Sequence insertion 3 of *Escherichia coli* RNA polymerase regulates transcriptional pausing

By

Yu Bao

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 07/27/2023

The dissertation is approved by the following members of the Final Oral Committee:

Robert C. Landick, Professor, Biochemistry and Bacteriology
Richard L. Gourse, Professor, Bacteriology
James L. Keck, Professor, Biomolecular Chemistry
Aaron A. Hoskins, Associate Professor, Biochemistry
Charlie Mo, Assistant Professor, Bacteriology

Sequence insertion 3 of *Escherichia coli* RNA polymerase regulates transcriptional pausing

Yu Bao

In the laboratory of Professor Robert C. Landick

University of Wisconsin – Madison

Madison, WI

Abstract

Gene expression across all three domains of life relies on multi-subunit RNA polymerase (RNAP), which transcribes DNA-encoded information into RNA with high quality and appropriate quantity. RNAP undergoes regulation *via* cis/trans regulatory elements throughout the various stages of transcription, including initiation, elongation, and termination. In addition to its primary function of synthesizing RNA rapidly and accurately, RNAP has evolved the ability to perceive signals from small molecules, regulatory factors, genomic sequences, and RNA structures. These signals enable transcriptional tuning in response to diverse cellular and environmental conditions, thereby facilitating processes of cellular competence, growth, and adaptation. Many of these regulatory events occur during the transcript elongation stage, where the productive elongation complex (EC) undergoes transition into a paused elongation complex (PEC), allowing regulation during the pausing time window.

The core structure and function of RNAP are highly conserved across cellular life, but functional diversity arises through the acquisition of point mutations and sequence insertions in key domains during evolution. Extensive research on transcriptional pausing has revealed that a lineage-specific sequence insertion known as SI3 in bacterial RNAP contributes to additional transcriptional regulation possibilities. Although the structural basis and physiological roles of transcriptional pausing in bacteria have been extensively characterized (as described in **Chapter 1**), less attention has been given to understanding the impact of SI3 on fundamental transcription processes. This work aims to establish an experimental system for studying the function of SI3 in transcriptional pausing both *in vitro* and *in vivo*.

Structural studies have revealed that SI3 assumes different positions within various RNA polymerase (RNAP) complexes, suggesting its flexible nature during the elongation stage of transcription. In **Chapter 2**, I will present the design of a Cys-triplet reporter (CTR) related to SI3 that enabled investigation of conformational changes. This

reporter incorporates three cysteine residues engineered at specific locations, that can form two distinct disulfide cross-linking species corresponding to different SI3 positions. By quantifying the relative amounts of these species, the positional bias of SI3 in different EC/PEC states can be determined. I also developed a derivative design of Cys-pair reporters (CPRs) to further restrict SI3 to extreme positions and test the impacts of this restriction. In addition to providing important insights into the function of this protein domain, this chapter highlights the optimization of the chemical cross-linking method by introducing hydrogen peroxide-catalase into the assay, significantly enhancing both cross-linking rate and efficiency.

Utilizing the Cys-triplet/pair reporter, I examined the positional biases of SI3 in various EC/PEC states. During the elongation stage, SI3 exhibited a concentration-dependent bias toward a closed conformation upon binding cognate nucleoside triphosphates (rNTPs). Interestingly, fixing SI3 in the closed conformation hindered efficient elongation, suggesting that SI3 undergoes obligatory open-close cycling during elongation. This cycling pattern connects the positional changes of SI3 to the conformational alterations of the catalytic active module, the trigger loop (TL). In pausing states, the positioning of SI3 exhibited biases dependent on the type of pause. For hairpin-stabilized pauses (hs-pauses), SI3 displayed a swiveled position, consistent with the structural studies of hairpin-stabilized pause complexes (hsPEC). However, for consensus elemental pauses, SI3 favored the closed position, revealing a novel conformation of the paused complex (see **Chapter Appendix A**). Confirming the role of SI3 in stabilizing the pause, fixing SI3 in the swiveled position supported the unfolded conformation of TL. Similarly, fixing SI3 in the closed position enhanced elemental pauses, potentially by stabilizing a pre-translocated register. Additionally, I identified a residue, β′F1199 in the jaw domain, that may mediate global conformational changes of RNAP through SI3-TL movements in hsPEC. This residue likely facilitates interactions between the jaw and SI3 domains via a hydrophobic "Phe pocket" on the surface of SI3. Mutating β′F1199 to alanine significantly reduced the strength of hs-pauses and the positional propensity of SI3 toward the swiveled position, highlighting the importance of the jaw-SI3 interaction in hs-pause regulation.

To investigate the *in vivo* function of SI3, I performed deletion of SI3 from the *E. coli* genome using the CRISPR-assisted recombineering method (no-SCAR), as described in **Chapter 3.** Although SI3 deletion is lethal to *E. coli*, my approach led to isolating a viable strain with SI3 deletion along with a point mutation (β′A941T-ΔSI3, ΔSI3*). In-depth *in vivo* characterization revealed the essential nature of the β′A941T mutation for the viability of the SI3-deleted strain. Biochemical assays demonstrated that SI3-deleted RNAP exhibited a reduced elongation rate

compared to the wildtype RNAP, but the β′A941T mutation restored a normal elongation rate. Interestingly, the mutant strain displayed smaller cell size but a comparable doubling rate to the wildtype strain during the mid-log phase. Preliminary phenotype microarray analysis indicated significant growth differences between the mutant strain and the wildtype strain under various culture conditions. Based on these findings, we propose that the deletion of SI3 in *E. coli* systematically alters the gene expression profile, leading to global changes in the physiological state of the cell.

Considering that the major alteration observed in SI3-deleted RNAP *in vitro* was a change in pause kinetics, we hypothesize that transcriptional pausing is affected *in vivo*. In **Chapter 3**, I describe the application of the native elongating transcript sequencing (NET-seq) method to investigate pause profiles in the SI3-deleted strain (ΔSI3*). The results of my work demonstrate that the SI3-deleted strain exhibits significant alterations in pause patterns within mRNA regions, with a lesser impact on rRNA regions. Comparative analysis of the altered pause signals between the wildtype (WT) and ΔSI3* datasets allowed categorization of these signals into ΔSI3-enhanced pauses and ΔSI3-suppressed pauses. The consensus sequence of the ΔSI3-enhanced pauses closely resembles the consensus elemental pause signal, while ΔSI3-suppressed pauses exhibit characteristics of both consensus elemental pausing and backtracked pausing. Investigation of the backtracked signals indicates that their origin lies in cleavage events during library preparation or the establishment of a pause signal *en route* to classic elemental pause sites. To further investigate to the hypothesized pause-stabilizing effects of the SI3 domain, we examined the ΔSI3-suppressed pause signals to identify potential hairpin-stabilized pauses based on predicted RNA secondary structures in the RNA exit channel. Several candidates for putative hairpin-stabilized pauses were validated through *in vitro* experiments. The abundance of these putative pauses within open reading frame (ORF) regions prompted further examination of the *in vivo* effects of these pauses on gene expression levels, and significant alterations were detected. The NET-seq results reveal the great potential of SI3 domain in regulating pause strength and pause pattern *in vivo*.

Collectively, my thesis research work provides new insights into the role of SI3 in shaping the pausing pattern in *E. coli* and highlights new features of hs-pause signals across the genome as well as potential regulatory functions of these pauses. The cysteine-H_2O_2 cross-linking system developed here could be applied to study other flexible systems in large protein complexes. Further, the use of NET-seq analysis of a mutant RNAP is an effective strategy that could be applied to study the function of other modules of RNAP.

Table of Contents

Chapter 1: Introduction to transcriptional pausing and SI3 domain

Adapted from:

Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proc Natl Acad Sci U S A* **118** (2021).

1.0 Why we study transcriptional pausing

Transcription plays a pivotal role in gene exprsssion by governing the quantity and variety of expressed genes through the context-dependent and time-sensitive generation of diverse RNA species. Across all cellular life forms, transcription proceeds through sequential stages of initiation, elongation, and termination. During initiation, a multi-subunit DNA-dependent RNA polymerase (RNAP) core enzyme assembles with initiation transcription factors to form a holoenzyme that can recognize and initiate transcription at specific DNA regions referred to as promoter regions. Subsequently, in the elongation stage, the RNAP elongation complex (EC) dynamically traverses DNA, utilizing ribonucleoside triphosphate (rNTP) molecules as substrates and producing single-stranded RNA. Finally, in the termination stage, the EC is inactivated and the RNA product is released (**Figure 1.1**). While the elongation process is generally efficient (~20–90 nt/second) (1, 2) at most DNA positions, RNAP encounters delays (ranging from seconds to minutes) at specific DNA sites known as pauses (3, 4). Work in recent decade to explore and elucidate transcriptional pausing has significantly enriched our comprehension of the intricate function of the transcriptional machinery. Pausing is not merely an offline state of the EC but realizes diverse regulatory functions in transcriptional processes: (i) in bacterial amino acid biosynthetic operons, pausing coordinates transcription with the trailing ribosome, facilitating the activity of attenuator elements in leader peptide regions (5, 6) (Figure 1.2); (ii) pausing provides temporal windows for the maturation of ribosomal RNAs (rRNAs) and riboswitches (7-10), allowing the formation of RNA secondary structures; (iii) pausing enhances transcription fidelity by enabling backtracking and recruiting cleavage factors in the event of misincorporation (11, 12); (iv) pausing ensures the preservation of genome integrity by recruiting the transcription-coupled DNA repair machinery upon encountering DNA damage (13, 14); (v) pausing facilitates the formation of the bacterial intrinsic termination complex (15); and (vi) pause strength can be modulated by global pause regulators such as ppGpp and H-NS (16-19) with global changes to transcriptional efficiency.

Despite considerable advancements, numerous intriguing questions surrounding the role of transcriptional pausing in tuning gene expression remain unsolved. Of particular interest is the impact of RNAP lineage-specific sequence insertion domains on the pausing process in distinct bacterial lineages. These protein insertions, acquired during the evolutionary course of bacterial RNAPs, dramatically influence pause kinetics *in vitro* and potentially modulate global pause profiles and gene expression levels. The aim of this study was to investigate the role of the

sequence insertion 3 (SI3) domain in *E. coli* RNAP and its contribution to transcriptional pausing through a range of experimental approaches.

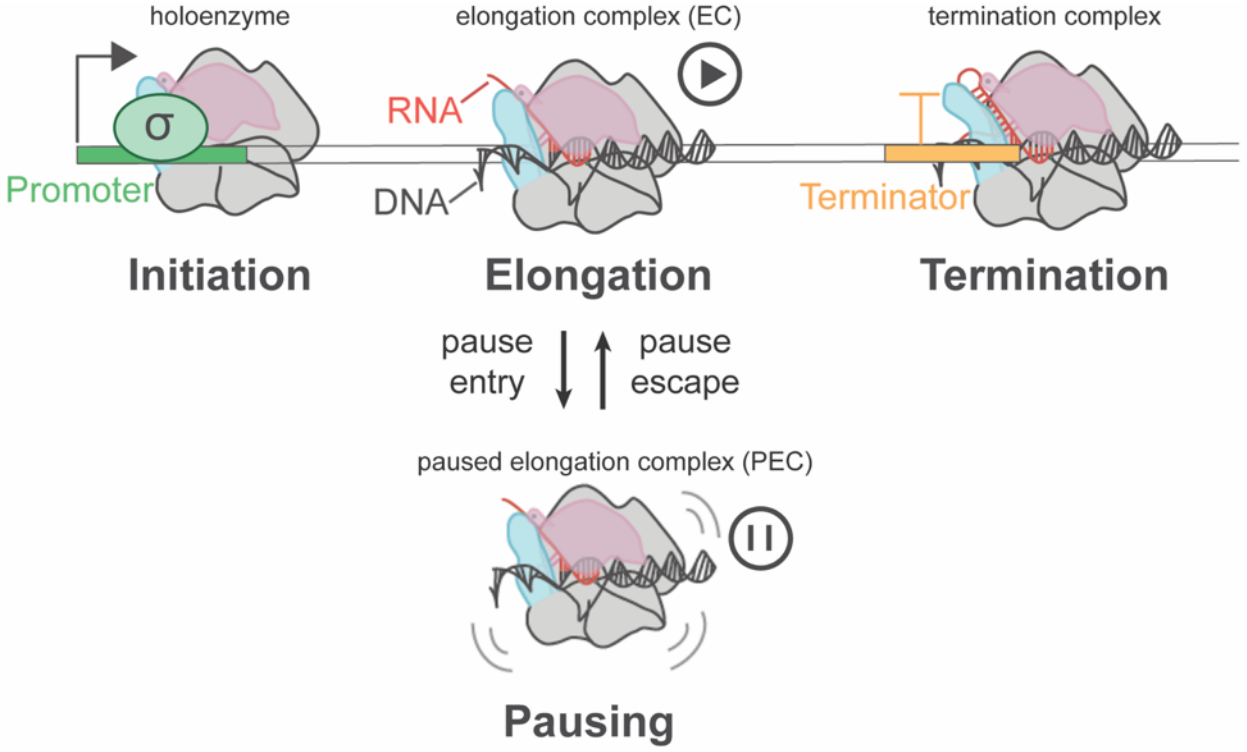


Figure 1.1. Schematic of the stages of bacterial transcription.

Bacterial transcription proceeds through stages of initiation, elongation, pausing, and termination. To initiate transcription, a sigma (σ) factor assembles with RNAP core enzyme into a holoenzyme, which recognizes the specific promoter region of a gene to be transcribed. After initiation, the core enzyme proceeds into the elongation stage with high processivity and the sigma factor normally disassociates. When encountering a pause signal in DNA, the elongation complex enters a pausing state to halt from seconds to minutes. Transcription is terminated when the transcribing enzyme reaches a termination site. Bacteria termination is normally mediated by sequence-specific intrinsic termination or assisted by rho termination factor. These events are all subjected to numerous regulatory mechanisms.

