrids that do not extend to this upstream position.

We suggest that the resemblance of the consensus sequence for initial-transcription pausing and the downstream portion of the consensus sequence for transcription-elongation pausing is real and reflects a functional relationship. We infer that initial-transcription pausing and transcription-elongation pausing share mechanistic commonalities. Consistent with this proposal, substitution of the RNAP  $\beta$ -subunit residue 446 ( $\beta^{D446A}$ ), increases sequence-dependent pausing in both initial transcription (Figure 3A; Dulin et al., 2018) and transcription elongation (Vvedenskaya et al., 2014).

## Prospect

XACT-seq (crosslink-between-active-center-and-template sequencing) provides a high-throughput, direct, single-nucleotide-resolution readout of RNAP-active-center A-site position relative to DNA during transcription initiation and transcription elongation *in vivo*, in living cells. In this work, using the reagent that underpins XACT-seq, an RNAP derivative which, upon photo-activation *in vitro* or *in vivo*, forms covalent crosslinks with DNA at a position exactly 5-nt downstream of the RNAP-active-center A-site we: (i) defined the RNAP-active-center A-site position at *placCONS* (Figure 3), (ii) demonstrated that initial-transcription pausing occurs at a substantial fraction of promoter initial-transcribed region sequences *in vivo* (Figures 4, S2), (iii) showed that initial-transcription pausing can occur at each promoter position in the promoter-initial-transcribed region from +3 to +9 (Figures 4, S2), (iv)

and showed that the  $\sigma$  finger contributes quantitatively to pausing at promoter positions +5 and +6, presumably through collision with the RNA 5' end (Figures 3, 5E,F). Next, using XACT-seq and sampling a library of  $4^{11}$  (~4,000,000) promoter initial-transcribed-region sequences, we: (i) confirmed that initial-transcription pausing occurs *in vivo* (Figures 5, S4; Table S1), (ii) confirmed and quantified that initial-transcription pausing occurs at a substantial fraction (at least ~20%) of initial-transcribed-region sequences (Figure 5C; Table S1), (iii) confirmed and quantified that initial-transcription pausing can occur at each position in the initial-transcribed-region sequence (high prevalence for positions +3 to +7; lower prevalence for positions +8 and +9; Figure 5C; Table S1), (iv) showed that initial-transcription pausing is determined primarily by promoter sequence (Figures 5, S4), (v) defined the consensus sequence for initial-transcription pausing as  $T_{P-1}N_P Y_A G_{A+1}$  (Figure 5A), and (vi) showed that the consensus sequence for initial-transcription pausing resembles the consensus sequence for transcription-elongation pausing, suggesting that initial-transcription pausing and transcription-elongation pausing share mechanistic commonalities (Figure 6). A major finding is that initial-transcription pausing and transcription-elongation pausing appear to be fundamentally similar, with the key mechanistic difference, a “scrunching” mechanism of RNAP-active-center translocation in the former vs. a “stepping” mechanism of RNAP-active-center translocation in the latter, having no detectable effect, and the key structural difference between an ITC and a TEC, namely, the presence of the initiation factor  $\sigma$  in the former but not the latter, having only a quantitative, modulatory effect on initial-transcription pausing.

Here we show XACT-seq overcomes the limitations of methods that infer the RNAP-active-center A-site position from the identity of the RNA 3' end, a process that, as described above, is subject to uncertainties due to translocation equilibration of TECs among

pre-translocated, post-translocated, reverse-translocated, and hyper-translocated states. We further show XACT-seq overcomes the limitations of methods such as ChIP-seq and ChIP-exo that can report RNAP position, but only with low resolution, and which define the overall RNAP position but not the RNAP-active-center A-site position.

We applied XACT-seq to analysis of a library of 4<sup>11</sup> (~4,000,000) initial-transcribed region sequences. In this report, we focused on results for initial-transcription pausing. However, the same experiments also provide information for pausing in subsequent stages of transcription, including pausing at the moment of promoter escape and formation of a TEC, promoter-proximal  $\sigma$ -dependent pausing by TECs, and sequence-dependent elongation pausing by TECs. The findings for these subsequent stages of transcription will be reported elsewhere.

The approach that was applied here to a library of promoter sequences can be applied to genome-wide analysis of RNAP-active-center A-site positions *ex vivo* on isolated DNA or chromatin, or *in vivo*, in living cells, providing a basis for “RNAP profiling” analogous to existing methods for ribosome profiling (Ingolia et al., 2009). Combining XACT-seq and NET-seq for genome-wide analysis of RNAP-active-center A-site positions and RNA 3' end positions, respectively, each with single-nucleotide resolution, should provide an unprecedentedly rich description of the sequence-dependent, factor-dependent transcriptional landscape.

## **Acknowledgements**

Work was supported by NIH grants GM124976 (PS), GM041376 (RHE), and GM118059 (BEN). JTW was supported by a fellowship from the Helen Hay Whitney Foundation.

## **Author Contributions**

Conceptualization, JTW, RHE, BEN

Methodology, JTW, CP, YZ, IOV, PS

Formal Analysis, JTW, CP, PS, YZ, DT

Investigation, JTW, CP, IOV

Writing, JTW, RHE, BEN

Visualization, JTW, RHE, BEN

Supervision, DT, RHE, BEN

Project Administration, RHE, BEN

Funding Acquisition, PS, RHE, BEN

## METHODS

### Plasmids

To generate plasmid pIA900- $\beta$ D446A RNAP- $\beta'$ <sup>R1148Bpa</sup>, we performed PCR using ~20 ng of pIA900-RNAP- $\beta'$ <sup>R1148Bpa</sup> (Winkelman et al., 2015) in 12.5  $\mu$ l reactions containing 0.8  $\mu$ M oligo JW44 and 1 X Phusion HF Master Mix (Thermo Fischer Scientific) (95°C for 2 min; 95°C for 15 s, 55°C for 15s, 72°C for 5 min; 30 cycles; 72°C for 6 min). Next, to remove the wild- plasmid template, 20 units of DpnI (New England Biolabs) was added, the reactions was incubated at 37°C for 16 h, 1  $\mu$ l of the reaction was introduced into electrocompetent DH10B cells (Thermo Fischer Scientific), and cells plated on LB agar plates containing 100  $\mu$ g/ml carbenicillin to select for transformants.

To generate plasmid p $\sigma$ <sup>70</sup>-His p $\sigma$ <sup>70</sup> $\Delta$  finger, we performed PCR using 4 ng of plasmid p $\sigma$ <sup>70</sup>-His (Marr and Roberts, 1997) in 50  $\mu$ l reactions containing 0.8  $\mu$ M oligo JW268, 0.8  $\mu$ M oligo JW269, and 1 X Phusion HF Master Mix (2 min at 95°C; 95°C for 15 s, 55°C for 15s, 72°C for 4 min (30 x); 2 min at 72°C). 30 units of DpnI was added, the reaction was incubated at 37°C for 16 h, and DNA was recovered using a PCR purification kit (Qiagen). The recovered DNA was treated with 10 units of T4 polynucleotide kinase (PNK) (New England Biolabs) in 1 X T4 PNK buffer containing 20  $\mu$ M ATP for 30 min at 37°C, 400 units of T4 DNA ligase (New England Biolabs) was added, reactions were incubated at 16°C for 16 h, 1  $\mu$ l of the reaction was introduced into electrocompetent DH10B cells, and cells plated on LB agar plates containing 50  $\mu$ g/ml kanamycin, and recombinant plasmid DNA was isolated from individual transformants.

Plasmid pCDF-CP (Yu et al., 2017) contains a CloDF13 replication origin, a selectable marker conferring resistance to spectinomycin, and two BglI recognition sites that are used to introduce DNA fragments upstream of transcription terminator tR2. Plasmid pCDF-*lac*CONS (Yu et al., 2017) is a derivative of pCDF-CP containing sequences from positions -88 to +70 of *plac*CONS inserted into BglI-digested pCDF-CP.

To generate plasmids pCDF-*lac*CONS-p5, -p6, -p7, -p8, and -p9, which contain the sequences between positions +3 and +13 shown in Figure 5D, we performed PCR using 1 ng of pCDF-*lac*CONS in 25  $\mu$ l reactions containing 1 X Phusion HF Master Mix, 0.8  $\mu$ M oligo JW80 and 0.8  $\mu$ M oligo JW433, JW408, JW409, JW580, or JW575, respectively. 1  $\mu$ l of this reaction was used as a template in PCR reactions (50  $\mu$ l) containing 1 X Phusion HF Master Mix, 0.8  $\mu$ M

JW79, and 0.8  $\mu$ M JW80. Amplicons were treated with BglI and BglI-digested fragments were ligated into BglI-digested pCDF-CP. The ligation mixture was transformed into NEB 5-alpha competent cells (New England Biolabs), cells plated on LB agar plates containing 50  $\mu$ g/ml spectinomycin, and recombinant plasmid DNA was isolated from individual transformants.

Plasmid-borne promoter libraries *plac*CONS-N11 and *plac*CONS-N20 were generated using a procedure described in (Vvedenskaya et al., 2015; Vvedenskaya et al., 2018b) that provides a “self-assembling barcode,” in which for each DNA molecule in the library a first randomized sequence in a region of interest is associated with a known corresponding second randomized sequence that serves as a barcode. The procedure involves synthesis of three oligos for use in PCR. One oligo, which serves as the template for PCR amplification, contains a first randomized sequence spanning the region of interest and a second randomized sequence that serves as a barcode. The other oligos serve as amplification primers. One of these oligos contains 5'-end sequences that introduce a BglI recognition sequence and the other oligo contains 3'-end sequences complementary to the template oligo. Libraries *plac*CONS-N11, containing up to 4<sup>11</sup> initial-transcribed region sequences from position +3 to position +13, was constructed using oligo JW153 as the template and oligos s1219 and s1220 as amplification primers. Library *plac*CONS-N20, containing up to 4<sup>20</sup> initial-transcribed region sequences from position +2 to position +21, was constructed using oligo JW203 as the template and oligos s1219 and s1220 as amplification primers. Amplicons were treated with BglI and BglI-digested fragments were ligated into BglI-digested pCDF-CP. The ligation mixture was transformed into NEB 5-alpha competent cells, cells plated on LB agar plates containing 50  $\mu$ g/ml spectinomycin, and recombinant plasmid DNA was isolated from  $\sim 10^7$  transformants.

Plasmid pEVOL-pBpF contains genes directing the synthesis of an engineered Bpa-specific UAG-suppressor tRNA and an engineered Bpa-specific aminoacyl-tRNA synthetase that charges the amber suppressor tRNA with Bpa (Addgene; Chin et al., 2002).

## Proteins

Bpa-containing RNAP core enzyme derivatives RNAP- $\beta'$ <sup>R1148Bpa</sup> and  $\beta$ D446A RNAP- $\beta'$ <sup>R1148Bpa</sup> were prepared from *E. coli* strain NiCo21(DE3) (New England Biolabs) containing plasmid pEVOL-pBpF and plasmid pIA900-RNAP- $\beta'$ <sup>R1148Bpa</sup> (Winkelman et al., 2015) or

plasmid pIA900- $\beta$ D446A RNAP- $\beta'$ <sup>R1148Bpa</sup>, using procedures described in Winkelman et al., 2015.

Wild-type  $\sigma^{70}$  or a  $\sigma^{70}$  derivative containing a deletion of the  $\sigma$  finger (residues 513-519), were prepared from *E. coli* strain NiCo21(DE3) containing plasmid p $\sigma^{70}$ -His (gift of J. Roberts; Marr and Roberts, 1997) or plasmid p $\sigma^{70}$ -His  $\Delta$  finger using procedures described in (Marr and Roberts, 1997). To form RNAP holoenzyme, 1  $\mu$ M RNAP core enzyme and 5  $\mu$ M wild-type  $\sigma^{70}$  or  $\Delta$  finger  $\sigma^{70}$  in 10 mM Tris-Cl (pH 8.0), 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol were incubated for 30 min at 25°C.

### Oligonucleotides

Oligodeoxyribonucleotides (Table S2) were purchased from IDT. Diribonucleotide ApA (HPLC-purified) was purchased from TriLink Biotechnologies.

### Templates for *in vitro* assays

Linear DNA templates used for *in vitro* transcription assays contain sequences from positions -88 to +70 of *plac*CONS (Figure 3A,B) or positions -88 to +70 of *plac*CONS-p5, *plac*CONS-p6, *plac*CONS-p7, *plac*CONS-p8, or *plac*CONS-p9 (Figure 5D-F). Templates were generated by PCR in reactions containing 1 X Phusion HF Master Mix, 0.8  $\mu$ M primer JW61, 0.8  $\mu$ M primer JW62, and ~1 pg of plasmids *plac*CONS-p5, -p6, -p7, -p8, or -p9. Reaction products were purified using a PCR purification kit.

### Determination of RNAP-active-center A-site positions by protein-DNA photo-crosslinking *in vitro*

*In vitro* photo-crosslinking and crosslink mapping experiments were done using procedures described in (Yu et al., 2017). For the experiments in Figure 3A,B, 50  $\mu$ l reactions containing 20 nM RNAP holoenzyme, 4 nM template, 100  $\mu$ M ApA dinucleotide (where indicated), and 1 X RB [10 mM Tris-Cl, pH 8.0, 70 mM NaCl, 10 mM MgCl<sub>2</sub>, and 0.1 mg/ml bovine serum albumin (BSA)] were incubated for 5 min at 25°C, 5  $\mu$ l of 100  $\mu$ M ATP, 100  $\mu$ M CTP, 100  $\mu$ M GTP, and 100  $\mu$ M UTP (final concentration of 9  $\mu$ M ATP, 9  $\mu$ M CTP, 9  $\mu$ M GTP, and 9  $\mu$ M UTP) were added (where indicated), reactions were incubated 2 min at 25°C, and subjected to UV irradiation for 2 min at 25°C in a Rayonet RPR-100 photochemical reactor



equipped with 16 x 350 nm tubes (Southern New England Ultraviolet). For the experiments in Figure 5C, 50  $\mu$ l reactions containing 20 nM RNAP holoenzyme, 4 nM template, and 1 X RB were incubated for 5 min at 25°C, 5  $\mu$ l of 1 mM ATP, 1 mM CTP, 1 mM GTP, and 1 mM UTP were added, reactions were incubated 2 min at 25°C, and subjected to UV irradiation for 2 min at 25°C. For the experiments in Figure 5D, 50  $\mu$ l reactions containing 50 nM RNAP holoenzyme, 20 nM template (4 nM of *plac*CONS-p5, -p6, -p7, -p8, and -p9), and 1 X RB were incubated for 5 min at 25°C, 5  $\mu$ l of 1 mM ATP, 1 mM CTP, 1 mM GTP, and 1 mM UTP was added, reactions were incubated 2 min at 25°C, and subjected to UV irradiation for 2 min at 25°C.

To denature RNAP-DNA complexes, reactions were mixed with 15  $\mu$ l 5 M NaCl and 6  $\mu$ l 100  $\mu$ g/ml heparin, incubated for 5 min at 95°C and then cooled to 4°C. Crosslinked RNAP-DNA complexes were isolated by adding 20  $\mu$ l MagneHis Ni-particles (Promega) equilibrated and suspended in 10 mM Tris-Cl, pH 8.0, 1.2 M NaCl, 10 mM MgCl<sub>2</sub>, 10  $\mu$ g/ml heparin, and 0.1 mg/ml BSA; MagneHis Ni-particles were collected using a magnetic microfuge tube rack; particles were washed with 50  $\mu$ l 10 mM Tris-Cl, pH 8.0, 1.2 M NaCl, 10 mM MgCl<sub>2</sub>, 10  $\mu$ g/ml heparin, and 0.1 mg/ml BSA, washed twice with 50  $\mu$ l 1 X *Taq* DNA polymerase buffer (New England Biolabs), and particles (which contained bound RNAP-DNA complexes) were resuspended in 10  $\mu$ l 1 X *Taq* DNA polymerase buffer. Primer extension reactions (12.5  $\mu$ l) were performed by combining 2  $\mu$ l of the recovered RNAP-DNA complexes, 1  $\mu$ l of 1  $\mu$ M <sup>32</sup>P-5'-end-labeled primer JW61 [200 Bq/fmol; prepared using [ $\gamma$ -<sup>32</sup>P]-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs) as described in (Sambrook et al., 2006)], 1  $\mu$ l 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM TTP, 0.5  $\mu$ l 5 U/ $\mu$ l *Taq* DNA polymerase (New England Biolabs), 5  $\mu$ l 5 M betaine, 0.625  $\mu$ l 100% dimethyl sulfoxide, and 1.25  $\mu$ l 10 X *Taq* DNA polymerase buffer; 40 cycles of 30 s at 95°C, 30 s at 53°C, and 30 s at 72°C. Reactions were stopped by addition of 12.5  $\mu$ l 1 X TBE, 8 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue; radiolabeled products were separated by electrophoresis on 10% 8M urea slab gels (equilibrated and run in 1 X TBE) and visualized by storage-phosphor imaging (Typhoon 9400 variable-mode imager; GE Life Science). Positions of RNAP-DNA crosslinking were determined by comparison to products of a DNA-nucleotide sequencing reaction generated using oligo JW61 and a linear DNA template containing sequences from positions -88 to +70 of *plac*CONS (Thermo Sequenase Cycle Sequencing Kit; Affymetrix).

## **Determination of RNAP-active-center A-site positions by protein-DNA photo-crosslinking *in vivo***

*In vivo* photo-crosslinking and crosslink mapping experiments of Figures 3B, 4B, and S5 were done essentially as described in (Yu et al., 2017) except the RNAP active site was not mutationally inactivated. Analysis of RNAP-active-center A-site positions *in vivo* for *placCONS* transcription complexes was performed by sequential introduction of plasmid pCDF-*lacCONS*, plasmid pIA900-RNAP- $\beta'$ <sup>R1148Bpa</sup>, and plasmid pEVOL-pBpF into *E. coli* strain NiCo21(DE3) by transformation. After the final transformation step, cells were plated on LB agar containing 100  $\mu$ g/ml carbenicillin, 50  $\mu$ g/ml spectinomycin, 50  $\mu$ g/ml streptomycin, and 25  $\mu$ g/ml chloramphenicol; at least 1,000 individual colonies were scraped from the plate, combined, and used to inoculate 250 ml LB broth containing 1 mM Bpa (Bachem), 100  $\mu$ g/ml carbenicillin, 50  $\mu$ g/ml spectinomycin, 50  $\mu$ g/ml streptomycin, and 25  $\mu$ g/ml chloramphenicol in a 1000 ml flask (Bellco) to yield OD<sub>600</sub> = 0.3; the culture was placed in the dark and shaken (220 rpm) for 1 h at 37°C; isopropyl- $\beta$ -D-thiogalactoside (IPTG) was added to 1 mM; and the culture was placed in the dark and shaken (220 rpm) for 3 h at 37°C.

Analysis of RNAP-active-center A-site positions *in vivo* for *placCONS*-N11 and *placCONS*-N20 transcription complexes was performed by sequential introduction into *E. coli* strain NiCo21(DE3) of the pCDF-*lacCONS*-N11 library (yielding ~2 million transformants) or the *placCONS*-N20 library (yielding ~2 million transformants), plasmid pIA900-RNAP- $\beta'$ <sup>R1148Bpa</sup> (yielding ~15 million transformants), and plasmid pEVOL-pBpF (yielding ~4 million transformants). After the final transformation step, cells were plated on ~10-15 LB agar plates containing 100  $\mu$ g/ml carbenicillin, 50  $\mu$ g/ml spectinomycin, 50  $\mu$ g/ml streptomycin, and 25  $\mu$ g/ml chloramphenicol to yield a lawn. Colonies were scraped from the surface of the plates, combined, and used to inoculate 250 ml LB broth containing 1 mM Bpa, 100  $\mu$ g/ml carbenicillin, 50  $\mu$ g/ml spectinomycin, 50  $\mu$ g/ml streptomycin, and 25  $\mu$ g/ml chloramphenicol in a 1000 ml flask to yield OD<sub>600</sub> = 0.3; the culture was placed in the dark and shaken (220 rpm) for 1 h at 37°C; IPTG was added to 1 mM; and the culture was placed in the dark and shaken (220 rpm) for 3 h at 37°C.

To enable analysis of static transcription complexes, a portion of the cell cultures containing pCDF-*lac*CONS, pCDF-*lac*CONS-N11, or pCDF-*lac*CONS-N20 were removed, Rifampin was added to a final concentration of 200 µg/ml, and the culture was shaken for 20 min prior to harvesting cells for UV irradiation. Cell suspensions (7 ml) were removed from each culture to a 13 mm x 100 mm borosilicate glass test tube (VWR) and subjected to UV irradiation for 10 min at 25°C. Cells were collected by centrifugation (3000 x g; 15 min at 4°C) and cell pellets were stored at -20°C. To lyse cells, the frozen cell pellets were incubated at 4°C for 30 min, re-suspended in 40 ml of 50 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 8.0), 1.4 M NaCl, 20 mM imidazole, 14 mM β-mercaptoethanol, 0.1% Tween 20, and 5% ethanol containing 2 mg egg white lysozyme, and sonicated for 5 min at 4°C. Lysates were centrifuged (23,000 x g; 40 min at 4°C), supernatants were added to 1 ml Ni-NTA-agarose (Qiagen) in 50 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 8.0), 1.4 M NaCl, 20 mM imidazole, 0.1% Tween 20, 5 mM β-mercaptoethanol, and 5% ethanol, and the mixture was incubated 30 min at 4°C with gentle rocking. The slurry was loaded into a 15 ml polyprep column (BioRad) to collect the Ni-NTA-agarose resin. The resin was washed with 10 ml of 1 X WB (50 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.1% Tween 20, 5 mM β-mercaptoethanol, and 5% ethanol) and His-tagged RNAP was eluted from the resin with 3 ml of 1 X WB containing 300 mM imidazole. The eluate was concentrated to 0.2 ml in 20 mM Tris-Cl (pH 8.0), 200 mM KCl, 20 mM MgCl<sub>2</sub>, 0.2 mM EDTA, and 1 mM DTT using a 1000 MWCO Amicon Ultra-4 centrifugal filter (EMD Millipore), 0.2 ml glycerol was added, and the sample was stored at -20°C.

To denature RNAP-DNA complexes, 25 µl of the concentrated eluate was mixed with 25 µl water, 15 µl 5 M NaCl, and 6 µl 100 µg/ml heparin, incubated at 95°C for 5 min, then incubated at 4°C for 30 min. Crosslinked RNAP-DNA complexes were isolated by adding 20 µl MagneHis Ni-particles equilibrated and suspended in 10 mM Tris-Cl, pH 8.0, 1.2 M NaCl, 10 mM MgCl<sub>2</sub>, 10 µg/ml heparin, and 0.1 mg/ml BSA; MagneHis Ni-particles were collected using a magnetic microfuge tube rack; particles were washed with 50 µl 10 mM Tris-Cl, pH 8.0, 1.2 M NaCl, 10 mM MgCl<sub>2</sub>, 10 µg/ml heparin, and 0.1 mg/ml BSA, washed twice with 50 µl 1 X *Taq* DNA polymerase buffer (New England Biolabs), and the particles (which contained bound RNAP-DNA complexes) were resuspended in 10 µl 1 X *Taq* DNA polymerase buffer. Primer extension reactions and analysis of radiolabeled products generated in these reactions were

performed using procedures identical to those used in the analysis of *placCONS in vitro* (see above) using material isolated from cells containing pCDF-*lacCONS* (Figure 3B), cells containing the pCDF-*lacCONS*-N11 library (Figure 4B), or cells containing the pCDF-*lacCONS*-N20 library (Figure S2).

XACT-seq experiments (see below) were performed using denatured RNAP-DNA complexes isolated from cells containing the pCDF-*lacCONS*-N11 library.