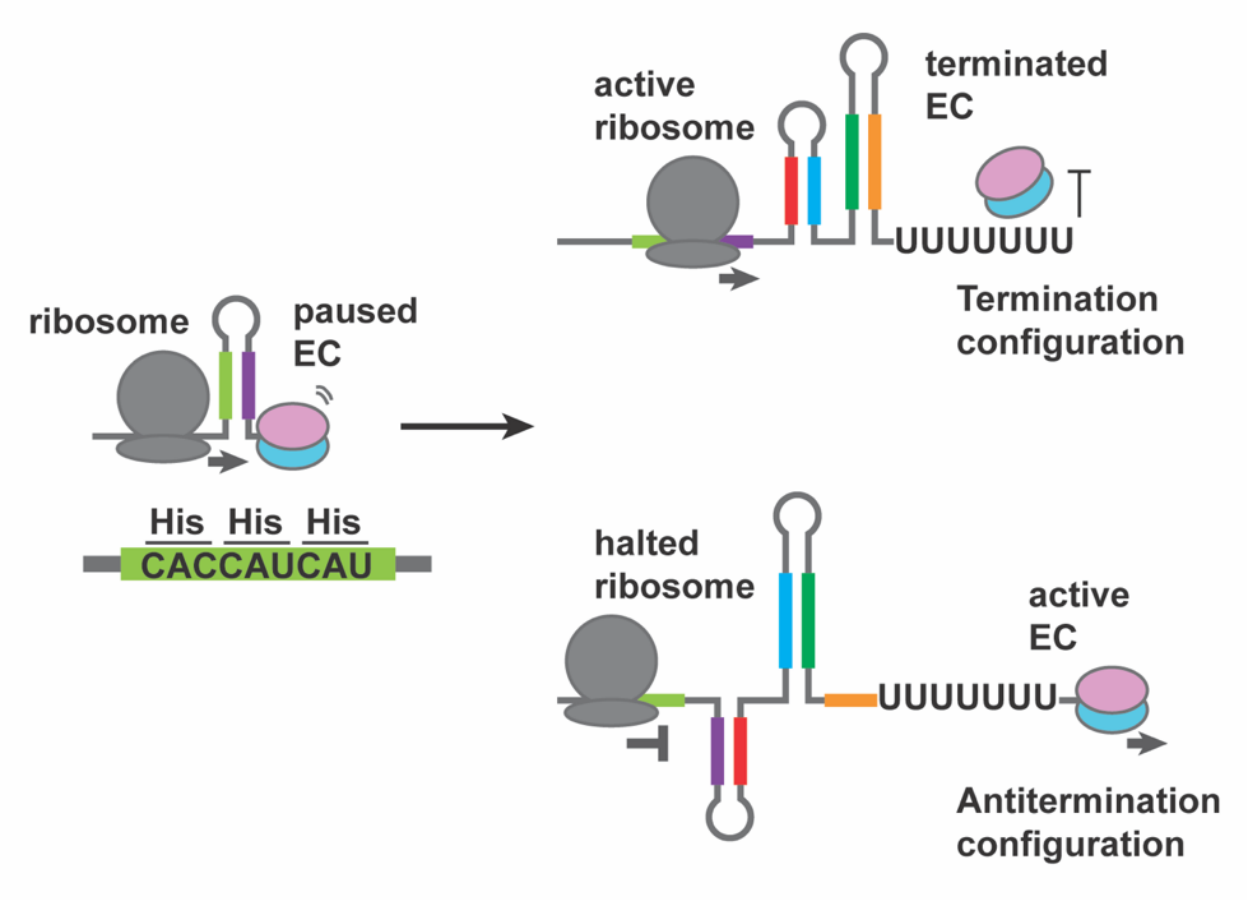


Figure 1.2. Regulatory roles of transcriptional pausing in transcription attenuation in *his* leader region.

Initial work on transcriptional pausing focused on the well-characterized hairpin-stabilized pause in the *his* leader region of the histidine biosynthetic operon. The paused EC can be reactivated by the trailing ribosome. The RNA folding patterns downstream of the pause site determine transcription attenuation of the biosynthetic genes. In the presence of efficient histidine-charged tRNAs (upper right panel), the active ribosome translates into the histidine codon region (light green) and allows formation of the termination configuration of the RNA structure, which contains an intrinsic termination signal and terminates transcription at the U-tract. With insufficient histidine-charged tRNAs (lower right panel), the ribosome halts and an alternative antitermination RNA configuration is formed, allowing EC to transcribe downstream genes.

1.1 *E. coli* RNA polymerase elongation complex overview

Most pausing events are associated with an elongation complex (EC) that encounters pause signals during the elongation stage (4). An EC is composed of a core RNAP and associated nucleic acids (Figure 1.3). A ten-base-pair double-stranded DNA molecule segment is melted into single strands inside the core and the template strand DNA (t-DNA) is used to guide RNA synthesis. RNA synthesis is comprised of repetitive cycles of biochemical reactions (aka the nucleotide addition cycle, NAC) (**Figure 1.4**), with one rNTP molecule consumed in each cycle and 1-nt extension of nascent RNA (20-22). This reaction is catalyzed by two magnesium ions, one stably held in the active center by a triad of aspartate residues and the other brought by the substrate rNTP triphosphate group. The substrate rNTP molecule enters the active center via the conserved secondary channel and binds the active site via Watson-Crick base pairing with the t-DNA base as well as polar interactions with surrounding residues in the binding pocket (23). The binding of the rNTP substrate triggers the chemical reaction of nucleotide addition, which is accelerated by multiple conformational changes of RNAP. A highly conserved, flexible trigger loop (TL) folds into a helical hairpin (trigger helix, TH) that forms a three-helix bundle with the bridge helix (BH) to close the active site (**Figure 1.5**). The side chains of the TH make specific contacts with the bound NTP, accelerating S_N2 reaction between the RNA 3′ hydroxyl group and the NTP α-phosphate by a factor of 10^4 (24). At this time point, with the NTP substrate incorporated into nascent RNA and no binding pocket available, the EC is in a pre-translocation state. The following translocation of RNAP in the downstream direction on the template DNA-RNA duplex and non-template DNA by 1 nt exposes a new substrate binding site and the EC is reset at the post-translocation state for the next round of nucleotide addition (20). The transition to the post-translocation state also involves TH unfolding, pyrophosphate release, and other uncharacterized RNAP conformational changes.

During elongation, EC can interact with multiple protein factors to regulate transcription. The conserved NusG transcription factor inhibits RNAP backtracking and promotes elongation in *E. coli* (25-27) but promotes pausing in *M. tuberculosis and B. subtilis* (28, 29). NusG can also assist rho-dependent termination by recruiting the rho termination factor (30-32). NusG's *E. coli* paralog RfaH makes specific interactions with the non-template strand DNA (nt-DNA) of the *ops* pause sequence to suppress both backtracking and hairpin-stabilized pause (26, 33, 34). Alone, the NusA transcription factor enhances hairpin-stabilized pause by chaperoning RNA hairpin structure outside the RNA exit channel (35, 36) but suppresses pause hairpin formation by reshaping the RNA exit channel in multiple antitermination complexes (37, 38). NusG and/or NusA facilitate recruitment of the trailing ribosome to the

EC to form a transcription-translation coupling complex, which may alleviate pausing events and prevent rho-dependent terminating and R-loop formation (39-41). When misincorporation occurs in transcription elongation, the stalled EC will be rescued by GreA/B factors by cleaving off the backtracked RNA fragments (11, 12). When a DNA damaging event occurs, the transcription-coupled repair (TCR) system will be recruited and assembled with the EC to fix the modified DNA bases by nucleotide excision repair (13).

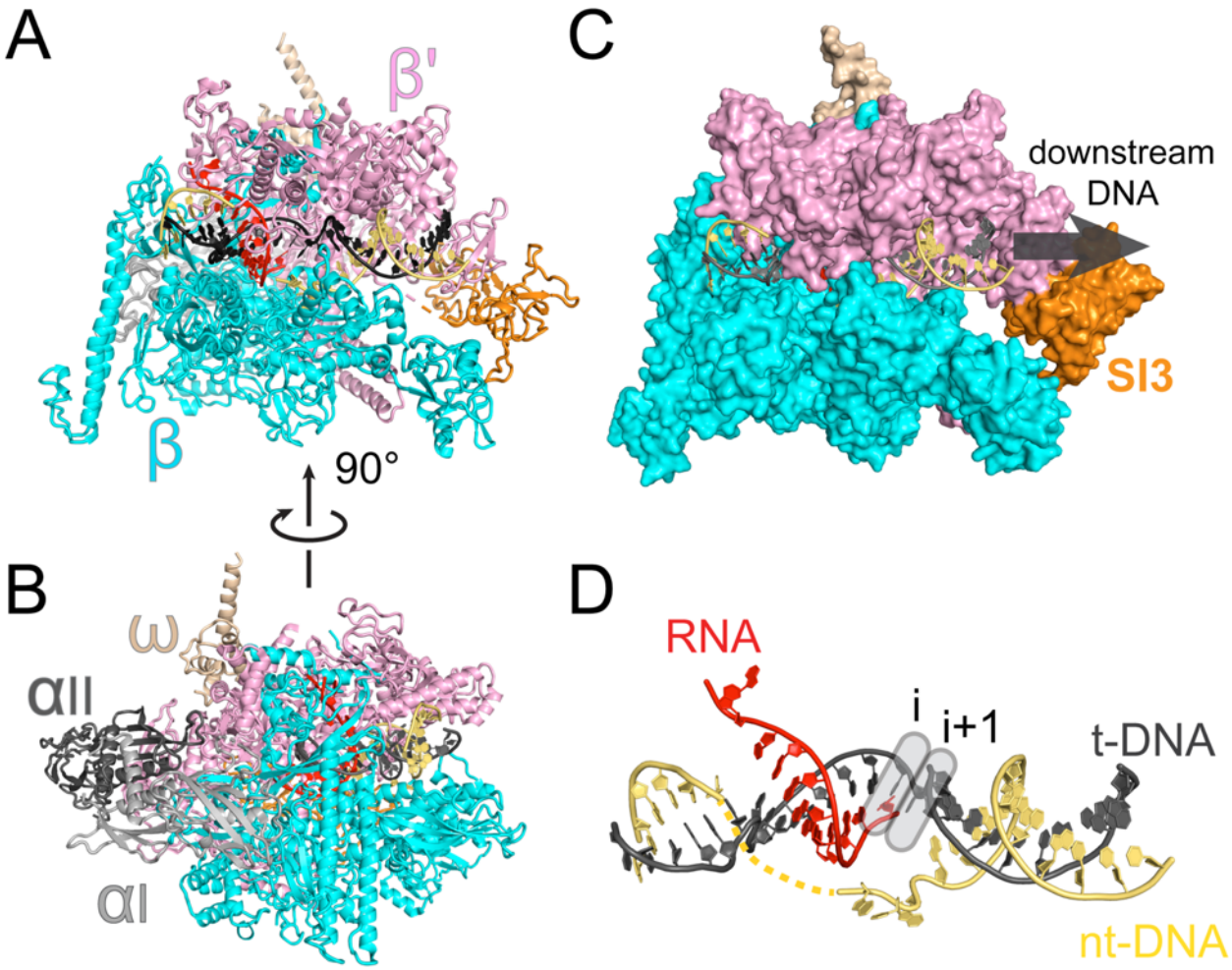


Figure 1.3. Structural overview of an EC (pdb: 6alf) (42).

(*A*) and (*B*) RNAP core enzyme is composed of 5 subunits, β (cyan), β′ (pink), αI (light grey), αII (dark grey) and ω (wheat). Nucleic acid scaffolds are embedded inside the groove formed by β and β′ subunits. The SI3 domain of β′ subunit is shown in orange. (*C*) Surface view of the EC structure. The downward movement direction of the EC during transcription on the DNA is indicated by the arrow. (*D*) Structures of the nucleic acid scaffold in the EC. Grey boxes indicate the location of i and i+1 sites in the active center. In this structure, the EC is in the post-translocated state and the 3′ end of the RNA resides in the i site. The unresolved part of the nt-DNA is represented in dashed lines.

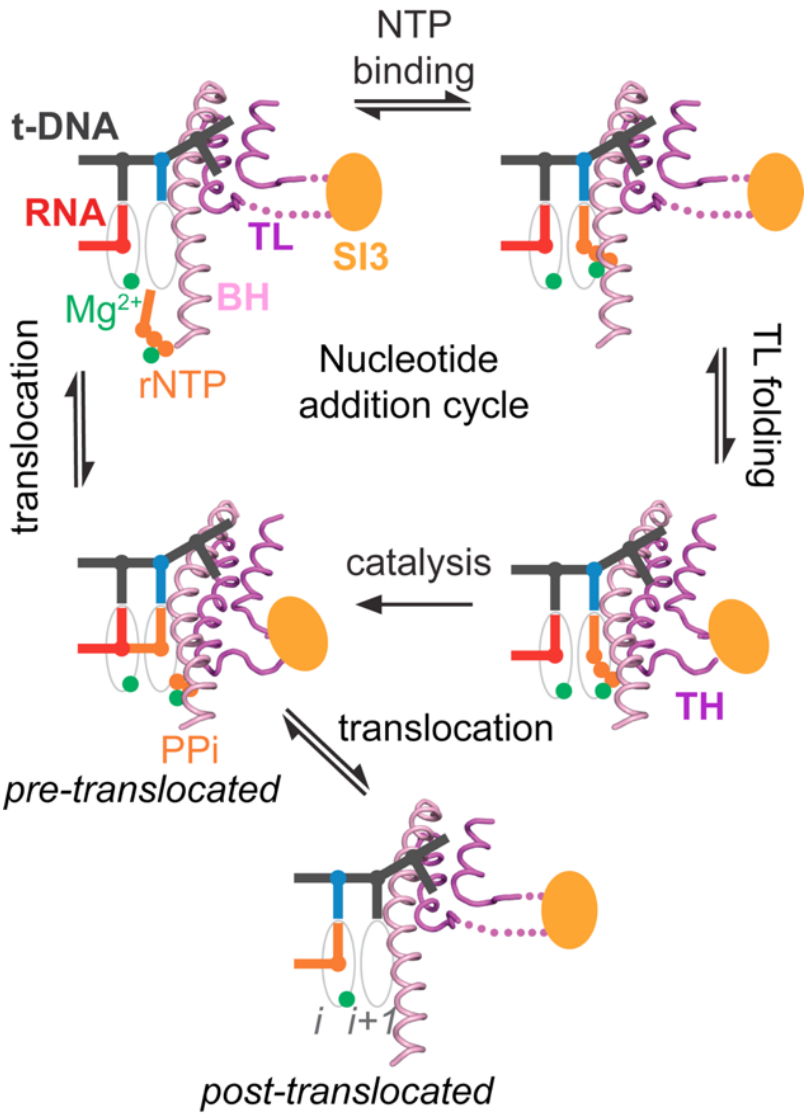


Figure 1.4. Schematic view of the nucleotide addition cycle (NAC).

The nucleotide addition cycle depicted in the above figure commences from a post-translocated register, wherein a base of t-DNA is loaded into the i+1 site. Subsequent to the binding of rNTP, interactions between the substrate and the unfolded trigger loop (TL) give rise to a conformational alteration of TL, resulting in the formation of the trigger helices (TH). The side chains present in TH function as positional factors that facilitate the catalysis of the nucleotide addition reaction. Conformational change of the BH also contributes to this process (43, 44). Following nucleotide addition, the pre-translocated register experiences translocation by 1 nt, thereby enabling the loading of the template base for the subsequent iteration of the nucleotide addition cycle.

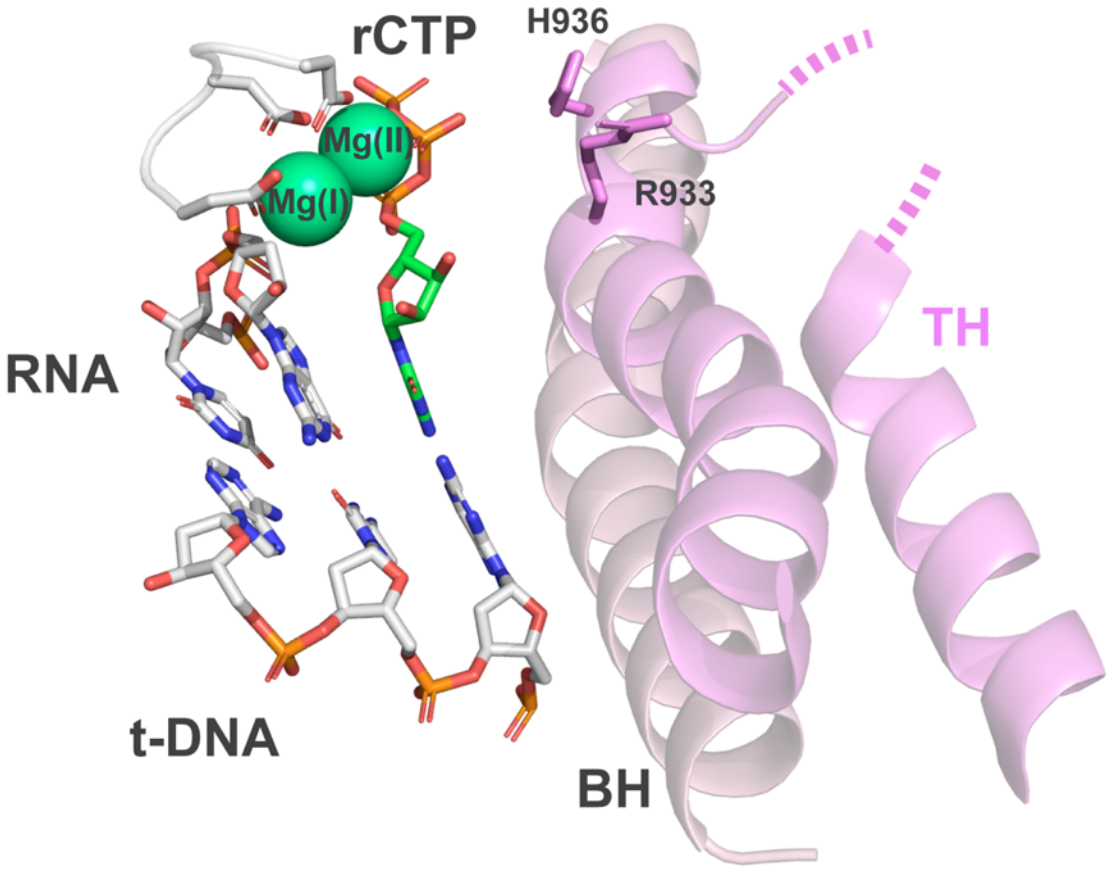


Figure 1.5. The active center upon rNTP binding.

The image (modified from pdb: 6rh3) (11) shows a rCTP molecule in the binding pocket poised for S_N2 reaction between the 3′ hydroxyl group of RNA and the α-phosphate of the rCTP. Note that in the actual structure, the 3’ hydroxyl group is removed to prevent reaction. H936 and R933, two conserved residues of TH, assist catalysis by positioning the substrate in the right configuration.

1.2 *E. coli* SI3 overview

The nucleotide addition process catalyzed by flexible TL/TH in prokaryotic, eukaryotic, and archaeal multi-subunit RNAPs indicates a highly conserved mechanism for RNA synthesis. However, certain bacterial phyla, such as proteobacteria, exhibit an interruption in the helical hairpin structure of TL due to lineage-specific sequence insertions of varying sizes. It is hypothesized that these insertions can influence diverse aspects of transcription and contribute to the regulation of transcription. In *E. coli* (Figure 1.6), the surface-exposed SI3 [β' (R943-G1130)] is connected to the TH helix2a section [β'(A941-T1135)] via two flexible linkers [linker1: β' (R943-S948), linker2: β' (I1124-G1130)] spanning a length of 14~18 Å (45). The major body of SI3 resides between the downstream DNA and the secondary channel, surrounded by rim helices (RH), sequence insertion 1 (SI1), and jaw domains. Structurally, SI3 consists of two Sandwich Barrel Hybrid Motifs (SBHMs), which are ancient and conserved motifs that typically serve as binding domains for biotin/lipoate in various enzymes and transporters (46). Unlike eukaryotes and archaea, TL sequence insertions are widely distributed in bacteria RNAPs. Different bacterial phyla possess TL sequence insertions with varying numbers of SBHMs. For instance, gammaproteobacteria, bacteroidetes, aquificea, and planctomycetes have TL sequence insertions composed of two SBHMs, while cyanobacteria can carry up to nine SBHMs within the TL sequence insertion (47, 48). In contrast, the two major phyla of gram-positive bacteria, actinobacteria and firmicutes, completely lack TL sequence insertions. Investigating the variability of TL sequence insertions among bacteria will enhance our understanding of bacterial evolution and adaptation.

In *E. coli*, SI3 has been shown to participate in multiple aspects of transcription. Alterations in SI3 or its binding by a monoclonal antibody significantly reduce the transcription elongation rate and intrinsic cleavage rate (49, 50). SI3 also plays a role in transcription initiation, as the open complex half-life of λP_R promoter is altered upon SI3 deletion (51). In cyanobacteria, the large SI3 region participates in extensive interactions with σ factor, promoting transcription initiation (48). SI3 is also involved in transcriptional pausing by interacting with the jaw domain, facilitating the stabilization of the *his* leader region hairpin-stabilized pause (*his* pause) (45). The swiveling conformational change of the paused elongation complex (EC) in *his* pause generates a steric clash between SI3 and a folded TL, thus preventing TL folding (52).

Structural studies of *E. coli* RNAPs suggest that SI3 occupies different positions in accordance with TL conformations in different states of ECs. In a transcription initiation complex (pdb: 4yln), TL is in the folded state

(TH), SI3 is brought to a closed position towards the direction of the secondary channel, and the tip loop (β′R1048–S1058) of SI3 is approaching RH (53). In a post-translocated EC (pdb: 6alf), TL is in an unfolded state and SI3 is in the open position, with the tip loop of SI3 sandwiched between SI1 and RH (42). In the *his* paused elongation complex (*his*PEC), TL is in an unfolded state and SI3, as part of the swiveling module, is in a swiveled position (see **1.2.2** for description of swiveling) with the tip loop near the SI1 domain (52). It remains unclear whether these positional changes reflect a movement pattern of SI3 during NAC and the formation of paused complexes, and more interestingly, whether the movement has implications for the function of SI3 in transcription regulation.

Despite the findings regarding SI3, several key questions remain unanswered. First, what is the movement pattern of SI3 during the conformational changes of TL in elongation and pausing? Second, what are the positional changes of SI3 in different types of transcriptional pausing? Third, how does SI3's existence/movement impact pausing strength and pausing regulation *in vivo*? Last, how can we explain the essentiality of SI3 in *E. coli* while its absence is observed in many bacterial phyla? Addressing these questions will undoubtedly provide insights into the mechanism of RNAP transcriptional pausing.

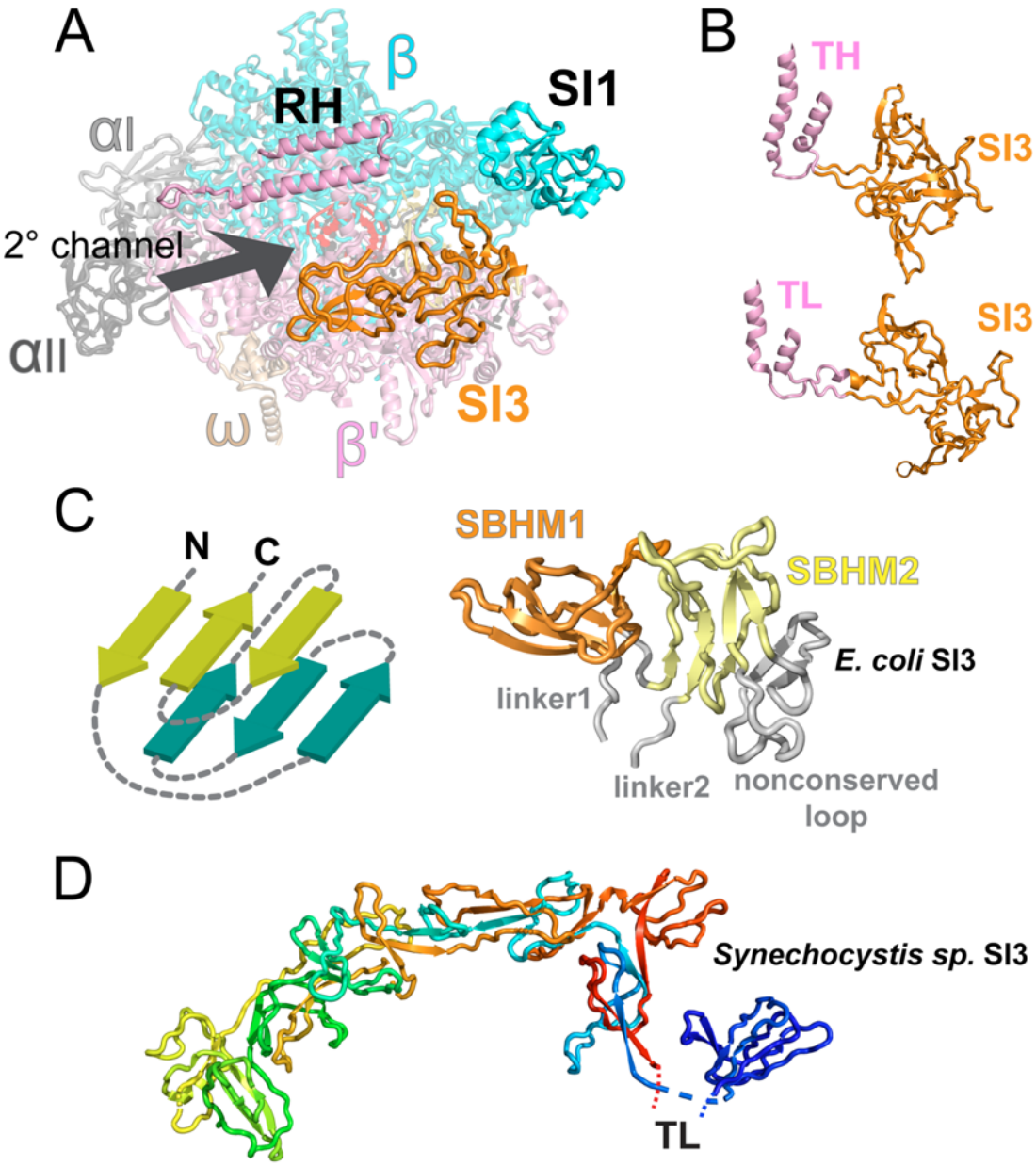


Figure 1.6. SI3 is composed of sandwich-barrel hybrid motifs (SBHMs).

(*A*) SI3 is surface-exposed and resides between RH and SI1 domains, gating the secondary channel (pointed by arrow). (*B*) SI3 domain is a sequence insertion in TL module. SI3 occupies distinct positions when TL is at different states. Structures are adapted from pdb: 8eg8 and pdb: 8gzg. (*C*) Left, the topology of SBHM, beta-sheet structures are represented with arrows and loops are represented by dotted lines. Right, *E. coli* SI3 contains two SBHMs. (*D*) SI3 of *Synechocystis* sp. PCC6803 contains nine SBHMs; only one shares sequence similarity with *E. coli* SI3 (pdb: 8gzg) (48).

1.3 Mechanisms of transcriptional pausing

The primary investigations into transcription mechanisms utilized biochemical assays, in which reconstituted transcription machinery can generate various RNA species from specific nucleic acid templates. RNA species are normally analyzed by separation on denaturing gels according to their lengths. In addition to full-length products, smaller RNA species are frequently detected, and these species are subsequently converted into larger RNAs with a half-life ranging from seconds to minutes (54). These discrete RNA species of specific sizes result from arrested elongation complexes (ECs) at particular DNA positions where RNA synthesis is halted. The existence of transcriptional pausing has been validated *in vivo* through the native elongating transcript sequencing (NET-seq) technique (55). NET-seq maps the 3′ end positions of nascent RNAs extracted from native ECs, thereby indicating the locations at which ECs tend to halt. In *E. coli*, NET-seq analyses revealed numerous pausing events across the genome (56, 57), suggesting that pausing is a common characteristic of ECs during gene expression. Structural studies with X-ray crystallography and cryo-EM revealed that paused ECs (PECs) undergo global conformational changes and adopt specific DNA-RNA duplex configurations (11, 35, 52, 58) that are distinct from those observed in the elongation state. Transcriptional pausing events are often associated with protein factors binding to RNA polymerase (35, 59-66). In the next section, we will review the proposed mechanisms of different types of transcriptional pausing.