### **XACT-seq: primer extension**

Primer extension reactions (50  $\mu$ l) were performed by combining 8  $\mu$ l of recovered RNAP-DNA complexes, 1  $\mu$ l of 1  $\mu$ M primer s128a, 5  $\mu$ l 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM TTP, 1  $\mu$ l 5 U/ $\mu$ l Taq DNA polymerase, 10  $\mu$ l 5 M betaine, 5  $\mu$ l 100% dimethyl sulfoxide, and 5  $\mu$ l 10 X Taq DNA polymerase buffer, and cycling 40 times through 30 s at 95°C, 30 s at 53°C, and 30 s at 72°C. Primer extension products were isolated by ethanol precipitation, washed twice with 80% cold ethanol, resuspended in 20  $\mu$ l water, and mixed with 20  $\mu$ l of 2 X RNA loading dye (95% deionized formamide, 18 mM EDTA, 0.25% SDS, xylene cyanol, bromophenol blue, amaranth).

Primer extension products were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1 X TBE). The gel was stained with SYBR Gold nucleic acid gel stain (Life Technologies) and ssDNA products ~35- to ~75-nt in size were excised from the gel. To elute nucleic acids from the gel, the gel fragment was crushed as described in (Vvedenskaya et al., 2018a), 300  $\mu$ l of 0.3 M NaCl in 1 X TE buffer was added, the mixture was incubated for 10 min at 70°C, and the supernatant was collected using a Spin-X column (Corning). The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 5  $\mu$ l of nuclease-free water.

### **XACT-seq: 3'-adapter ligation and library amplification**

The recovered primer extension products (5  $\mu$ l) were combined with 1  $\mu$ l 10 X NEB buffer 1, ~0.8  $\mu$ M 3'-adapter oligo s1248 [5' adenylylated and 3'-end blocked oligo containing ten randomized nucleotides (10N) at the 5' end], 5 mM MnCl<sub>2</sub> and 1  $\mu$ M of 5'-AppDNA/RNA ligase (New England Biolabs) in a final volume of 10  $\mu$ l. The mixture was incubated for 1 h at 65°C

followed by 3 min at 90°C, and cooled to 4°C for 5 min. The reaction was combined with 15 µl of mixture containing 10 U of T4 RNA ligase 1 (New England Biolabs), 1 X T4 RNA ligase 1 reaction buffer, 12% PEG 8000, 10 mM DTT, 60 µg/mL BSA. Reactions were incubated at 16°C for 16 h.

To determine the effect of sequence on the efficiency of ligation between primer extension products and the 3'-adapter oligo s1248, we performed a control ligation reaction containing oligo s1248 and oligo JW402, which contains 15 random nucleotides (15N) at the 5' end.

Adapter-ligated products were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1 X TBE). The gel was stained with SYBR Gold nucleic acid gel stain and species ranging from ~60 to ~100-nt (for reactions containing oligo s1248 and primer extension products) or ~50 and 90 nt (for reactions containing oligo s1248 and oligo JW402) were isolated by gel excision. The gel fragment was crushed, 600 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column (Corning). The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 20 µl of nuclease-free water.

Adapter-ligated DNA (5 µl) was added to PCR reactions containing 1 X Phusion HF reaction buffer, 0.2 mM dNTPs, 0.25 µM Illumina RP1 primer, 0.25 µM Illumina RPI index primer and 0.02 U/µl Phusion HF polymerase [40 µl total volume; 98°C for 30 s, 98°C for 10 s, 62°C for 20 s, 72°C for 10 seconds (11 cycles), 72°C for 5 min]. Reaction products were separated by electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE), and amplicons between ~160 bp and ~170 bp were isolated by gel excision. The gel fragment was crushed, 600 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 20 µl of nuclease-free water.

Libraries generated by this procedure are: CP21/CP21D, CP23/CP23D, CP27/CP27C, CP22/CP22D, CP24/CP24D, and CP28/CP28C.

## **XACT-seq: generation of library for analysis of template sequences in the *placCONS-N11* library**

We used emulsion PCR (ePCR) to generate a DNA library to identify template sequences present in the *placCONS-N11* library. ePCR reaction mixtures contained  $\sim 10^9$  molecules of the *placCONS-N11* plasmid library, 1 X Detergent-free Phusion HF reaction buffer containing 5  $\mu\text{g/ml}$  BSA, 0.4 mM dNTPs, 0.5  $\mu\text{M}$  Illumina RP1 primer, 0.5  $\mu\text{M}$  Illumina index primer and 0.04 U/ $\mu\text{l}$  Phusion HF polymerase [95°C for 10 s, 95°C for 5 s, 60°C for 5 s, 72°C for 15 seconds (20 cycles), 72°C for 5 min]. DNA was isolated from ePCR reactions using a Micellula DNA Emulsion and Purification Kit. The emulsion was broken, DNA was purified according to the manufacturer's recommendations, recovered by ethanol precipitation, and resuspended in 20  $\mu\text{l}$  of nuclease-free water. Products were subjected to electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE), the 204 bp fragment was excised from the gel. The gel fragment was crushed, 600  $\mu\text{l}$  of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 20  $\mu\text{l}$  of nuclease-free water. The library generated by this procedure is CP26T.

## **XACT-seq: quantitation of template sequences**

To determine the relative abundance of each template sequences, we isolated plasmid DNA from 2 ml of cell suspensions harvested prior to UV irradiation. Plasmid DNA ( $\sim 120$  ng) was mixed with 1 X Phusion HF reaction buffer, 0.1 mM dNTPs, 0.125  $\mu\text{M}$  biotinylated oligonucleotide JW401, and 0.01 U/ $\mu\text{l}$  Phusion HF polymerase in a 20  $\mu\text{l}$  reaction. Primer extension reactions we performed by denaturing plasmid DNA for 30 s at 98°C, followed by 2 cycles of denaturation for 10 s at 98°C, annealing for 20 s at 62°C, extension for 15 s at 72°C, and a final extension for 5 min at 72°C. Reactions were subjected to electrophoresis using a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE). The gel was stained with SYBR Gold nucleic acid gel stain and the region of the gel containing ssDNA products larger than  $\sim 40$ -nt was excised. The gel fragment was crushed, 600  $\mu\text{l}$  of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids



were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 10 µl of nuclease-free water. 1 µl of this solution was amplified by ePCR, products were recovered from emulsions as described above, and subjected to electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE). The 204 bp fragment was excised from the gel, the gel fragment was crushed, 600 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 20 µl of nuclease-free water.

Libraries generated by this procedure are: Vv1362, Vv1363, and Vv1364.

### **XACT-seq: HTS sample serial numbers**

CP21, CP21D, CP23, CP23D, CP27, and CP27C are samples used for the identification of RNAP-active-center A-site positions in active transcription complexes *in vivo* (- Rifampin). CP22, CP22D, CP24, CP24D, CP28, and CP28C are samples used for the identification of RNAP-active-center A-site positions in static transcription complexes *in vivo* (+ Rifampin). CP21/CP21D, CP23/CP23D, CP27/CP27C, CP22/CP22D, CP24/CP24D, and CP28/CP28C are technical replicates derived from the same biological sample.

Sample CP26T was used to identify template sequences present in the *plac*CONS-N11 library. Vv1362, Vv1363, and Vv1364 are samples used quantify template sequences for each biological replicate (Vv1362 was used to quantify template sequences for CP21/CP21D/CP22/CP22D; Vv1363 was used to quantify template sequences for CP23/CP23D/CP24/CP24D; Vv1364 was used to quantify template sequences for CP27/CP27C/CP28/CP28C). Samples CP40, CP41, and CP42 are libraries generated to determine the effect of sequence on the efficiency of adapter ligation.

All samples were sequenced using an Illumina NextSeq with custom sequencing primer s1115.

### **XACT-seq: data analysis**

Sequencing reads derived from sample CP26T that were a perfect match to non-randomized sequences of *plac*CONS-N11 were identified:





RNAP-active-center A-site at positions +3, +4, +5, +6, +7, +8, or +9 were aligned to generate the sequence logo (Wagih, 2017) for the A-site position (A), the position 1 bp downstream of the A-site (A<sub>+1</sub>), and the position 2 bp downstream of the A-site (A<sub>+2</sub>); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +4, +5, +6, +7, +8, or +9 were aligned to generate the sequence logo for the P-site position (P); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +5, +6, +7, +8, or +9 were aligned to generate the sequence logo for the position immediately upstream of the P-site (P<sub>-1</sub>); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +6, +7, +8, or +9 were aligned to generate the sequence logo for the position 2 bp upstream of the P-site (P<sub>-2</sub>); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +7, +8, or +9 were aligned to generate the sequence logo for the position 3 bp upstream of the P-site (P<sub>-3</sub>); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +8, or +9 were aligned to generate the sequence logo for the position 4 bp upstream of the P-site (P<sub>-4</sub>); the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at position +9 were aligned to generate the sequence logo for the position 5 bp upstream of the P-site (P<sub>-5</sub>).

To generate the plot shown in Figure 5B, values of RNAP occupancy for promoter-positions +5, +6, +7, +8, or +9 over sequences from position +5 to position +9 were calculated for nontemplate-strand tetranucleotide sequences at the positions corresponding to P<sub>-1</sub>PAA<sub>+1</sub> (+3 to +6 for A-site position +5; +4 to +7 for A-site position +6; +5 to +8 for A-site position +7; +6 to +9 for A-site position +8; +7 to +10 for A-site position +9). Plot shows the mean and SEM calculated from three biological replicates.

To define the consensus sequence for initial-transcription pausing at each position (Figure 5C), we identified *plac*CONS-N11 template sequences for which RC<sub>total</sub> >10 (with RC<sub>total</sub> = sum of RC at each position from -3 to +20). Pause sites were identified as positions with RNAP occupancy >50%. The position-sequence logos for pausing at positions +3, +4, +5, +6, +7, +8, or +9 were generated by aligning the sequences of nontemplate-strand nucleotides of pause sites with the RNAP-active-center A-site at positions +3, +4, +5, +6, +7, +8, or +9, respectively.

**Table S1. *plac*CONS-N11 template sequences containing pauses at positions +3 to +9 identified by XACT-seq**

| A-site position | Sequences with pausing | Percentage of sequences in library |
|-----------------|------------------------|------------------------------------|
| +3              | 77,293                 | 9.90                               |
| +4              | 22,783                 | 2.92                               |
| +5              | 10,686                 | 1.37                               |
| +6              | 23,192                 | 2.97                               |
| +7              | 10,190                 | 1.30                               |
| +8              | 286                    | 0.03                               |
| +9              | 103                    | 0.01                               |
| +3 to +9        | 144,533                | 18.5                               |