1.3.1 Elemental pause

A generally accepted view about pausing is that most pause events arise from an elemental pause, in which elongating RNAP recognizes pause signals embedded in DNA sequence and enters an inactive state (**Figure 1.7**). The movement of RNAP on a DNA template during transcription elongation involves numerous physical and chemical processes: the melting (breaking of hydrogen bonds in Watson-Crick base pairs) of double stranded DNA at the downstream fork-junction (dsFJ), the annealing of DNA at the upstream fork-junction (usFJ), the melting of the DNA-RNA hybrid at the upstream fork-junction, the formation of RNA secondary structures at the RNA exit channel, the chemical energy transferred during nucleotide incorporation in the active center, and the polar and Van der Waals interactions between nucleic acids and RNAP. Extensive biochemical assays established that the elemental pause is multipartite, with dsFJ/usFJ/hybrid/dsDNA sequences all contributing to the pause strength. The utilization of NET-seq method revealed that pause sites are universally distributed across the *E. coli* genome with a

frequency of around one pause site per 100 base pairs (56). Alignment of strong pause sites yielded a consensus elemental pause sequence with a 5′-$G_{-11}G_{-10}t_{-3}g_{-2}Y_{-1}G_{+1}$ (–1 corresponds to 3′ end position of nascent RNA) motif, consistent with reported elemental pause sequences validated by single molecule experiments (67).

The first structure of the elemental paused elongation complex (ePEC, pdb: 4gzy) was determined using *Thermus thermophilus* RNAP (58). The paused configuration features a half-translocation state in the active center, in which the 3′ end of the RNA is translocated by 1 nt while the t-DNA backbone remains unmoved. The upcoming t-DNA base cannot load into the active site, as it is blocked by a kinked BH. This ePEC structure suggests a hinderance of translocation in nucleotide addition cycle is the primary explanation of the pause.

A cryo-EM structure of a *his* elemental paused elongation complex (*his*-ePEC, pdb: 6bjs) was solved without the *his* RNA hairpin (stem-loop secondary structure) that characterizes the *his* paused elongation complex (*his*PEC, pdb: 6asx). Comparison of this structure with that of the EC (pdb: 6alh) reveals minor swiveling (see **1.2.2** for description of swiveling) and a half-translocated DNA-RNA register (52). This implies swiveling and half-translocation might be common features of paused elongation complexes. However, some biochemical evidence suggests that elemental paused complexes may have alternative conformations. For example, *ops* elemental pause's susceptibility to pyrophosphate-induced cleavage implied a pre-translocated register (59). Reducing RNAP swiveling by biasing SI3 to the closed conformation via a disulfide cross-link enhanced the pause strength on a *his* elemental pause scaffold (68), indicating SI3 closure/TL folding contributes to the stabilization of specific elemental pause sequences.

According to recently published results of the comparison between two cryo-EM structures of *his* elemental paused elongation complex (*his*-ePEC) and consensus elemental paused elongation complex (*con*-ePEC), the two types of elemental paused complexes are both composed of multiple populations with distinct conformation distributions (69). Unlike previously reported PEC cryo-EM structures, *con*-ePEC was observed exclusively in the pretranslocated state. There were two major populations, with the major one (~55%) having a folded TL and SI3 in the closed position and the minor one (~45%) having an unfolded TL and SI3 in the open position. The *his*-ePEC structures reached higher resolution than previously determined ones, with analysis of more particles, thus providing more accurate information about conformations. The dominant state of this new *his*-ePEC was half-translocated (~73%) with unfolded TL and swiveled conformation. The less populated state (~27%) of *his*-ePEC was pre-

translocated with folded TL and SI3 in the closed position. The observation that a pretranslocated state is coupled with a folded TL in both ePECs indicates that DNA-RNA sequence-specific interactions stabilize TL in the folded state and translocation is prohibited in this state.

Biochemical results were also consistent with the distinct conformation distribution for the two ePECs. By using a Cys-triplet reporter (will be discussed in detail in **Chapter 2**) that probes multiple disulfide cross-link species of SI3 with peripheral domains, it was shown that SI3 is more biased toward the closed position when a *con*-ePEC was formed from *con*-ePEC$_{-1}$ by 1-nt extension. In contrast, SI3 is more biased toward the swiveled position when a *his*-ePEC was reconstituted in a similar manner. As for translocation states, *con*-ePEC was more resistant to NTP addition and more sensitive to pyrophosphate-induced RNA cleavage than *his*-ePEC, consistent with the cryo-EM results that *con*-ePEC was majorly pre-translocated while *his*-ePEC was majorly half-translocated.

Collectively, the results from cryo-EM and biochemistry assays indicate that multiple ePEC states exist in various ePECs and the DNA-RNA sequence differences in those ePECs altered the equilibrium among these ePEC states, resulting in the divergence of the kinetics of ePECs. The DNA-RNA sequence-induced energy barrier that inhibits successful translocation after nucleotide addition might be the essence of the elemental pause, and multiple ePEC states in this process offer more regulatory options to tune transcription.

1.3.2 Hairpin-stabilized pause

The hairpin-stabilized pause was first discovered and characterized in the leader region of gammaproteobacteria amino acid biosynthetic operons (5, 70). At the end of the leader region, RNAP will either be terminated or keep transcribing into structural gene regions. This decision is made by the coupling between the transcribing RNAP and the translation ribosome, with the formation of mutually exclusive RNA structures with or without a terminator structure. When the free amino acid produced by the operon is sufficient, the ribosome will elongate on the nascent RNA and facilitate the formation of the RNA structure with a terminator hairpin and thus induce termination. In this context, hairpin-stabilized pause signals play a pivotal role in synchronizing translation with transcription in transcription attenuation (see **Figure 1.2**).

Previous studies show that the nascent RNA stem-loop secondary structure (RNA hairpin) formed at the RNA exit channel stabilizes the pause, and the pause effect can be recapitulated with antisense DNA/RNA oligos to form duplex with nascent RNA that mimics the pause RNA hairpin (34, 71). The distance between the stem loop and the

3′ end of the nascent RNA also influences pause strength (34, 71). Taking the *his* pause as an example, the last base pair of the RNA hairpin forms at –12 position (–1 position corresponding to the 3′ end). When the RNA hairpin was altered to have larger or smaller distance from the 3′ end, the pause strength was dramatically reduced. Although not all hairpin-stabilize pauses position the RNA hairpin at –12, the interactions between RNA hairpin and RNAP force the ePEC to enter a new state with enhanced pause strength.

Two sets of cryo-EM structures revealed mechanisms underlying hairpin-stabilized pause in *E. coli* (Figure 1.8). The structure of *his*PEC provided insights into a novel global conformation rearrangement of RNAP called swiveling, which was not observed in earlier RNAP structures. The swivel module contains multiple domains including the clamp, shelf, dock, jaw, SI3 and β′ CTD, all of which undergo rotation along an axis parallel to the BH. The rotation causes significant positional change of the distal SI3 and impedes the folding of TL (52). Moreover, the *his*PEC has a half-translocated nucleic acid configuration, as seen in multiple paused complexes (35, 58, 72). Another *his*PEC structure includes a bound NusA factor that makes multiple interactions with RNAP β flap and αII-CTD regions and shields the RNA hairpin backbone with a positively charged surface (35).

Generally, studies on hairpin-stabilized pause have been restricted to specific cases with limited functional implications in gene regulation. Compared to elemental and backtrack pauses, hairpin-stabilized pauses may be less abundant and more enriched in 5′ UTR regions than in the bodies of genes (56). NET-seq does not distinguish among different types of pausing mechanisms, so cannot specifically capture hairpin-stabilized pauses across the genome. Instead, more advanced computational algorithms are needed for optimal RNA secondary structure prediction to probe hairpin-stabilized pause signals across the genome.

To address the gap between the limited study of hairpin-stabilized pauses and the great potential of these sequences in transcription regulation, in **Chapter 3** I describe work utilizing mutant RNAPs and NET-seq to probe putative hairpin-stabilized pause signals in the *E. coli* genome.

1.3.3 Backtrack pause

Early biochemical experiments unveiled anomalous footprints of elongation complexes (ECs) on the DNA template, giving rise to the concept of "inchworming" as a mode of EC movement during elongation (73). Subsequent investigations into EC elongation revealed that the apparent "inchworming" motion represents a distinctive state of ECs at specific DNA sites (74). In this state, the EC retrogresses, or backtracks, to extrude the 3′ end of the nascent

RNA, temporarily stalling until rescued by multiple factors. The phenomenon of backtrack pausing is widely observed throughout the eukaryotic genome, as evidenced by NET-seq results, where this kind of pausing has been found in both the promoter-proximal regions and bodies of genes (comprising 75% of total pause signals in yeast) (75).

In *E. coli*, backtracking pausing is commonly associated with translocation barriers. Backtracked species rapidly equilibrate with pre-translocated/half-translocated species upon the formation of the elemental paused complex (76). A well-documented case of this is observed in the operon polarity suppressor (*ops*) pause signal (59). The paused complex facilitates the recruitment of the anti-pausing factor RfaH, which binds to the EC clamp region, establishing sequence-specific contacts with nt-DNA (62). Backtracking can also occur when the misincorporation of non-cognate NTPs during the elongation cycle weakens the stability of the DNA-RNA hybrid at the 3′ end, resulting in backtracking of the EC and the threading of the RNA 3′ end toward the secondary channel. This form of backtracking pausing is rescued by the binding factors GreA/GreB, which promote cleavage of the mismatched 3′ nucleotides (77). Lastly, backtracking pausing emerges when the EC encounters DNA damage, such as DNA adducts and lesions. These obstacles are resolved by transcription-coupled DNA repair machineries (78). Collectively, backtracking pausing represents a prevalent form of pausing that not only relates to elemental pausing processes but also ensures the fidelity of transcription and the stability of the genome.

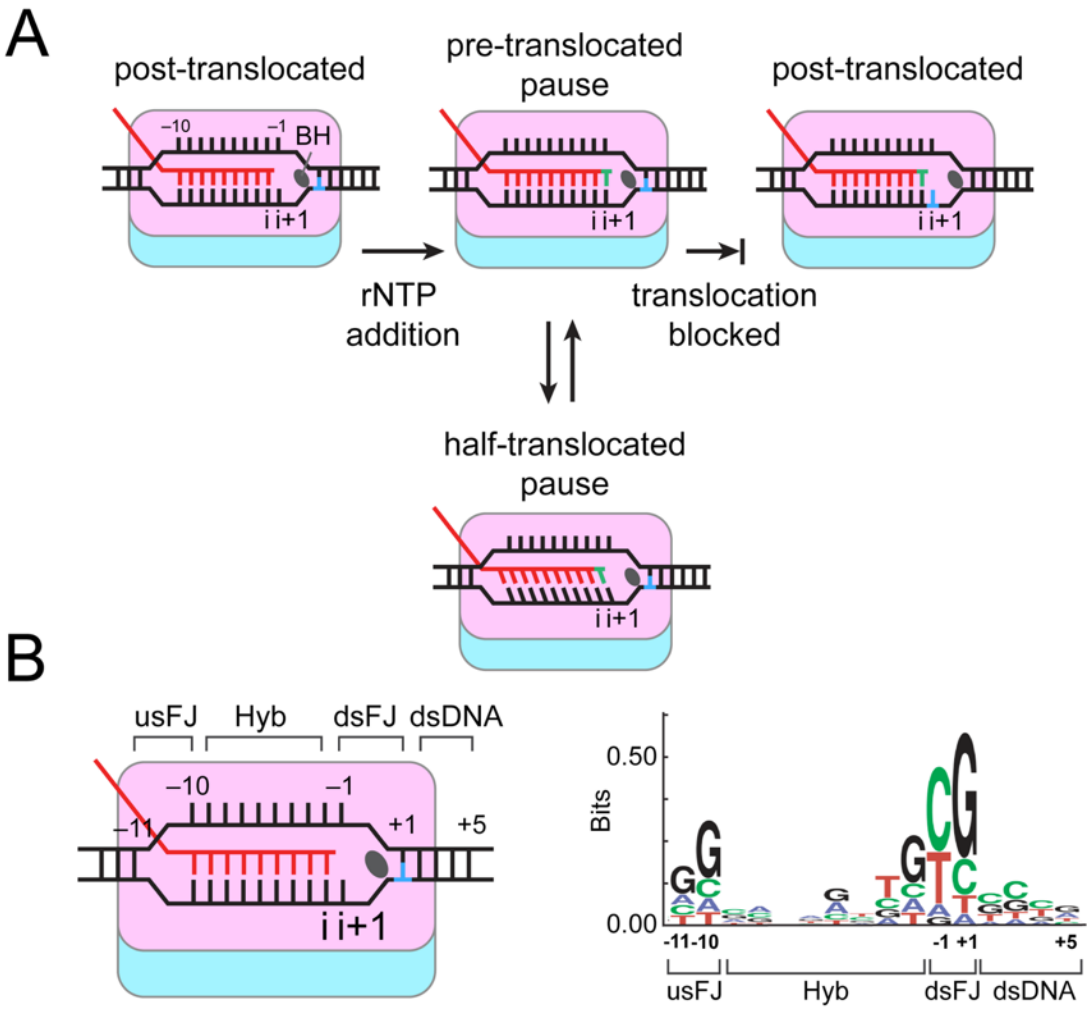


Figure 1.7. Elemental pause is an offline state of transcriptional elongation.

(*A*) EC enters the elemental paused state after nucleotide addition. Paused complexes can either stay in the pre-tranlocated state or in a half-translocated state. (*B*) The elemental pause signal is determined by multiple sequence factors. Upstream fork junction, usFJ; DNA-RNA hybrid, Hyb; downstream fork junction, dsFJ; downstream DNA, dsDNA. Right, the consensus sequence logo of elemental pause signals identified in Larson et al (56).