**Table S2. Oligonucleotides**

| Name  | Sequence (5' to 3')                                                                                                                                                                                                                                        | Description                                                                                                                                                             |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| JW44  | GTCGATGATATCGCGCACCT<br>CGGCAACCG                                                                                                                                                                                                                          | oligo for site-directed mutagenesis to generate rpoB D446A                                                                                                              |
| JW61  | CGGCATCACCATCGGCATTG<br>AC                                                                                                                                                                                                                                 | oligo for amplifying <i>plac</i> CONS and derivatives and for primer extension reactions                                                                                |
| JW62  | CTCCAGGTACCCGCAATAAA<br>TGTG                                                                                                                                                                                                                               | oligo for amplifying <i>plac</i> CONS and derivatives                                                                                                                   |
| JW79  | AATTAAGCCGCTGGGGCCCT<br>TGGCACCCGAGAATTCCCGG<br>CATCACCATCGGCATTGAC                                                                                                                                                                                        | oligo used to clone <i>plac</i> CONS derivatives                                                                                                                        |
| JW80  | TATAATGCCTGACCGGCGTT<br>CAGAGTTCTACAGTCCGACG<br>ATCCTCCAGGTACCCGCAAT                                                                                                                                                                                       | oligo used to clone <i>plac</i> CONS derivatives                                                                                                                        |
| JW153 | GTTCAGAGTTCTACAGTCCG<br>ACGATCNNNNNNNNNNNNNN<br>NNNNAGTGAGCGCAACGCA<br>ATAACAGTCATCTAGATAGA<br>ACTTTAGGCACCCCAGGCTT<br><b>GACACTTTATGCTTCGGCTC</b><br><b>GTATAATGTGTGGAANNNN</b><br>NNNNNNNgataacaattcAACAA<br>TNNNNNNNNNNNNNNNNNT<br>GGAATTCTCGGGTGCCAAGG | Template oligo carrying <i>plac</i> CONS sequence a 11-nt randomized region and a 17-nt randomized region. Promoter -10 (TATAAT) and -35 elements (TTGACA) are in bold. |
| JW203 | GTTCAGAGTTCTACAGTCCG<br>ACGATCNNNNNNNNNNNNNN<br>NNNNAGTGAGCGCAACGCA<br>ATAACAGTCATCTAGATAGA<br>ACTTTAGGCACCCCAGGCTT<br><b>GACACTTTATGCTTCGGCTC</b><br><b>GTATAATGTGTGGAANNNN</b><br>NNNNNNNNNNNNNNNNNtttcAA<br>CAATNNNNNNNNNNNNNNNN                        | Template oligo carrying <i>plac</i> CONS sequence a 20-nt randomized region and a 17-nt randomized region. Promoter -10 (TATAAT) and -35 elements (TTGACA) are in bold. |

|       |                                                                                                            |                                                                                                  |
|-------|------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
|       | NNTGGAATTCTCGGGTGCCA<br>AGG                                                                                |                                                                                                  |
| JW268 | GGGGATTTCATCGAGGATAC<br>CACC                                                                               | oligo for generating $\sigma$ $\Delta$ 513-519 ( $\Delta$ finger)<br>mutant                      |
| JW269 | ACCGATCGGCGTTTCCATG                                                                                        | Reverse mutagenesis primer for generating $\sigma$<br>$\Delta$ 513-519 ( $\Delta$ finger) mutant |
| JW401 | /5deSBioTEG/GTTCAGAGTTC<br>TACAGTCCGACGATCNNNN<br>NNNNNNNNNGCTTCGGCT<br>CGTATAATGTG                        | First strand synthesis oligo for counting<br>templates                                           |
| JW402 | CCTTGGCACCCGAGAATTCC<br>ANNNNNNNNNNNNNNNNN                                                                 | 3' adaptor used in ligation bias experiments                                                     |
| JW408 | CGGCATCACCATCGGCATTG<br>ACGATGCCGTGTATCCCGAA<br>TTCTTGAAATTGTTATCCCA<br>AATCATATTTCCACACATTA<br>TACGAGCCG  | Template used to clone <i>placCONS</i> -p6                                                       |
| JW409 | CGGCATCACCATCGGCATTG<br>ACGATGCCGTGTATCCCGAA<br>TTCTTGAAATTGTTATCGTG<br>AACGCATCTTCCACACATTA<br>TACGAGCCG  | Template used to clone <i>placCONS</i> -p7                                                       |
| JW433 | CGGCATCACCATCGGCATTG<br>ACGATGCCGTGTATCCCGAA<br>TTCTTGAAATTGTTATCTCG<br>AGTTTCGCCTTCCACACATTA<br>TACGAGCCG | Template used to clone <i>placCONS</i> -p5                                                       |
| JW575 | CGGCATCACCATCGGCATTG<br>ACGATGCCGTGTATCCCGAA<br>TTCTTGAAATTGTTATCAGA<br>CATACCCGTTCCACACATTA<br>TACGAGCCG  | Template used to clone <i>placCONS</i> -p9                                                       |
| JW580 | CGGCATCACCATCGGCATTG<br>ACGATGCCGTGTATCCCGAA<br>TTCTTGAAATTGTTATCAAT<br>ACGTTTCGCTTCCACACATTA<br>TACGAGCCG | Template used to clone <i>placCONS</i> -p8                                                       |
| s1115 | CTACACGTTTCAGAGTTCTAC<br>AGTCCGACGATC                                                                      | custom Illumina sequencing primer (HPLC<br>purified)                                             |
| s1219 | tataatgcctgaccggcGTTTCAGAGT<br>TCTACAGTCCGACGATC                                                           | oligo for amplifying and cloning <i>placCONS</i><br>libraries                                    |
| s1220 | aattaagcgcgtggggcCCTTGGCAC<br>CCGAGAATTCC                                                                  | oligo for amplifying and cloning <i>placCONS</i><br>libraries                                    |
| RP1   | AATGATACGGCGACCACCG<br>AGATCTACACGTTTCAGAGTT<br>CTACAGTCCGA                                                | Illumina PCR primer (HPLC purified)                                                              |
| RPI1  | CAAGCAGAAGACGGCATA<br>GAGATCGTGATGTGACTGGA<br>GTTCTTGGCACCCGAGAAT<br>TCCA                                  | Illumina indexing PCR primer 1 (HPLC<br>purified)                                                |

|       |                                                           |                                                |
|-------|-----------------------------------------------------------|------------------------------------------------|
| s128a | CCTTGGCACCCGAGAATTCC                                      | Used for primer extension on library templates |
| s1248 | /5'Phos/NNNNNNNNNGATC<br>GTCGGACTGTAGAACTCTGA<br>AC/3ddC/ | 3' adapter with 10N at 5' end (HPLC purified)  |

## References

- 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.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837.
- Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *The Journal of biological chemistry* **289**, 24549-24559.
- Belogurov, G.A., and Artsimovitch, I. (2019). The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *Journal of Molecular Biology*
- Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **99**, 11020-11024.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373.
- Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). *Current protocols in molecular biology* / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 14 11-17.
- Duchi, D., Bauer, D.L., Fernandez, L., Evans, G., Robb, N., Hwang, L.C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., *et al.* (2016). RNA Polymerase Pausing during Initial Transcription. *Molecular Cell* **63**, 939-950.
- Dulin, D., Bauer, D.L., Malinen, A.M., Bakermans, J.J., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., and Kulbachinskiy, A. (2018). Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nature Communications* **9**, 1478.
- Erie, D.A., Yager, T.D., and von Hippel, P.H. (1992). The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu Rev Biophys Biomol Struct* **21**, 379-415.

Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., and Weixlbaumer, A. (2018). Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* 69, 816-827 e814.

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.

Imashimizu, M., Takahashi, H., Oshima, T., McIntosh, C., Bubunenko, M., Court, D.L., and Kashlev, M. (2015). Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biology* 16, 98.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.

Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* 69, 802-815 e805.

Kang, J.Y., Mishanina, T.V., Landick, R., and Darst, S.A. (2019). Mechanisms of Transcriptional Pausing in Bacteria. *Journal of Molecular Biology*.

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.

Kulbachinskiy, A., and Mustaev, A. (2006). Region 3.2 of the  $\sigma$  subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *The Journal of biological chemistry* 281, 18273-18276.

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.

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.

Latif, H., Federowicz, S., Ebrahim, A., Tarasova, J., Szubin, R., Utrilla, J., Zengler, K., and Palsson, B.O. (2018). ChIP-exo interrogation of Crp, DNA, and RNAP holoenzyme interactions. *PLoS One* 13, e0197272.

Lerner, E., Chung, S., Allen, B.L., Wang, S., Lee, J., Lu, S.W., Grimaud, L.W., Ingargiola, A., Michalet, X., Alhadid, Y., *et al.* (2016). Backtracked and paused transcription initiation

intermediate of *Escherichia coli* RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America *113*, E6562-E6571.

Margeat, E., Kapanidis, A.N., Tinnefeld, P., Wang, Y., Mukhopadhyay, J., Ebricht, R.H., and Weiss, S. (2006). Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. Biophysical journal *90*, 1419-1431.

Marr, M.T., and Roberts, J.W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science *276*, 1258-1260.

Mazumder, A., and Kapanidis, A.N. (2019). Recent Advances in Understanding sigma70-Dependent Transcription Initiation Mechanisms. J Mol Biol.

McClure, W.R., and Cech, C.L. (1978). On the mechanism of rifampicin inhibition of RNA synthesis. J Biol Chem *253*, 8949-8956.

Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? Molecular Cell *20*, 335-345.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science *296*, 1280-1284.

Pupov, D., Kuzin, I., Bass, I., and Kulbachinskiy, A. (2014). Distinct functions of the RNA polymerase  $\sigma$  subunit region 3.2 in RNA priming and promoter escape. Nucleic Acids Research *42*, 4494-4504.

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

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Ruff, E.F., Record, M.T., Jr., and Artsimovitch, I. (2015). Initial events in bacterial transcription initiation. Biomolecules *5*, 1035-1062.

Saba, J., Chua, X.Y., Mishanina, T.V., Nayak, D., Windgassen, T.A., Mooney, R.A., and Landick, R. (2019). The elemental mechanism of transcriptional pausing. Elife *8*.

Sambrook, J., Russell, D.e., and Sambrook, J. (2006). The condensed protocols from molecular cloning: a laboratory manual. NY: Cold Spring Harbor Laboratory Press v.

Severinov, K., Fenyö, D., Severinova, E., Mustaev, A., Chait, B.T., Goldfarb, A., and Darst, S.A. (1994). The sigma subunit conserved region 3 is part of the "5'-face" of the active center of *Escherichia coli* RNA polymerase. The Journal of biological chemistry *269*, 20826-20828.

Srivastava, D.B., Leon, K., Osmundson, J., Garner, A.L., Weiss, L.A., Westblade, L.F., Glickman, M.S., Landick, R., Darst, S.A., Stallings, C.L., *et al.* (2013). Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* *110*, 12619-12624.

Vvedenskaya, I.O., Bird, J.G., Zhang, Y., Zhang, Y., Jiao, X., Barvik, I., Krasny, L., Kiledjian, M., Taylor, D.M., Ebright, R.H., *et al.* (2018a). CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD. *Molecular Cell* *70*, 553-564 e559.

Vvedenskaya, I.O., Goldman, S.R., and Nickels, B.E. (2018b). Analysis of Bacterial Transcription by "Massively Systematic Transcript End Readout," MASTER. *Methods in Enzymology* *612*, 269-302.

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.

Vvedenskaya, I.O., Zhang, Y., Goldman, S.R., Valenti, A., Visone, V., Taylor, D.M., Ebright, R.H., and Nickels, B.E. (2015). Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Molecular Cell* *60*, 953-965.

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* *33*, 3645-3647.

Winkelman, J.T., Chandrangsu, P., Ross, W., and Gourse, R.L. (2016a). Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proceedings of the National Academy of Sciences of the United States of America* *113*, E1787-E1795.

Winkelman, J.T., Nickels, B.E., and Ebright, R.H. (2019). The transition from transcription initiation to transcription elongation: start-site selection, initial transcription, and promoter escape. In *RNA Polymerase as a Molecular Motor, Second Edition*, R. Landick, J. Wang, and T. Strick, eds. (Cambridge, UK: RSC Publishing).

Winkelman, J.T., Vvedenskaya, I.O., Zhang, Y., Zhang, Y., Bird, J.G., Taylor, D.M., Gourse, R.L., Ebright, R.H., and Nickels, B.E. (2016b). Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* *351*, 1090-1093.

Winkelman, J.T., Winkelman, B.T., Boyce, J., Maloney, M.F., Chen, A.Y., Ross, W., and Gourse, R.L. (2015). Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol Cell* *59*, 768-780.

Yu, L., Winkelman, J.T., Pukhrambam, C., Strick, T.R., Nickels, B.E., and Ebright, R.H. (2017). The mechanism of variability in transcription start site selection. *eLife* *6*, e32038.