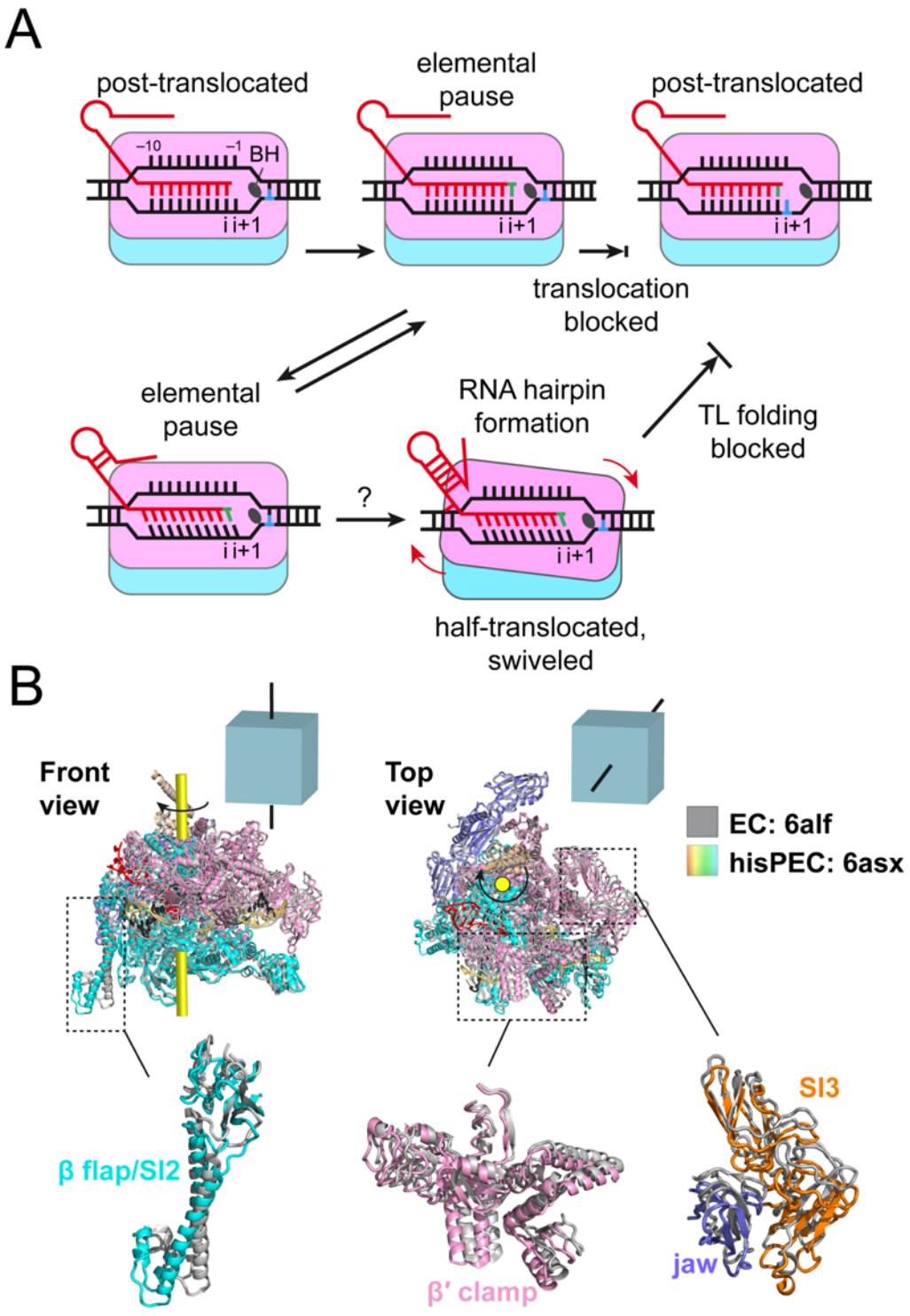
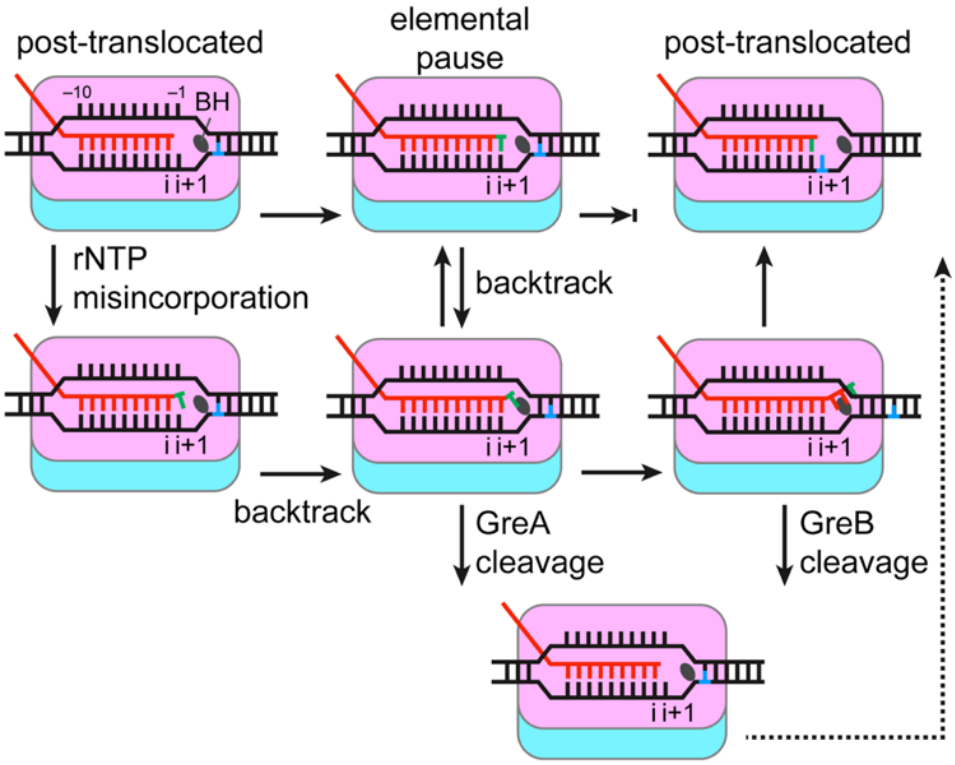

A
post-translocated
elemental pause
post-translocated
−10
−1
BH
i i+1
translocation blocked
elemental pause
RNA hairpin formation
TL folding blocked
?
half-translocated, swiveled
B
Front view
Top view
EC: 6alf
hisPEC: 6asx
β flap/SI2
β′ clamp
SI3
jaw

Figure 1.8. Hairpin-stabilized pause induces RNAP global conformational change.

(*A*) The formation of a hairpin-stabilized paused complex involves several steps. When an elongation complex (EC) reaches an elemental pause signal, an RNA stem-loop secondary structure is formed at the RNA exit channel within the time window. This RNA hairpin structure plays a crucial role in stabilizing the pause by potentially inducing a swiveling conformational change in the RNAP. As a result of this swiveling motion, the RNAP prevents the folding of the transcriptional leader (TL), thereby strengthening the pause. (*B*) The swiveling of the RNAP occurs in the aligned models of an elongation complex (PDB: 6alf) with a *his*PEC (PDB: 6asx). The swivel module undergoes a clockwise rotation along the axis parallel to the BH. The major components of the swivel module consist of the flap/SI2, clamp, and SI3/jaw regions, as shown in detail at the bottom.

A


post-translocated
elemental pause
post-translocated
−10
−1
BH
i i+1
rNTP misincorporation
backtrack
backtrack
GreA cleavage
GreB cleavage

B

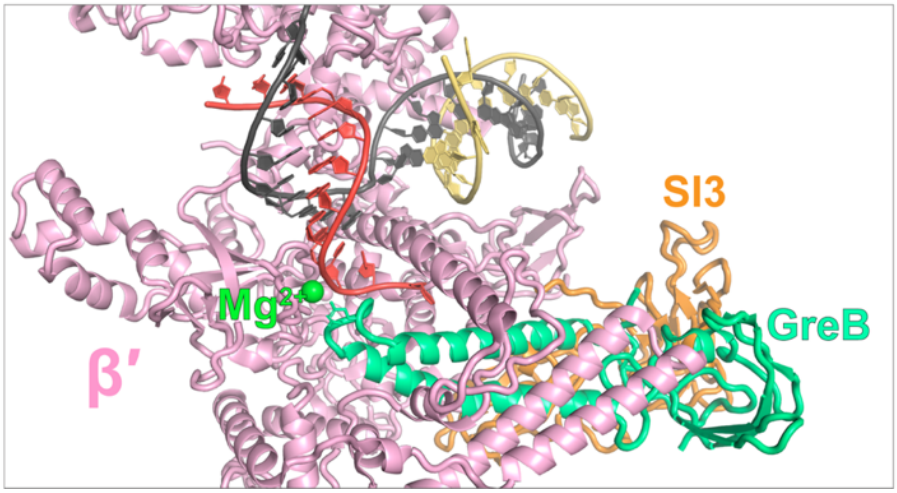

SI3
Mg2+
GreB
β′

Figure 1.9. Backtracked pause can be released by cleavage factors.

(*A*) The formation of a backtracked paused complex. When an elongation complex (EC) reaches an elemental pause signal or incorporates a noncognate rNTP, the EC halts and backtracks. GreA/GreB factors can access the active center via the secondary channel to promote RNAP's cleavage of the backtracked 3′ RNA. (*B*) The structure of GreB bound to a backtracked paused complex (PDB: 6rin). The α and β subunits are hidden in the presentation. GreB is sandwiched between SI3 (orange) and RH. The side chains of the two conserved residues responsible for cleavage catalysis are shown in GreB.

1.4 Key questions addressed in this work

To gain insights into the role of the SI3 domain of *E. coli* RNAP in transcriptional elongation and pausing, we designed a series of experiments to address the following questions:

First, we emplyed biochemical assays to examine the positional changes and movement patterns of SI3 in various states of the EC, including states in NAC and pausing states. Our strategy involved the development of a molecular reporter integrated with in vitro transcription assays, supported by a well-defined rationale and experimental design.

Second, we utilized genetic methods to engineer *E. coli* strains with altered or deleted SI3 domains, enabling us to investigate the changes in the profile of pausing events *in vivo*. This approach involved the construction of mutant strains, comprehensive phenotype characterization, and the utilization of the NET-seq technique. Subsequently, extensive data analysis was performed to reveal the implications of SI3 domain manipulation on pausing dynamics.

The molecular reporter designed to monitor SI3 movement can also be used in studies focused on characterizing RNAP properties. Specifically, this should allow exploration of the effects of ppGpp on elemental pausing, thereby expanding our understanding of the regulatory mechanisms governing transcriptional control.

1.5 References

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

2. M. D. Wang *et al.*, Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902-907 (1998).

3. R. Landick, The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**, 1062-1066 (2006).

4. R. Landick, Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol* **75**, 291-314 (2021).

5. R. Landick, J. Carey, C. Yanofsky, Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc Natl Acad Sci U S A* **82**, 4663-4667 (1985).

6. C. L. Chan, R. Landick, The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J Biol Chem* **264**, 20796-20804 (1989).

7. A. Chauvier, J. F. Nadon, J. P. Grondin, A. M. Lamontagne, D. A. Lafontaine, Role of a hairpin-stabilized pause in the Escherichia coli thiC riboswitch function. *RNA Biol* **16**, 1066-1073 (2019).

8. A. Chauvier *et al.*, Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat Commun* **8**, 13892 (2017).

9. G. A. Perdrizet, 2nd, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci U S A* **109**, 3323-3328 (2012).

10. A. K. Huffines *et al.*, Rate of transcription elongation and sequence-specific pausing by RNA polymerase I directly influence rRNA processing. *J Biol Chem* **298**, 102730 (2022).

11. M. Abdelkareem *et al.*, Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell* **75**, 298-309 e294 (2019).

12. D. Wang *et al.*, Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203-1206 (2009).

13. B. K. Bharati *et al.*, Crucial role and mechanism of transcription-coupled DNA repair in bacteria. *Nature* **604**, 152-159 (2022).

14. J. Y. Kang *et al.*, Structural basis for transcription complex disruption by the Mfd translocase. *Elife* **10** (2021).

15. L. You *et al.*, Structural basis for intrinsic transcription termination. *Nature* **613**, 783-789 (2023).

16. B. A. Boudreau *et al.*, StpA and Hha stimulate pausing by RNA polymerase by promoting DNA-DNA bridging of H-NS filaments. *Nucleic Acids Res* **46**, 5525-5546 (2018).

17. B. A. Boudreau, M. V. Kotlajich, R. Landick, In Vitro Transcription Assay to Quantify Effects of H-NS Filaments on RNA Chain Elongation by RNA Polymerase. *Methods Mol Biol* **1837**, 351-386 (2018).

18. C. M. Hustmyer, M. B. Wolfe, R. A. Welch, R. Landick, RfaH Counter-Silences Inhibition of Transcript Elongation by H-NS-StpA Nucleoprotein Filaments in Pathogenic Escherichia coli. *mBio* **13**, e0266222 (2022).

19. J. W. Weaver *et al.*, Control of transcription elongation and DNA repair by alarmone ppGpp. *Nat Struct Mol Biol* **30**, 600-607 (2023).

20. F. W. Martinez-Rucobo, P. Cramer, Structural basis of transcription elongation. *Biochim Biophys Acta* **1829**, 9-19 (2013).

21. D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, I. Artsimovitch, Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157-162 (2007).

22. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**, 1876-1882 (2001).

23. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941-954 (2006).

24. J. Zhang, M. Palangat, R. Landick, Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* **17**, 99-104 (2010).

25. G. A. Belogurov *et al.*, Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* **26**, 117-129 (2007).

26. J. Y. Kang *et al.*, Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e1614 (2018).

27. R. A. Mooney, K. Schweimer, P. Rosch, M. Gottesman, R. Landick, Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol* **391**, 341-358 (2009).

28. M. Delbeau *et al.*, Structural and functional basis of the universal transcription factor NusG pro-pausing activity in Mycobacterium tuberculosis. *Mol Cell* **83**, 1474-1488 e1478 (2023).

29. A. V. Yakhnin, P. Babitzke, Mechanism of NusG-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the Bacillus subtilis trp leader. *Mol Microbiol* **76**, 690-705 (2010).

30. C. M. Burns, W. L. Nowatzke, J. P. Richardson, Activation of Rho-dependent transcription termination by NusG. Dependence on terminator location and acceleration of RNA release. *J Biol Chem* **274**, 5245-5251 (1999).

31. J. Chalissery *et al.*, Interaction surface of the transcription terminator Rho required to form a complex with the C-terminal domain of the antiterminator NusG. *J Mol Biol* **405**, 49-64 (2011).

32. V. Molodtsov, C. Wang, E. Firlar, J. T. Kaelber, R. H. Ebright, Structural basis of Rho-dependent transcription termination. *Nature* **614**, 367-374 (2023).

33. I. Artsimovitch, R. Landick, The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193-203 (2002).

34. K. E. Kolb, P. P. Hein, R. Landick, Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J Biol Chem* **289**, 1151-1163 (2014).

35. X. Guo *et al.*, Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e814 (2018).

36. K. S. Ha, I. Toulokhonov, D. G. Vassylyev, R. Landick, The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J Mol Biol* **401**, 708-725 (2010).

37. Y. H. Huang *et al.*, Structure-Based Mechanisms of a Molecular RNA Polymerase/Chaperone Machine Required for Ribosome Biosynthesis. *Mol Cell* **79**, 1024-1036 e1025 (2020).

38. N. Said *et al.*, Structural basis for lambdaN-dependent processive transcription antitermination. *Nat Microbiol* **2**, 17062 (2017).

39. F. J. O'Reilly *et al.*, In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554-557 (2020).

40. C. Wang *et al.*, Structural basis of transcription-translation coupling. *Science* **369**, 1359-1365 (2020).

41. M. W. Webster *et al.*, Structural basis of transcription-translation coupling and collision in bacteria. *Science* **369**, 1355-1359 (2020).

42. J. Y. Kang *et al.*, Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife* **6** (2017).

43. P. P. Hein, R. Landick, The bridge helix coordinates movements of modules in RNA polymerase. *BMC Biol* **8**, 141 (2010).

44. R. O. Weinzierl, The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain. *BMC Biol* **8**, 134 (2010).

45. T. A. Windgassen *et al.*, Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res* **42**, 12707-12721 (2014).

46. L. M. Iyer, E. V. Koonin, L. Aravind, Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* **3**, 1 (2003).

47. W. J. Lane, S. A. Darst, Molecular evolution of multisubunit RNA polymerases: structural analysis. *J Mol Biol* **395**, 686-704 (2010).

48. L. Shen *et al.*, An SI3-sigma arch stabilizes cyanobacteria transcription initiation complex. *Proc Natl Acad Sci U S A* **120**, e2219290120 (2023).

49. N. Zakharova, I. Bass, E. Arsenieva, V. Nikiforov, K. Severinov, Mutations in and monoclonal antibody binding to evolutionary hypervariable region of Escherichia coli RNA polymerase beta' subunit inhibit transcript cleavage and transcript elongation. *J Biol Chem* **273**, 24912-24920 (1998).

50. I. Artsimovitch, V. Svetlov, K. S. Murakami, R. Landick, Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* **278**, 12344-12355 (2003).

51. E. F. Ruff *et al.*, E. coli RNA Polymerase Determinants of Open Complex Lifetime and Structure. *J Mol Biol* **427**, 2435-2450 (2015).

52. J. Y. Kang *et al.*, RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* **69**, 802-815 e801 (2018).

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

54. N. M. Maizels, The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of Escherichia coli. *Proc Natl Acad Sci U S A* **70**, 3585-3589 (1973).

55. A. Mayer *et al.*, Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541-554 (2015).

56. M. H. Larson *et al.*, A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).

57. I. O. Vvedenskaya *et al.*, Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285-1289 (2014).

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

59. I. Artsimovitch, R. Landick, Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**, 7090-7095 (2000).

60. P. P. Hein *et al.*, RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* **21**, 794-802 (2014).

61. J. Zhou, K. S. Ha, A. La Porta, R. Landick, S. M. Block, Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. *Mol Cell* **44**, 635-646 (2011).

62. A. Sevostyanova, G. A. Belogurov, R. A. Mooney, R. Landick, I. Artsimovitch, The beta subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol Cell* **43**, 253-262 (2011).

63. Z. F. Mandell *et al.*, NusG is an intrinsic transcription termination factor that stimulates motility and coordinates gene expression with NusA. *Elife* **10** (2021).

64. A. V. Yakhnin *et al.*, NusG controls transcription pausing and RNA polymerase translocation throughout the Bacillus subtilis genome. *Proc Natl Acad Sci U S A* **117**, 21628-21636 (2020).

65. A. V. Yakhnin, K. S. Murakami, P. Babitzke, NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J Biol Chem* **291**, 5299-5308 (2016).

66. C. M. Burns, L. V. Richardson, J. P. Richardson, Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol* **278**, 307-316 (1998).

67. K. M. Herbert *et al.*, Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* **125**, 1083-1094 (2006).

68. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proc Natl Acad Sci U S A* **118** (2021).

69. J. Y. Kang *et al.*, An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. *Proc Natl Acad Sci U S A* **120**, e2215945120 (2023).

70. I. Artsimovitch, R. Landick, Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev* **12**, 3110-3122 (1998).

71. I. Toulokhonov, I. Artsimovitch, R. Landick, Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**, 730-733 (2001).

72. S. M. Vos, L. Farnung, H. Urlaub, P. Cramer, Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601-606 (2018).

73. B. Krummel, M. J. Chamberlin, Structural analysis of ternary complexes of Escherichia coli RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J Mol Biol* **225**, 239-250 (1992).

74. E. Nudler, A. Goldfarb, M. Kashlev, Discontinuous mechanism of transcription elongation. *Science* **265**, 793-796 (1994).

75. L. S. Churchman, J. S. Weissman, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373 (2011).

76. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *Elife* **8** (2019).

77. S. Borukhov, V. Sagitov, A. Goldfarb, Transcript cleavage factors from E. coli. *Cell* **72**, 459-466 (1993).

78. P. Sivaramakrishnan, A. J. E. Gordon, J. A. Halliday, C. Herman, How Acts of Infidelity Promote DNA Break Repair: Collision and Collusion Between DNA Repair and Transcription. *Bioessays* **40**, e1800045 (2018).

Chapter 2: Obligate movements of SI3 controls transcriptional elongation and pausing

Adapted from:

Bao, Y., & Landick, R. (2021). Obligate movements of an active site–linked surface domain control RNA polymerase elongation and pausing via a phe pocket anchor. Proceedings of the National Academy of Sciences, 118(36).

2.1 Abstract

The catalytic trigger loop (TL) in RNA polymerase (RNAP) alternates between unstructured and helical hairpin conformations to admit and then contact the NTP substrate during transcription. In many bacterial lineages, the TL is interrupted by insertions of 2–5 surface-exposed, sandwich-barrel hybrid motifs (SBHMs) of poorly understood function. The 188-aa, 2-SBHM *E. coli* insertion, called SI3, occupies different locations in halted, NTP-bound, and paused transcription complexes, but its dynamics during active transcription and pausing are undefined. Here we report design, optimization, and use of a Cys-triplet reporter to measure the positional bias of SI3 in different transcription complexes and to determine the effect of restricting SI3 movement on nucleotide addition and pausing. We describe use of H_2O_2 as a superior oxidant for RNAP disulfide reporters. NTP binding biases SI3 toward the closed conformation whereas transcriptional pausing biases SI3 toward a swiveled position that inhibits TL folding. We find that SI3 must change location in every round of nucleotide addition and that restricting its movements inhibits both transcript elongation and pausing. These dynamics are modulated by a crucial Phe pocket formed by the junction of the two SBHM domains. This SI3 Phe pocket captures a Phe residue in the RNAP jaw when the TL unfolds, explaining the similar phenotypes of alterations in the jaw and SI3. Our findings establish that SI3 functions by modulating the TL folding to aid transcriptional regulation and to reset secondary channel trafficking in every round of nucleotide addition.

2.2 Introduction

Gene expression in all cellular life forms is accomplished by a conserved, multi-subunit RNA polymerase (RNAP) via a highly regulated nucleotide addition cycle (NAC; **Fig. 2.1**) that extends RNA transcripts by reaction with DNA-templated nucleoside triphosphates (NTPs). A posttranslocated RNAP first samples NTPs until the correct NTP binds in its active site. A flexible trigger loop (TL) then folds into a helical hairpin called the trigger helices (TH) that is stabilized by contacts to the complementary NTP–DNA pair, causing active-site closure (1-4). The closed active site accelerates by $\geq 10^4$ S_N2 displacement of pyrophosphate from the NTP by RNA 3′-hydroxyl-group attack on the α-phosphate, resulting in extension of the RNA by one nucleotide (2, 5, 6). Pyrophosphate release, RNA–DNA translocation, and TH unfolding then reset the NAC to the posttranslocated state for the next round of NTP binding. In many bacterial lineages, TL–TH cycling occurs in every round of nucleotide addition despite the presence of a large insertion of 2–5 surface-exposed, sandwich-barrel hybrid motifs (SBHMs) (7-9). The 188-aa, 2-SBHM insertion in *E. coli* RNAP is called sequence insertion 3 (SI3) (10).

Nucleotide addition is rapid (30–100 $nt \cdot s^{-1}$) at most DNA positions. However, the elongation complex (EC) is controlled in part by pause sequences that halt the EC for ≥1 s by converting it to a paused EC (PEC) (11-13). These transcriptional pauses aid regulation and RNA biogenesis in both prokaryotes and eukaryotes by enabling transcriptional attenuation, transcription–translation coupling, RNA folding, RNA splicing, termination, and antitermination (14-20). Different types of pauses involve distinct PEC conformations, some of which can interconvert (Figure 1A). The elemental pause occurs at specific DNA sequences that inhibit DNA template base loading (i.e., DNA translocation) during the NAC (11). A modest conformational change called ‘swiveling’ and base-pairing energetics may explain slow translocation in the elemental PEC (ePEC), which can occupy multiple states and may rearrange further into backtracked or hairpin-stabilized PECs (21-24). Backtracked PECs are prevalent in eukaryotes, also occur in prokaryotes, and can be triggered by nucleotide misincorporation (25-28). At a hairpin-stabilized pause (e.g., the well-studied *his* pause from the attenuation control region of the *E. coli* histidine biosynthetic operon) (16, 29), a nascent RNA hairpin formed 11–12 nt from the RNA 3′ end remodels the PEC and stabilizes the swiveled conformation (22, 23). Swiveling involves an ~4° rotation of a module composed of the clamp, dock, shelf, β′ C-terminal region, jaw, and SI3 (the swivel module). The swivel-module rotation distorts an SI3-complementary depression in the RNAP surface so that RNAP can no longer accommodate the folded TH position of SI3. Backtrack pauses also induce swiveling similar to that seen in *his*PEC (24).

SI3 connects to the apex of the TL–TH via two flexible linkers (7), contacts the β′ jaw in the TL conformation (5), and mediates hairpin-stabilized pausing in *Eco*RNAP by inhibiting TL folding in the swiveled PEC conformation. SI3 and jaw deletions abrogate hairpin-stabilized pausing with no additional effect when combined (5, 30, 31). SI3 shifts to a different location when the TH form, disrupting the SI3–jaw interface (23, 24, 32, 33).

Several aspects of SI3 function remain unclear. Although the TL folds and unfolds in every round of the NAC as expected (34), it is unclear whether SI3 also fluctuates between open and closed locations during rapid transcription or only opens when the EC is halted or paused. Second, it is unknown if SI3 swiveling is required for pausing or just a consequence of other RNAP conformational changes responsible for pausing. Finally, the structural basis by which the jaw domain modulates SI3 movements is undefined.

To address these questions, we sought ways to exploit engineered disulfide crosslinks both to measure and to test SI3 conformational changes. Previously used methods to form RNAP disulfides proved unable to generate a key SI3 disulfide efficiently, leading us to develop a new approach based on H_2O_2 oxidation. The facile formation of stable disulfides with limited damage to RNAP using H_2O_2 followed by catalase allowed us to determine SI3 positional biases in different EC states and to analyze how SI3 locations affect transcription and pausing, including in mutant RNAPs that define a role for SI3–jaw interaction.

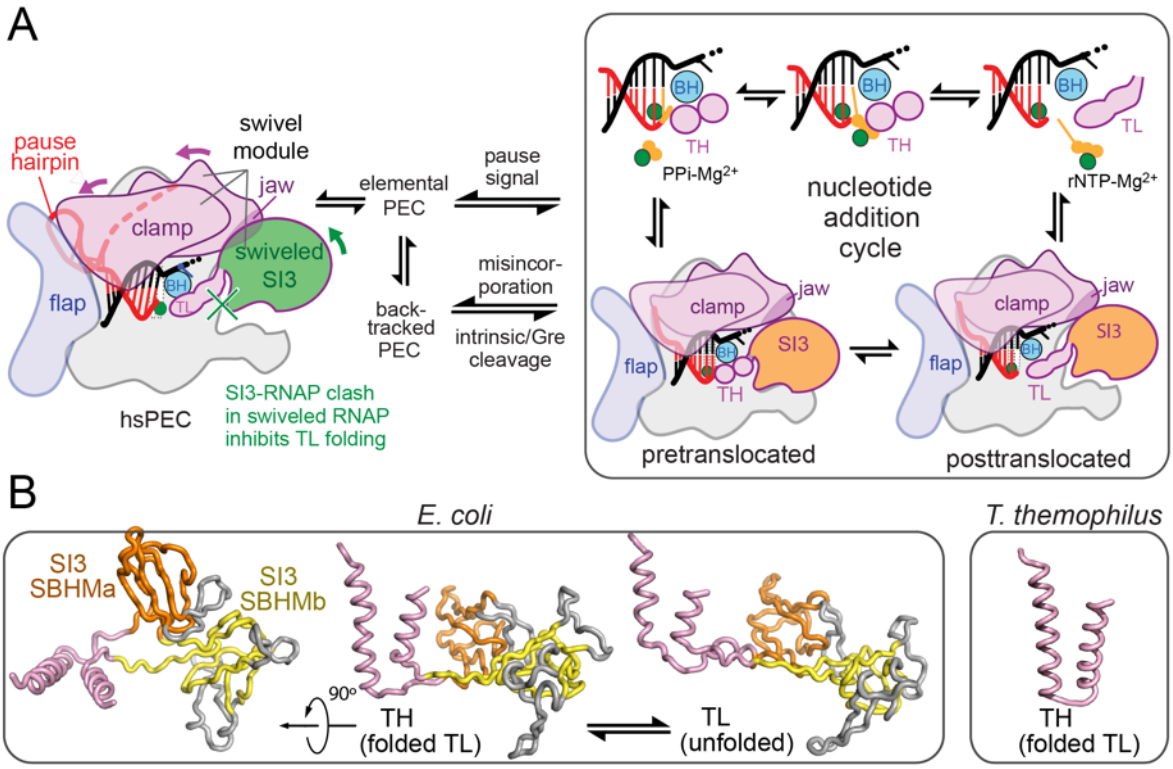


Figure 2.1. *E. coli* RNAP transcription cycle and SI3.