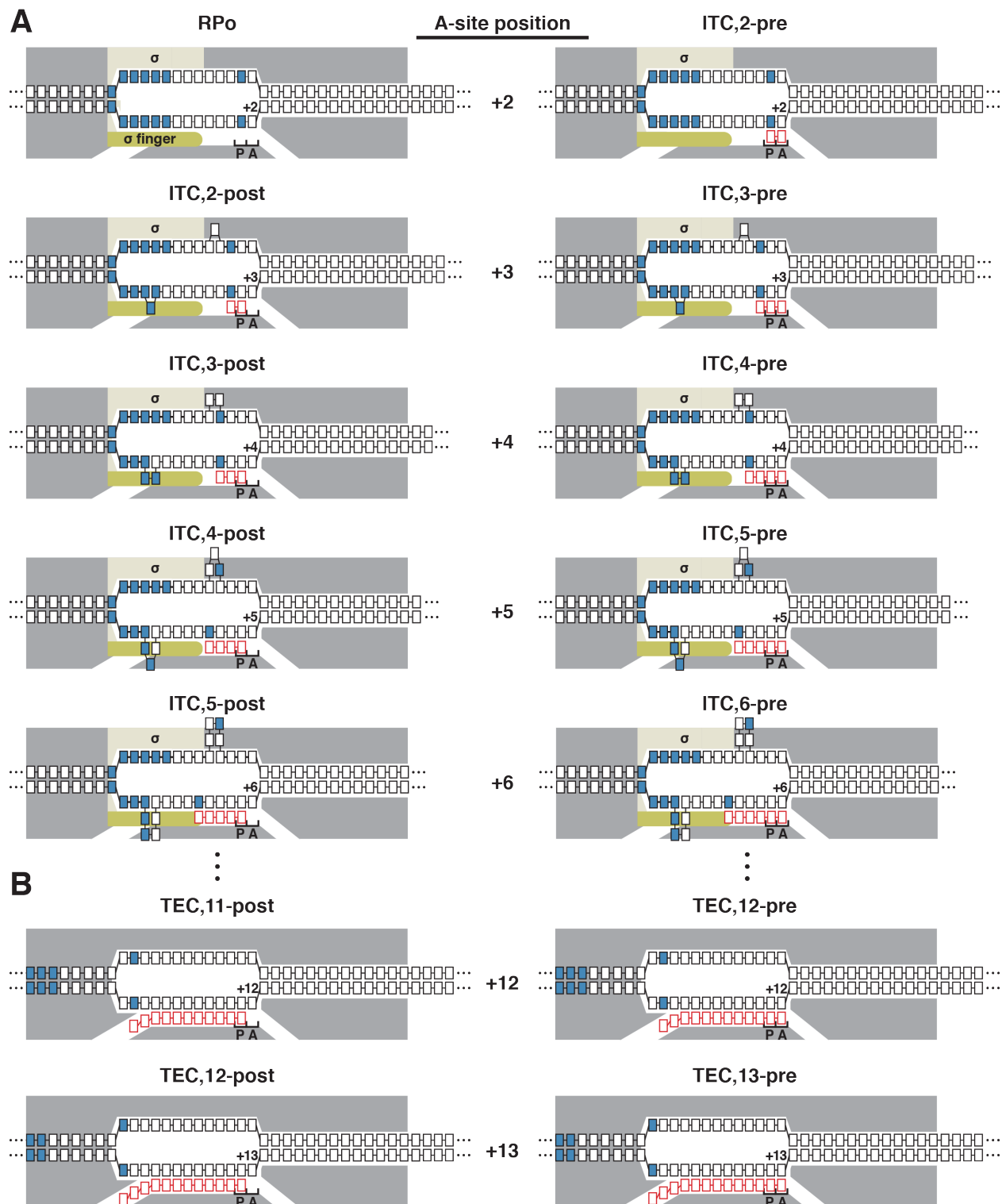


Zhang, L.R., and Landick, R. (2009). Substrate loading, nucleotide addition, and translocation by RNA polymerase. In RNA Polymerases as Molecular Motors, H. Buc, and S. Strick, eds. (London, UK: Royal Society of Chemistry).

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* 338, 1076-1080.

Zuo, Y., and Steitz, T.A. (2015). Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Molecular Cell* 58, 534-540.



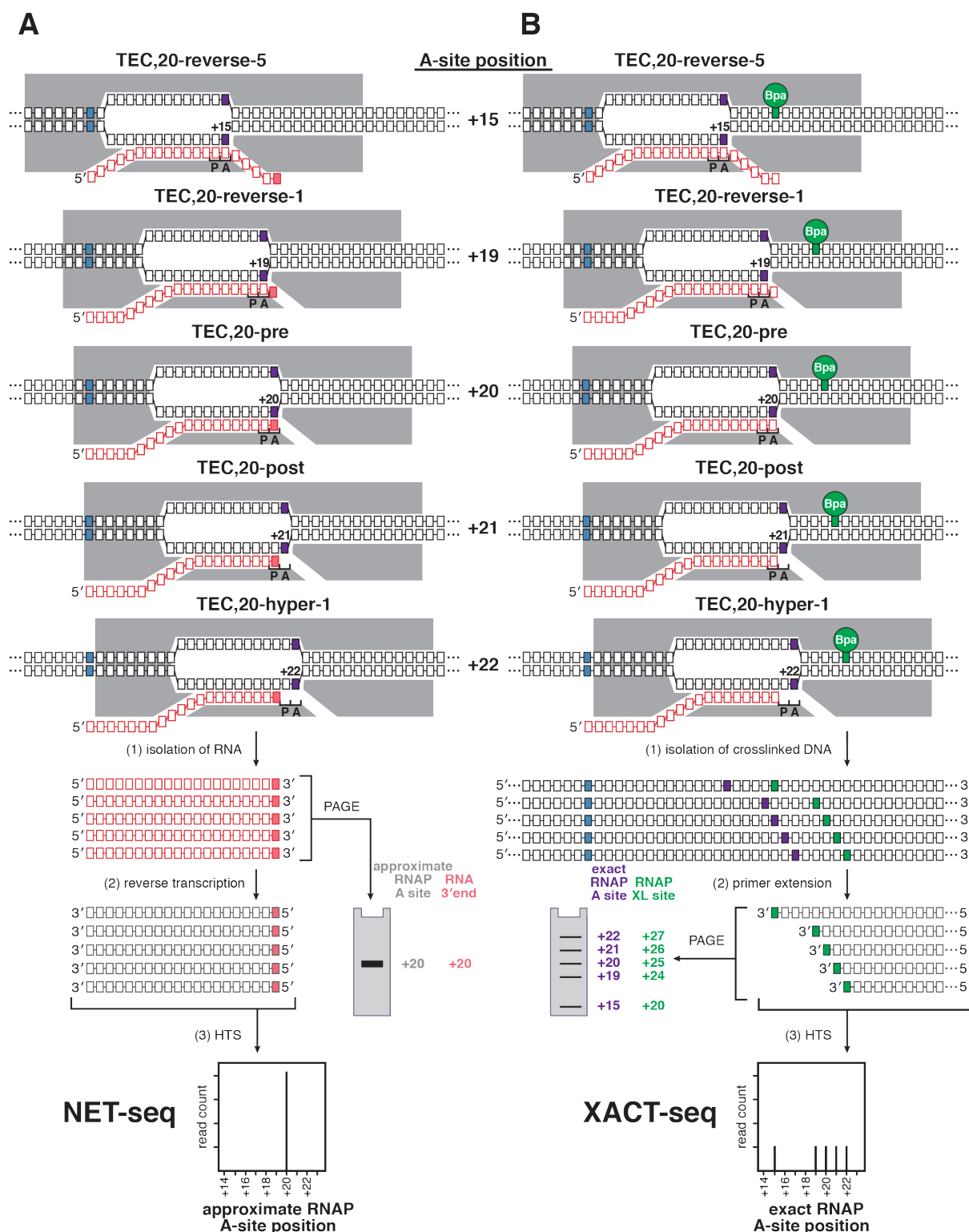


**Figure 1. Mechanisms of initial transcription and transcription elongation**

(A) Mechanism of initial transcription, with RNAP-active-center translocation through DNA scrunching mechanism and with RNAP in complex with transcription initiation factor  $\sigma$ . Panel shows first six nucleotide-addition steps of initial transcription, starting from the

RNAP-promoter open complex (RPO), and yielding a series of successive RNAP-promoter initial transcribing complexes containing 2, 3, 4, 5, and 6 nt RNA products (ITC,2, ITC,3, ITC,4, ITC,5, and ITC,6). Following each nucleotide addition, the RNAP active center translocates forward through a DNA scrunching mechanism, from a pre-translocated state (pre; right column) to a post-translocated state (post; left column), and the RNAP-active-center A-site position advances by 1 bp (numbered positions between left and right columns). Grey, RNAP; yellow,  $\sigma$ ; dark yellow,  $\sigma$  finger; blue, -10 element nucleotides and TSS nucleotides; P and A, RNAP-active-center P-site and A-site, respectively; black boxes, DNA nucleotides (nontemplate-strand nucleotides above template-strand nucleotides); red boxes, RNA nucleotides; promoter positions are numbered relative to the TSS position, +1.

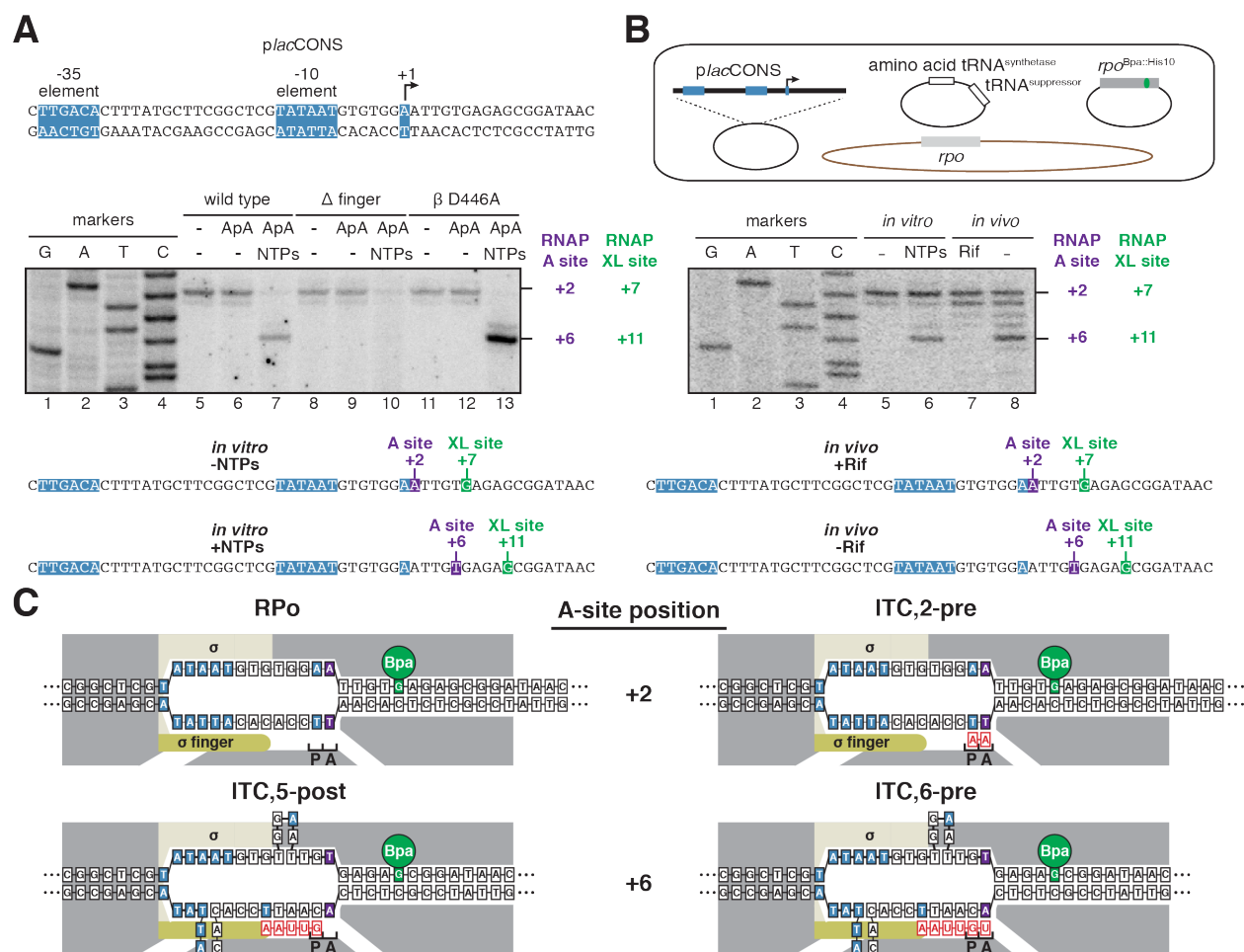
(B) Mechanism of transcription elongation, with RNAP-active center translocation through DNA stepping mechanism and with RNAP not containing  $\sigma$ . Panel shows two nucleotide-addition steps of transcription elongation, from a transcription elongation complex containing 11 nt of RNA (TEC,11) to a transcription elongation complex containing 12 nt of RNA (TEC,12). Following each nucleotide addition, RNAP, as a whole, translocates forward through a DNA stepping mechanism, between a pre-translocated state (pre; right column) and a post-translocated state (post; left column), and the RNAP-active-center A-site position advances by 1 bp (numbered positions between left and right columns). Symbols and colors as in panel A.



**Figure 2. Approaches to map RNAP-active-center A-site position: NET-seq and XACT-seq**  
 (A) Approach to map RNAP-active-center A-site position through analysis of RNA 3' ends. The procedure entails isolating nascent RNA from transcription complexes and identifying and quantifying RNA 3' ends by polyacrylamide gel electrophoresis (PAGE) or by analysis of primer

extension products by high-throughput sequencing (HTS; NET-seq). The RNAP-active-center A-site position (purple boxes on nontemplate- and template-strand nucleotides; numbered relative to the TSS position, +1) of the promoter position in the RNAP-active-center A-site is approximated based on the identification of the RNA 3' end (pink boxes); other symbols and colors as in Figure 1. The multiple different translocational states adopted by a transcription elongation complex (TEC) containing a single defined RNA product--e.g. reverse-translocated by 5 bp (TEC,20-reverse-5), reverse-translocated by 1 bp (TEC,20-reverse-1), pre-translocated (TEC, 20-pre), post-translocated (TEC,20-post), and hyper-translocated by 1 bp (TEC,20-hyper-1)--all yield the same RNA 3' end and therefore cannot be distinguished.

(B) Approach to map the RNAP-active-center A-site position by site-specific protein-DNA photo-crosslinking with an RNAP derivative that contains a photo-activatable agent (green circle labeled Bpa) that crosslinks to DNA a defined distance from the RNAP active center. The procedure entails UV irradiating transcription complexes, isolating RNAP-crosslinked DNA, and identifying and quantifying crosslinking sites by analysis of primer extension products by PAGE or HTS (XACT-seq). The RNAP-active-center A-site position (purple boxes on template and nontemplate strands; numbered relative to the TSS position, +1) is defined based on the identity of the crosslinking site (green boxes; XL site); other symbols as in panel A. The multiple different translocational states adopted by a transcription elongation complex (TEC) containing a single defined RNA product--e.g. reverse-translocated by 5 bp (TEC,20-reverse-5), reverse translocated by 1 bp (TEC,20-reverse-1), pre-translocated (TEC, 20-pre), post-translocated (TEC,20-post), and a hyper-translocated by 1 bp state (TEC,20-hyper-1)--all yield different crosslinking sites and therefore can be distinguished.

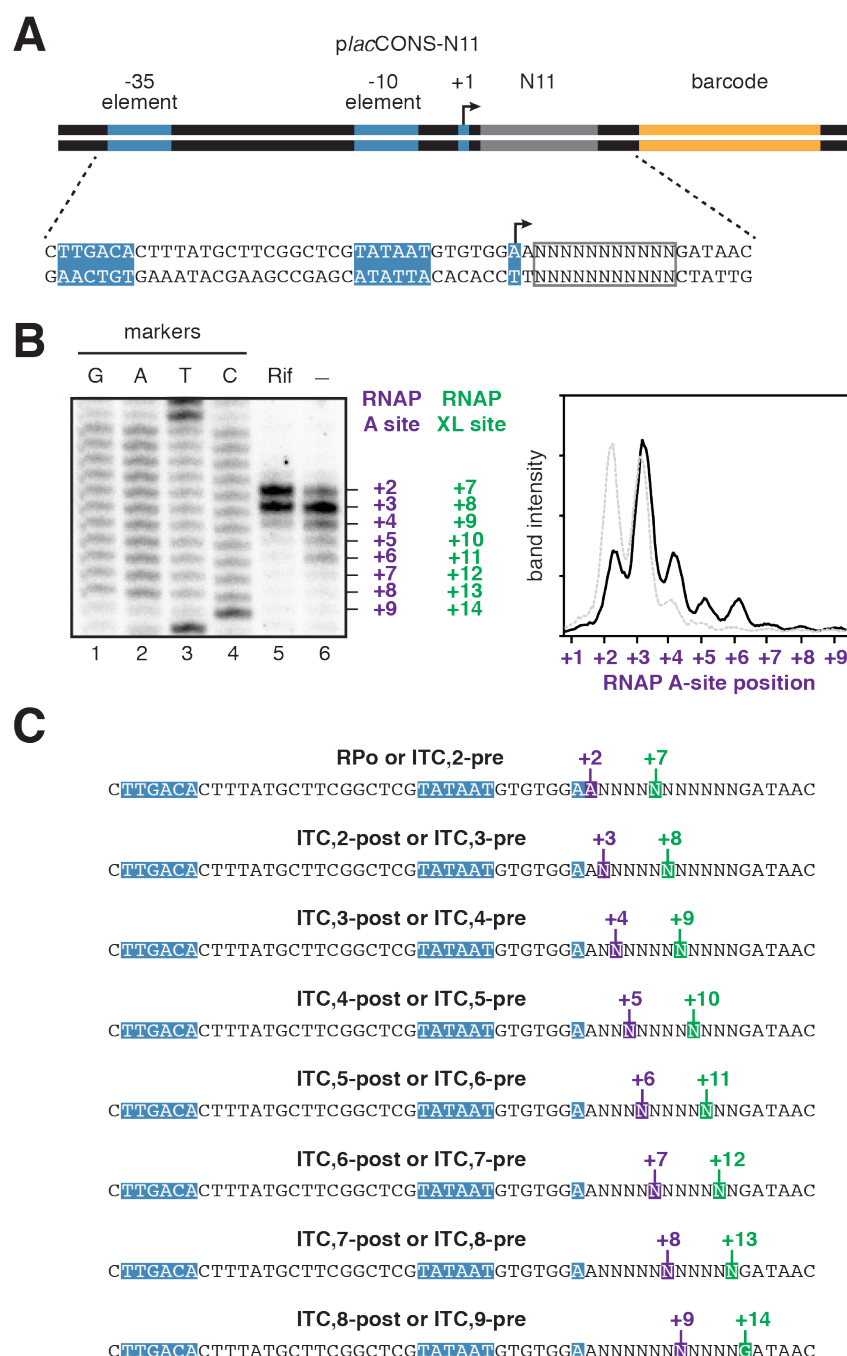


**Figure 3. RNAP-active-center A-site positions in initial-transcription pausing at the lacCONS promoter *in vitro* and *in vivo***

(A) Analysis of initial-transcription pausing at lacCONS by RNAP-β<sup>R1148Bpa</sup> *in vitro*. Top, lacCONS promoter (lacCONS). Middle, results. Bottom, position of RNAP-active-center A-site (purple) and nucleotide crosslinked to Bpa (green) defined relative to the TSS position. Markers, DNA sequence ladder generated using lacCONS template.

(B) Analysis of initial-transcription pausing at lacCONS by RNAP-β<sup>R1148Bpa</sup> *in vivo*. Top, three-plasmid merodiploid system for co-production, in *E. coli* cells, of decahistidine-tagged, RNAP-β<sup>R1148Bpa</sup>, in the presence of untagged wild-type RNAP. The first plasmid carries gene for RNAP subunit with a nonsense (TAG) codon at residue β<sup>R1148</sup> (grey rectangle; green indicates nonsense codon); The second plasmid carries genes for engineered Bpa-specific nonsense-suppressor tRNA and Bpa-specific amino-acyl-tRNA synthetase (white rectangles); the third plasmid carries lacCONS; and the chromosome (brown oval) carries wild-type RNAP subunit genes (light grey rectangle). Middle, results. Bottom, position of RNAP-active-center A-site (purple) and nucleotide crosslinked to Bpa (green) defined relative to the TSS position. Rif, Rifampin; Markers, DNA sequence ladder generated using lacCONS template.

(C) Interpretation of results in (A) and (B). Symbols and colors as in Figures 1,2.

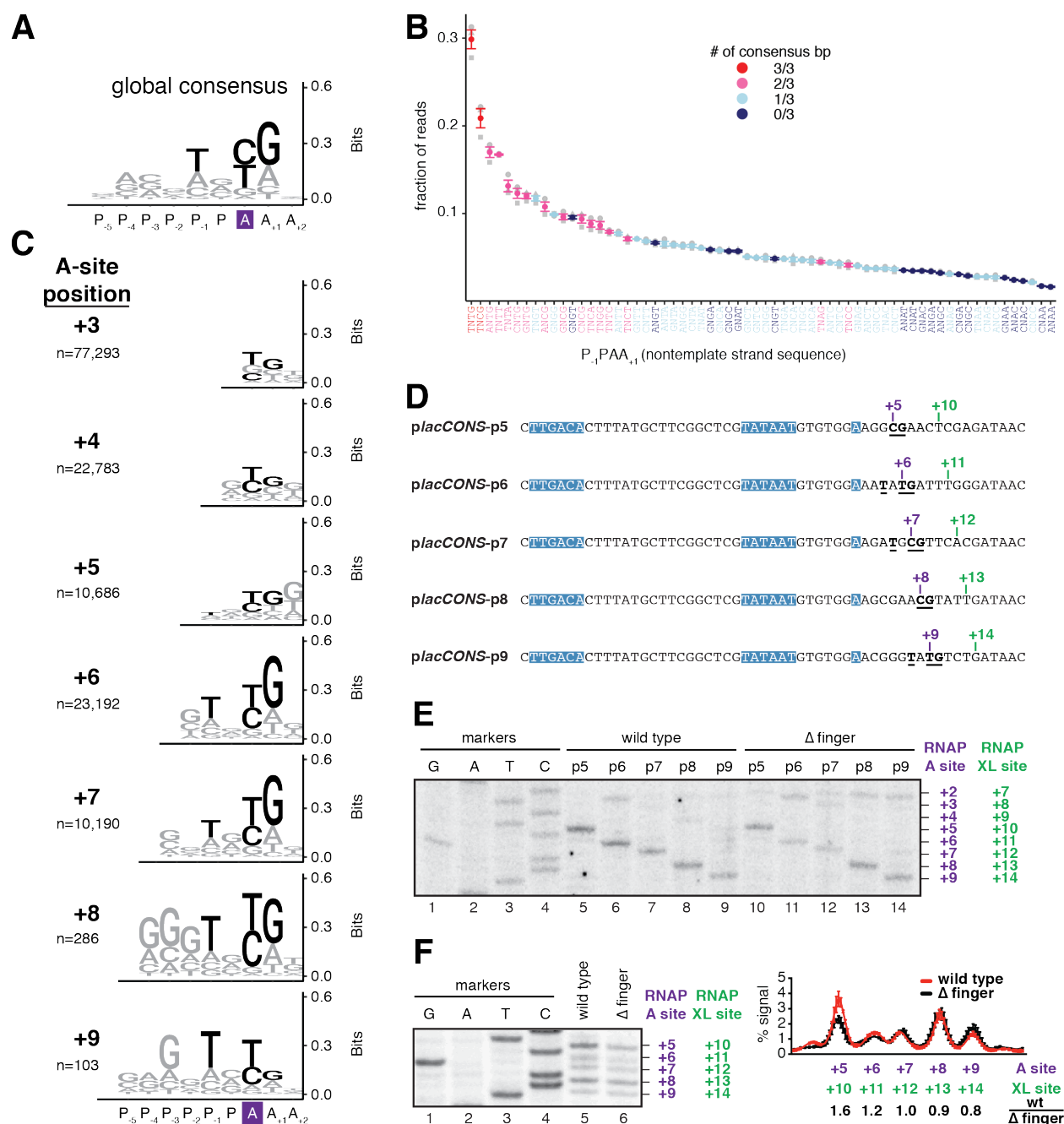


**Figure 4. Promoter-position dependence of initial-transcription pausing for a library of  $4^{11}$  (~4,000,000) promoters *in vivo***

(A) *placCONS* template library containing all possible sequences from promoter position +3 to promoter position +13 (*placCONS-N11*).