(*A*) The *E. coli* EC alternates between pre- and post-translocated states and between TL and TH states during the nucleotide addition cycle (right) but enters alternate conformations in response to pause signals in the DNA and RNA that can slow translocation and induce swiveling or backtracking. SI3 occupies different locations in these states and inhibits TH formation in the swiveled state due to steric clash. (*B*) SI3 in *E. coli* TL (pink) is composed of two SBHM domains (orange and yellow), which include several loops (white) in addition the core SBHM fold. Structures are from pdb 6rh3 (TH) and 6rin (TL) (24). The *T. thermophilus* TH (right panel; pdb 2o5j) (2), which lacks a TL sequence insertion, is shown for comparison.

2.3 Results

2.3.1 Design and optimization of an SI3 positional reporter – H_2O_2 is a superior thiol oxidant

To design disulfide reporters of SI3 location, we examined SI3 residues that changed potential contacts in RNAP structures in which different SI3 positions have been resolved (closed SI3, TH form; open SI3, TL form; swiveled SI3, hairpin-stabilized pause form; **Fig. S2.1*A***) (23, 32, 33). We first sought an SI3 residue that would make different contacts in the closed, open, and swiveled SI3 positions so that potential disulfides created by substitutions could report which location predominates in different EC states. Unfortunately, no SI3 residue exhibits this property. However, the closed and swiveled SI3 states could be distinguished by the location of residues in a β-ribbon-resembling loop in SI3-SBHM1 (β′1046–1061). This loop inserts between the rim helices (RH) and sequence insertion 1 (SI1) such that β′D1051 approaches β′G671 in the closed state, whereas β′D1051 approaches a different part of SI1 (βC267) in the swiveled state. We predicted that Cys substituted for β′D1051 would compete for disulfide formation with Cys residues substituted for β′G671 and βC267 in the closed and swiveled states, respectively (**Fig. 2.2*A*, *B***).

To test this prediction, we generated an RNAP bearing β′D1051C, β′G671C, and βR267C substitutions. We call this system in which disulfides that report two different conformations compete with each other for formation a Cys-triplet reporter (CTR) (23, 35). Conveniently, the disulfides formed by the SI3 CTR are readily distinguishable because one is internal to the β′ subunit, altering its electrophoretic mobility, and the other links β′ to β, causing a much larger retardation of electrophoretic mobility (**Fig. 2.2*C*,*D*; Fig. S2.2**).

To date, cystamine (AED), diamide (TMAD), or Cu^{2+} have been used as oxidants to form disulfide bonds in Cys-substituted RNAPs (5, 23, 35-38). However, these oxidants did not work well for the SI3 CTR (**Fig. 2.2*C*; Fig. S2.2*A*,*B***) and each has disadvantages for probing RNAP conformations. Disulfide formation requires activation of one Cys sulfur by conjugation to the oxidant followed by S_N2 nucleophilic displacement of the activating group by the second Cys thiolate via a trigonal bipyramidal transition state (**Fig. S2.1*C***). This mechanism can reduce overall crosslink efficiency in two ways (37, 39). First, although high concentrations of oxidant favor disulfide formation electrochemically, some oxidants (e.g., cystamine) also react with the activated intermediate causing accumulation of mixed disulfides in competition with formation of the desired disulfide. Second, steric clash of the conjugated activating group with the surrounding protein environment may inhibit formation of the trigonal bipyramidal

geometry in the transition state required for disulfide formation. This latter issue is a particular problem in the complex structure of RNAP where larger or less flexible activating groups often reduce crosslink efficiency. Finally, oxidants can also be disadvantageous because they attack other exposed Cys residues on RNAP and alter its properties (e.g., heavy metals like Cu^{2+} can bind and inactivate enzymes) (40). Such effects may explain why diamide causes smearing of subunit bands during PAGE, complicating quantitation of disulfide bond formation (**Fig. 2.2*C***).

In search of better oxidants for disulfide probes of RNAP conformation, we tested hydrogen peroxide (H_2O_2). Compared to cystamine or diamide, H_2O_2 generated SI3 disulfides with higher efficiency (>80%) and at relatively high rate (k_{obs} = ~1.5 min^{-1}; **Fig. S2.2**). H_2O_2-induced crosslinks also form in two steps (41). A cysteine thiolate reacts with H_2O_2 to form a sulfenic acid (-CSOH) followed by displacement of OH^- by a second cysteine thiolate. However, disulfides can also form by conjugation of two sulfenic acids via a thiosulfinate (41), which coupled with minimal steric constraints on the transition state may explain why H_2O_2 produces disulfide crosslinks with higher efficiency.

The higher efficiency of H_2O_2-induced disulfide formation allowed us to create an SI3 CTR in which β′C671 in the RH and βC267 in SI1 compete to form disulfides with β′C1051 in SI3, thus reporting the occupancy ratio of the closed vs. swiveled SI3 conformations (**Fig. S2.1*E***). In contrast, cystamine was unable to generate the β′C1051-βC267 at useful levels irrespective of RNAP conformation (**Fig. 2.2*C***).

H_2O_2 offers an additional advantage over other oxidants. It can be rapidly eliminated from a solution by addition of catalase, which converts H_2O_2 to H_2O and O_2 with a maximal turnover number of 16,000,000–44,000,000 s^{-1}(42). Thus, catalase treatment minimizes oxidative damage to the RNAP once disulfides are formed, allowing *in vitro* transcription or other assays of crosslinked RNAP in the absence of undesired effects of the added oxidant. Catalase treatment also generated clearer bands in SDS-PAGE for crosslink quantitation (compare to diamide, **Fig. 2.2*C***, or H_2O_2 without catalase, **Fig. S2.1*F***). Finally, H_2O_2 treatment had minimal effects on RNAP activity (Fig. S3). We conclude that H_2O_2 is a superior oxidant for disulfide probing of RNAP conformations.

2.3.2 SI3 samples multiple locations in a resting EC

Using the optimized CTR system, we found that both crosslink species (SI3–RH and SI3–SI1) form in an EC, a paused complex (*his*PEC), and core RNAP (**Figs. 2.2*D*, 2.2*E***). This result indicates that SI3 samples multiple locations by thermal fluctuation including the closed and swiveled states even in the absence of pause signals or bound NTP substrate. Differences in the ratio of SI3–RH to SI3–SI1 crosslinks should be proportional to changes in the relative occupancies of these states in different transcription complexes (*i.e.*, proportional to differences in ΔG° of the states in different transcription complexes). This view of RNAP as a dynamic system sampling diverse conformational states with different probabilities is consistent with both MD simulations (43) and with the typically lower resolution observed in cryo-EM analyses of RNAP complexes compared to those achievable in globular enzymes (23, 24, 32, 35).

Before using the SI3 CTR to study RNAP, we considered two questions. First, does the formation of both crosslink species by the CTR RNAP reflect non-interconverting conformations of SI3 in separate populations on RNAP? Using RNAPs containing only two Cys substitutions so that only one or the other disulfide could form (i.e., either the SI3–RH or SI3–SI1 disulfide in RNAPs called Cys-pair reporters, CPRs) (37), we found that SI3 conformations are in dynamic equilibrium. Fluctuations in SI3 position allowed each CPR disulfide to form at high efficiency absent conformation with the alternative disulfide (>80%; **Fig. S2.2*E***), which would not be possible if the conformations did not interconvert. Thus, SI3 readily samples the closed or swiveled conformation even in the absence of an NTP substrate or a pause signal.

Second, how is the ratio of SI3–SI1/SI3–RH disulfides in the CTR RNAP related the ratio of the corresponding SI3 conformations? Although we cannot determine this relationship without an independent measure of the SI3 conformations, it seems unlikely that the SI3–RH/SI3–SI1 disulfide ratio and the closed/swiveled conformation ratio are identical. The rate at which each crosslink forms in the CTR RNAP will be a function not only of the ratio of the SI3 conformations but also of the chemical environment around the Cys residues and the probability that the side chains can achieve the transition state sterically. These environmental effects are unlikely to be identical for the competing disulfides. Nonetheless, because the SI3 states are dynamically interconverting, a change in the SI3–RH/SI3–SI1 disulfide ratio should be proportional to changes in the bias of the SI3 positional equilibrium for a given RNAP or EC state. We defined ‘SI3 Positional Bias’ (SPB) as the ratio of the SI3–RH/SI3–SI1 disulfides and

an relative measure of closed/swiveled SI3 occupancy (**Fig. 2.2*E***). A higher SPB indicates that SI3 is more biased toward the closed position and a lower SPB indicates SI3 is more biased toward the swiveled position.

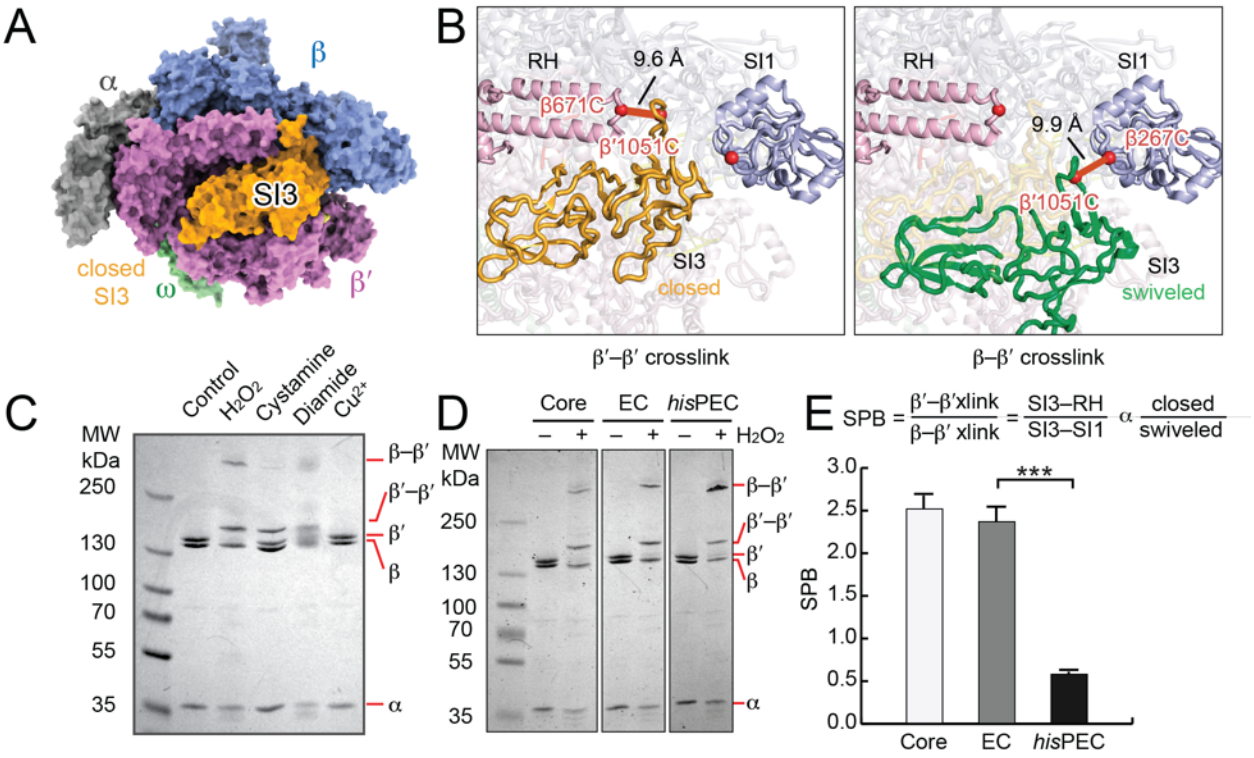


Figure 2.2. Alternative SI3 locations and detection by a Cys-triplet reporter (CTR).

(*A*) RNAP in TH conformation with SI3 in closed position (pdb 4yln) (33). (*B*) Cα–Cα distances of residues changed to Cys to generate the CTR (red spheres): β′SI3–β′RH (closed, orange, pdb 4yln) and β′SI3–βSI1 (swiveled, green, pdb 6asx). (32, 33). RH, rim helices; SI1, sequence insertion 1. (*C*) Comparison of ability of different oxidants to form disulfides in the CTR (Materials and Methods). (*D*) Comparison of CTR disulfides formed in apo RNAP core enzyme, an EC, and a paused EC (*his* hsPEC). Both β′–β′and β′–β disulfides are formed in all three complexes. (*E*) SI3 positional bias (SPB) calculated from the ratio of the 2 crosslink species in apo RNAP, the EC and the *his* hsPEC. Results are mean ± SD (n = 3). Unpaired t test, ***p < 0.001.

2.3.3 Cognate NTP binding stabilizes SI3 shift to closed conformation

We first sought to explore how SI3 position changes upon substrate binding. Structural studies suggest that SI3 favors the closed position when cognate NTP is bound by an EC (24, 33), conditions known to stabilize TH formation to promote catalysis (23, 32, 33, 44). However, the positional bias of SI3 upon binding of cognate or noncognate NTPs in physiologic conditions where SI3 is dynamic remains untested. To ask how NTP binding affects SI3 location, we used SI3-CTR RNAP and an RNA–DNA scaffold to form an EC that allowed NTP binding but not nucleotide addition. This scaffold contained complementary 50mer DNA strands and a 19-nt RNA that was reacted with ATP to form A20 ECs or with 3′dATP to form dA20 ECs (**Fig. 2.3*A*; Fig. S2.4*A***). Upon oxidation, A20 EC gave an SPB of 1.8 ± 0.1, similar to U19 (1.2 ± 0.1), but the SPB shifted to 5.9 ± 0.5 upon addition of the unreactive CTP analog CMPCPP (**Fig. 2.3*B***). This shift in SBP indicates that SI3 increases occupancy of the closed position by a factor of ~3.3 upon cognate NTP binding. Similarly, when CTP was added to dA20 EC, which lacked the 3′ OH required for S_N2 attack during nucleotide incorporation, SPB shifted from 2.4 ± 0.2 to 6.7 ± 0.9, an increase of ~2.8x. The slightly higher SPB of dA20 EC vs. A20 EC is consistent with prior observations that a 3′ deoxy substitution favors the pretranslocated register, which favors TL folding (45).