(B) Left, PAGE analysis of RNAP-active-center A-site positions *in vivo*. Right, histogram showing signals detected in the absence (black line) or presence (grey line) of Rif. Markers, DNA sequence ladder generated using *placCONS-N11*.

(C) Position of RNAP-active-center A-site (purple) and nucleotide crosslinked to Bpa (green) defined relative to the TSS position, +1.



**Figure 5. Promoter-sequence determinants for initial-transcription pausing in a library of  $4^{11}$  (~4,000,000) promoters *in vivo***

(A) Position-independent, global, consensus sequence for initial-transcription pausing

(B) RNAP occupancy at each initial-transcribed region tetranucleotide sequence. Red, pink, cyan, and blue denote pause-site sequences with 3 of 3, 2 of 3, 1 of 3, and 0 of 3 matches to the global consensus sequence. Mean  $\pm$  SEM (n = 3).

(C) Position-specific consensus sequences for initial-transcription pausing for RNAP-active-center A-site positions +3, +4, +5, +6, +7, +8, +9.



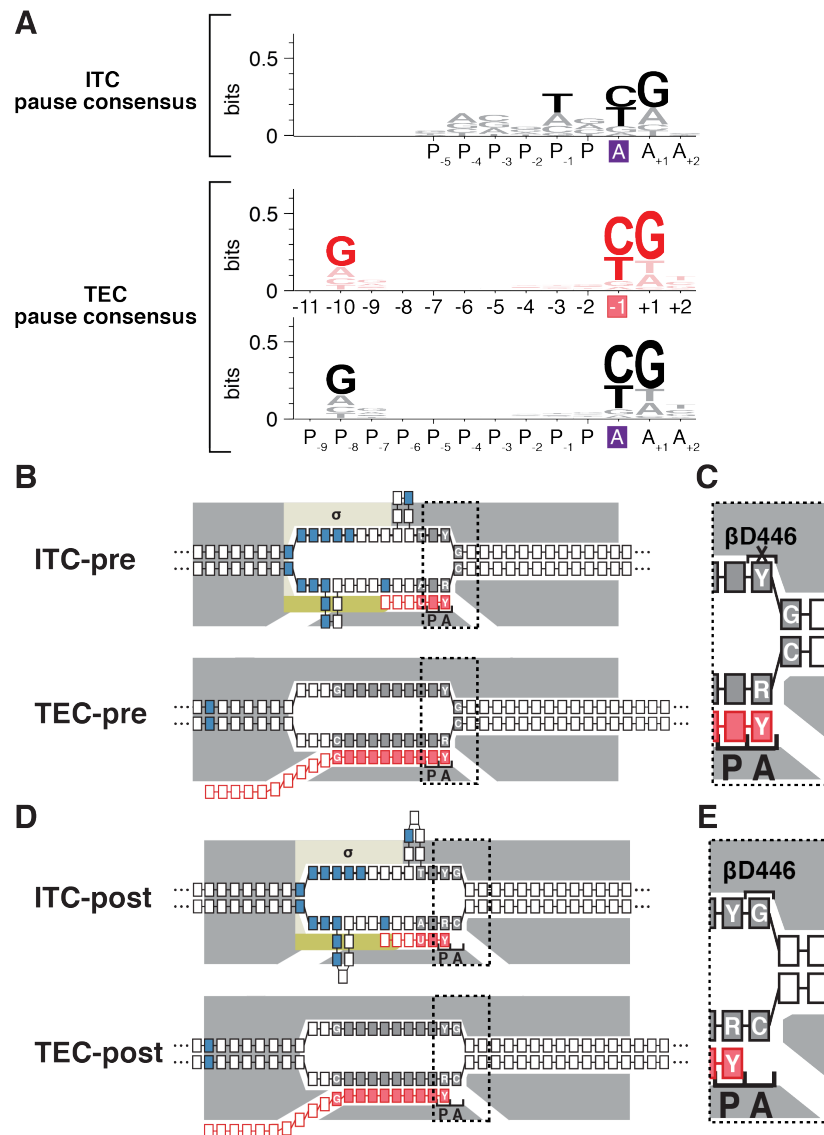
(D) Representative sequences yielding high RNAP occupancy at RNAP-active-center A-site positions +5, +6, +7, +8, +9, respectively. Colors as in Figure 4C.

(E) RNAP-active-center A-site positions and crosslinking positions *in vitro* for the sequences of (C) with high RNAP occupancy at RNAP-active-center A-site positions +5, +6, +7, +8, +9.

Lanes 1-4 present sequence markers, lanes 5-9 present data for wild-type RNAP, and lanes 10-14 present data for an RNAP derivative containing a deletion of the  $\sigma$  finger ( $\Delta$  finger).

(F) RNAP-active-center A-site positions and crosslinking positions *in vitro* in which sequences of (C) are pooled and analyzed as a pool. Left, gel; right subpanel presents quantitation of RNAP occupancies at +5, +6, +7, +8 and +9, with data for wild-type RNAP in red and data for  $\Delta$  finger RNAP in black.





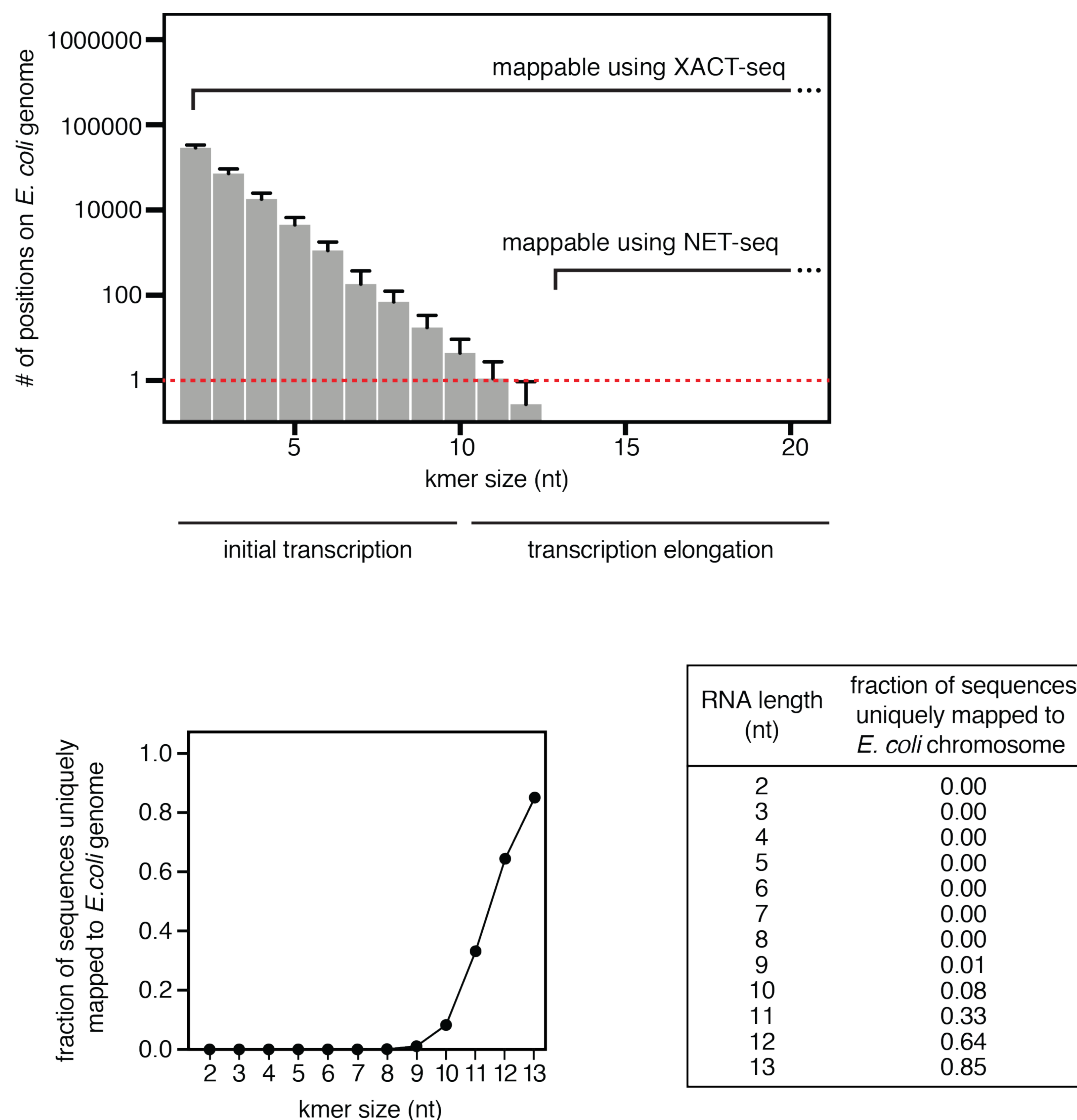
**Figure 6. Relationship between sequence determinants for initial-transcription pausing and sequence determinants for transcription-elongation pausing**

(A) Consensus sequence for initial-transcription pausing. Positions are labeled relative to RNAP-active-center A-site (purple box) and P-site.

(B) Consensus sequence for transcription-elongation pausing. Positions are numbered relative to the RNA 3' end (pink box) in upper axis label and numbered relative to RNAP-active-center A-site (purple box) and P-site in lower axis label.

(C) Positions of consensus-sequence nucleotides in pre-translocated-state complexes in initial-transcription pausing (ITC-pre) and transcription elongation pausing (TEC-pre). Grey boxes with white lettering, high-information content DNA nucleotides of consensus sequence; grey boxes without white lettering; low-information content nucleotides of consensus-sequence; Pink boxes with white lettering, high-information content RNA nucleotides of consensus sequence. Dotted rectangle, region enlarged at right. Other symbols and colors as in Figure 1.

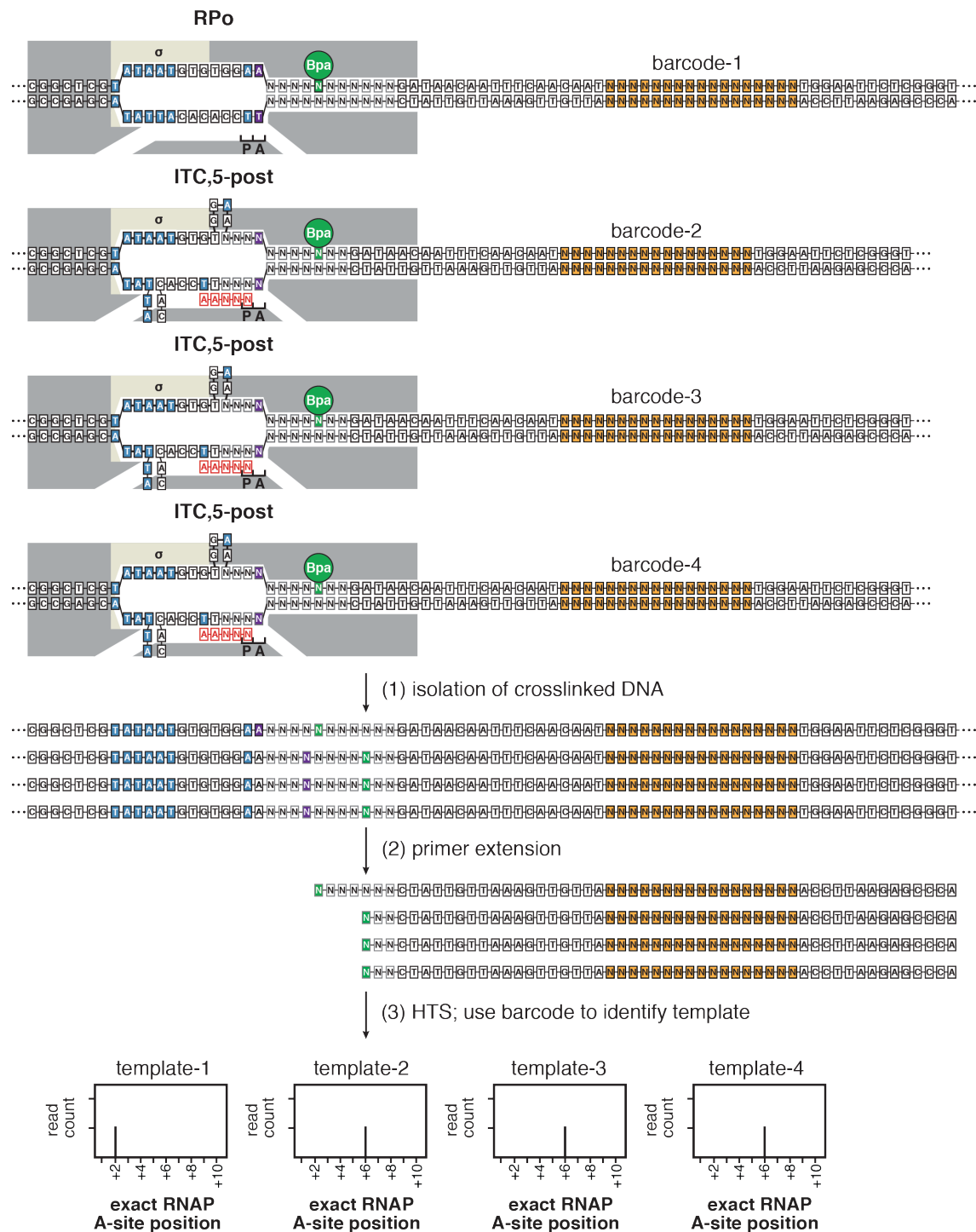
- (D) Enlarged view of RNAP-active-center region of pre-translocated state complexes, showing unfavorable interaction of RNAP  $\beta$  residue D446 with nontemplate-strand pyrimidine (Y) at A-site (crossed out bracket labeled  $\beta$ -D446).
- (E) Positions of consensus-sequence nucleotides in post-translocated-state complexes in initial transcription pausing (ITC-post) and transcription elongation pausing (TEC-post).
- (F) Enlarged view of RNAP-active-center region of post-translocated state complexes, showing favorable interaction of RNAP  $\beta$  residue D446 with nontemplate-strand purine (R) at A-site (bracket labeled  $\beta$ -D446).



**Figure S1. Inapplicability of RNA-based sequencing methods, such as NET-seq, for analysis of initial transcription**

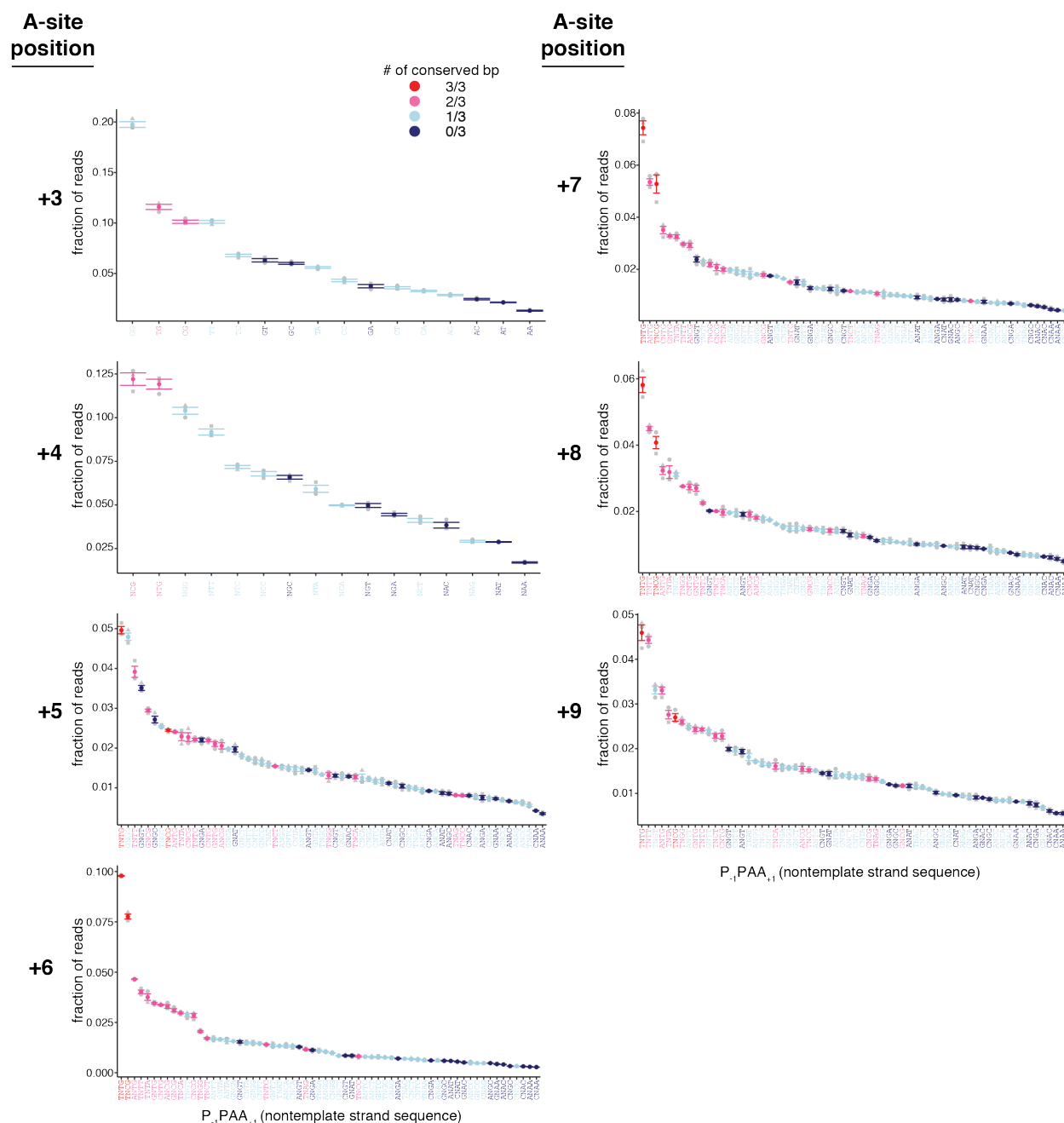
Relationship between RNA product length and number of occurrences of sequence on the *E. coli* genome. Top, plot of mean  $\pm$  SD of the number of occurrences of sequences on the *E. coli* genome as a function of RNA length. Bottom, fraction of sequences uniquely mappable to the *E. coli* genome as a function of RNA length. Because essentially all RNA products  $\leq 10$  nt, and a substantial number of RNA products  $\leq 13$  nt, have more than one occurrence on the *E. coli* genome, and because many RNA product sequences  $\leq 13$  nt are not uniquely mappable to the *E. coli* genome, RNA-based sequencing methods cannot be used to study initial transcription. In contrast, because RNAP-active-center A-site positions for RNA products of any length are mappable to the *E. coli* genome using XACT-seq, this method can be used for analysis of initial transcription.





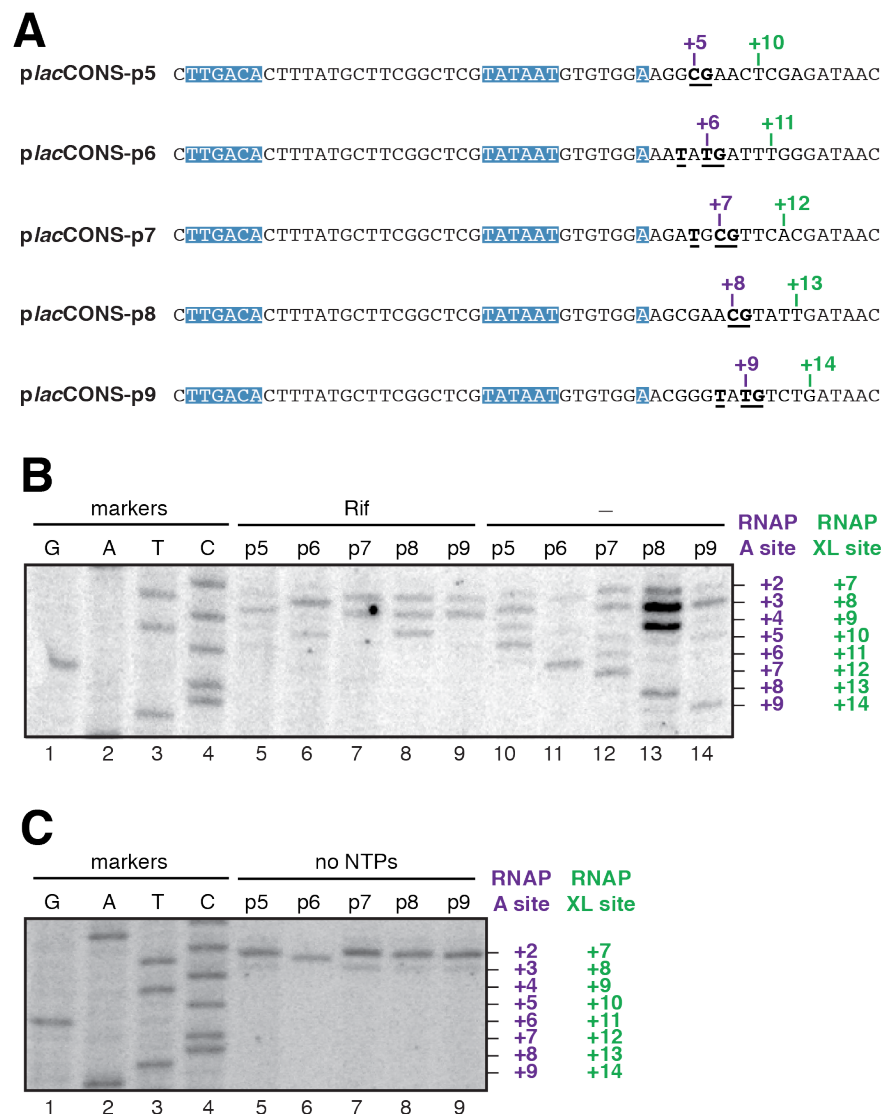
**Figure S3. Identification of RNAP-active-center A-site positions in initial transcription by XACT-seq**

Identification of RNAP-active-center A-site positions for transcription complexes formed on *placCONS-N11* template sequences. Grey boxes, nucleotides of the randomized initial-transcribed region (+3 to +13); orange-filled boxes, nucleotides of the randomized barcode region (+32 to +48). Other colors and symbols as in Figure 3C.



**Figure S4. Promoter-sequence determinants for initial-transcription pausing in a library of  $4^{11}$  (~4,000,000) promoters *in vivo*: additional data**

RNAP occupancy at each dinucleotide sequence (for RNAP-active-center A-site position +3), at each trinucleotide sequence (for RNAP-active-center A-site position +4), or at each tetranucleotide sequence (for RNAP-active-center A-site positions +5 to +9). Colors as in Figure 5C. Mean  $\pm$  SEM ( $n = 3$ ).



**Figure S5. Promoter-sequence determinants for initial-transcription pausing in a library of  $4^{11}$  (~4,000,000) promoters *in vivo*: additional data**

(A) Representative sequences yielding high RNAP occupancy at RNAP-active-center A-site positions +5, +6, +7, +8, or +9, respectively. Colors as in Figure 4C.

(B) RNAP-active-center A-site positions and crosslinking positions *in vivo* for the sequences of (A) with high RNAP occupancy at RNAP-active-center A-site positions +5, +6, +7, +8, or +9. Lanes 1-4 present sequence markers, lanes 5-9 present data for wild-type RNAP in the presence of the RNAP inhibitor Rifampin, and lanes 10-14 present data for wild-type RNAP in the absence of the RNAP inhibitor Rifampin.

(C) RNAP-active-center A-site positions and crosslinking positions *in vitro* for the sequences of (A) with high RNAP occupancy at RNAP-active-center A-site positions +5, +6, +7, +8, +9. Lanes 1-4 present sequence markers, lanes 5-9 present data for wild-type RNAP in the absence of NTPs.