We next asked how cognate vs. noncognate NTP would affect SI3 positional bias. At 100 μM, noncognate ATP and GTP marginally increased dA20 SBP whereas UTP decreased SPB possibly because rU-dG base-pairing disfavors TL folding (**Fig. 2.3*B***). Similarly, at higher concentration (5 mM) noncognate GTP decreased SPB to 1.8 ± 0.3 with a concentration dependence similar to that by which cognate CTP raised SPB from 4.5 ± 0.5 at 10 μM to 7.9 ± 0.6 at 5 mM (**Fig. 2.3*C***). These concentration profiles with half-maximal but opposite effects at ~180 μM are similar to previous measurements of NTP binding to 3′deoxy ECs (37, 45) and are consistent with a pre-insertion model for NTP selection in which NTP binding is indiscriminate and substrate discrimination arises principally from the ability of a complementary rNTP-dNMP base pairing to support TL folding (3, 46). The decrease in SPB caused by GTP suggests that either non-Watson–Crick base pairing or NTP binding in the E site may decrease occupancy of the TH state due of steric clash with TL side chains. Taken together, these results suggest that cognate NTPs shift SI3 positional bias toward the closed position, likely in concert with TL folding, whereas noncognate NTPs may shift SI3 away from the closed positions, likely linked to decreased TH formation. Thus, the energetics of SI3 positioning may aid in NTP discrimination by modulating TL folding, which directly affects NTP discrimination (47-49), although effects of SI3 on NTP selectivity remain to be investigated in detail.

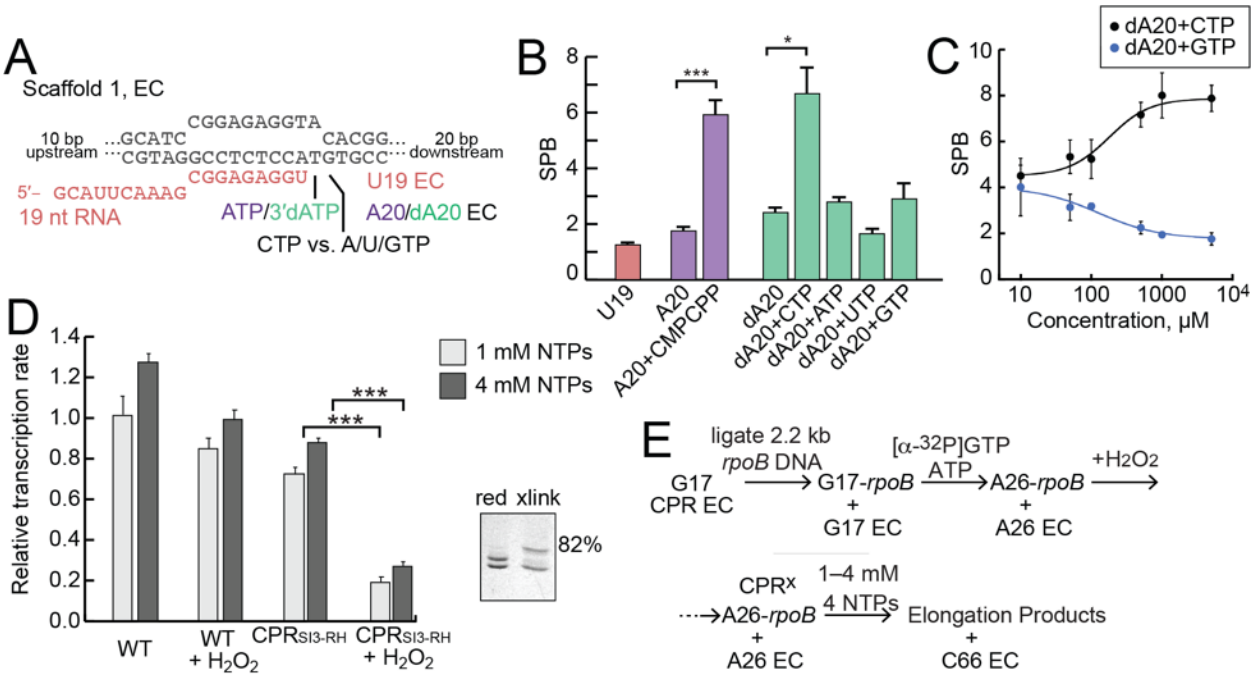


Figure 2.3. SI3 is biased toward the closed position by cognate but not noncognate NTPs and inhibits transcript elongation when crosslinked in the closed position.

(*A*) RNA–DNA scaffold used to reconstitute ECs. Either ATP or 3′ deoxy ATP (3′dATP) can be incorporated at position 20 (see also **Fig. S2.4**). 3′dA20 EC supports NTP binding but not nucleotide addition. (*B*) SPB of A20 EC, 3′dA20 EC, and ECs incubated with different NTPs (at 100 μM each NTP). (*C*) SPB of 3′dA20 EC incubated with different concentrations of cognate CTP or non-cognate GTP. (*D*) Average elongation rates (relative to WT RNAP in reducing condition) at 1 mM each NTP or 4 mM each NTP. Data represent mean ± SD (n = 3). Unpaired t test, ***$p < 0.001$. Inset, a typical assay of crosslink efficiency (82% of $CPR_{SI3\text{-}RH}$ was crosslinked by 5 mM H_2O_2). (*E*) Schematic of ligated-scaffold transcription assay. The ligation reaction is incomplete, but only products longer than C66 are used to estimate elongation rate (see **Fig. S2.5** and ***Materials and Methods***).

2.3.4 Nucleotide addition cycling requires SI3 open–close cycling

A key question is whether SI3 moves back and forth between the closed and open position in every round of the NAC. Although the TL clearly undergoes folding unfolding cycles in each round of nucleotide addition (34) and blocking TH unfolding by disulfide crosslinking inhibits rapid transcription (37), it is unknown if SI3 also moves in every round of the NAC or might only shift to the open or swiveled state when an EC halts or pauses. To determine whether SI3 movement is required for rapid nucleotide addition, we asked how biasing SI3 to the closed position with a disulfide would affect transcription rate during multi-round nucleotide addition. For this purpose, we used a previously described ligated-scaffold transcription assay that enables transcription by reconstituted ECs over multi-kb of DNA (50). We ligated a 2.2 kb DNA from the *E. coli rpoB* gene to reconstituted $CPR_{SI3–RH}$ ECs and then measured transcript elongation at 1 mM each of all four NTPs with and without the SI3–RH disulfide (**Fig. 2.3 *D*** and ***E***). Biasing SI3 to the closed position with the disulfide allowed transcript elongation but at an average rate that was slower by a factor of ≥3.7 compared to the non-crosslinked EC (**Fig. 2.3*D*; Fig. S2.5**). Pausing by the SI3-crosslinked RNAP was enhanced at some locations (e.g., at G31, 5 nt downstream of A26) and only a few ECs were able to extend past the point of ligation (+66; **Fig. S2.5*B***). The inhibition of multi-round transcript elongation when SI3 is crosslinked in the closed position is unlikely due only to reduced affinity of ECs for substrate NTP, since increasing Mg^{2+}·NTP concentrations 4-fold to 4 mM each had little effect on transcription rate (**Fig. 2.3D**).

These results indicate that SI3 must ordinarily alternate between the closed and open positions in each round of nucleotide addition rather than remaining closed while accommodating TL–TH cycling (note that biasing SI3 open is even more inhibitory as expected based on a TH-containing EC structure; see below and **Fig. 2.5**). Biasing SI3 to the closed position likely influences transcription in multiple ways. First, it should favor TH formation, which narrows the RNAP secondary channel and should restrict NTP access to active site. Second, the closed SI3–TH may stabilize the pretranslocated state during the NAC, which could slow transcription by inhibiting translocation as observed for a TH-stabilizing crosslink (37). Third, the closed SI3–TH appeared to strengthen sequence-specific pausing at certain sites (e.g., the pause sites evident in **Fig. S2.5*B***). The SI3–RH crosslink must allow some extent of TH–TL interconversion, however, since it decreased transcription rate by less than a factor of ten whereas deletion of the TL inhibits transcription rate by $\geq 10^4$ and a direct stabilization of the TH inhibits transcription rate by factors of ≥10–200 (2, 5, 6, 44). We conclude that alternation of SI3 between open and closed locations likely occurs in

every round of nucleotide addition, and that inhibition of these oscillations explains the effect of the SI3–RH disulfide.

2.3.5 SI3 location bias shifts in paused ECs

We next examined the effect of RNAP pausing on the location bias of SI3 using the well-characterized paused ECs (PECs) that form at U102 in the attenuation-controlled leader region of the *E. coli his* biosynthetic leader region (*his*PECs) (22, 23, 29). RNAP forms either ePECs in the absence of a nascent RNA pause hairpin or hairpin-stabilized PECs (hsPECs) when a nascent RNA structure forms in the RNA exit channel 11 nt from the RNA 3′ end. We examined ECs (G18), ePECs (U20), and hsPECs (C19 and U20) reconstituted with CTR RNAP on the same scaffold using an antisense RNA to form the pause hairpin, limited nucleotide addition to advance G18 ECs to C19 (with CTP) and U20 (with CTP and UTP), and either a complementary or noncomplementary (bubble) NT DNA strand (**Fig. 2.4*A***). We observed higher SPBs on the bubble-containing ECs and PECs, but the same overall trends in SI3 positional bias (**Figs. 2.4 *B*,*C***). The higher SPBs on bubble-containing complexes may reflect lesser thermal fluctuations of SI3 when translocation is constrained by the noncomplementary bubble, allowing greater formation of the closed SI3 crosslink. In both types of complexes, however, the ePEC exhibited a modest shift toward the swiveled conformation relative to G18 and C19 ECs whereas the hsPEC increased the bias toward the swiveled conformation more dramatically (**Fig. 2.4 *B*,*C***).

These changes in SI3 positional bias for the *his* elemental and hairpin-stabilized pauses confirm that SI3 swiveling imaged by cryoEM of PECs reflects changes in the dynamics of SI3 in physiological condition and that the extent of swiveling correlates with pause duration since SI3 in the *his* ePEC is less biased toward the swiveled state than in longer-lived *his* hsPEC. Hairpin promotion of swiveling is evident at C19 in addition to U20, consistent with prior observations of that the *his* pause hairpin can form in the pause –1 complex (51, 52); this result also supports the conclusion exit-channel RNA duplexes directly promote SI3 swiveling. A shift in SI3 toward the closed conformation in C19 ECs relative to G18 ECs also was evident (**Fig. 2.4 *B*,*C***), consistent with prior observations that 3′ C favors the pretranslocated state, 3′ G favors the posttranslocated state, and a pretranslocated EC stabilizes the TH vs. TL conformation (45, 53).

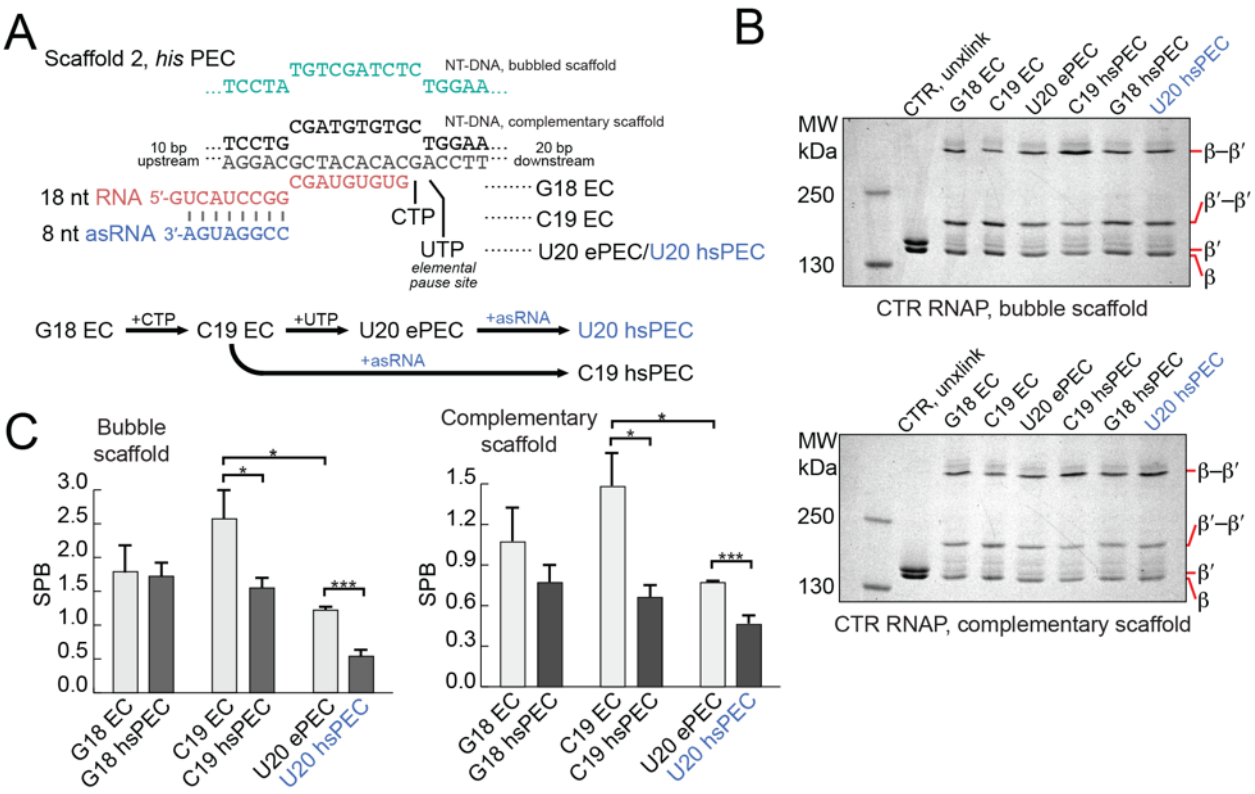


Figure 2.4. SI3 becomes biased to the swiveled position as RNAP in the *his* ePEC and hsPEC.

(*A*) RNA–DNA scaffold used to reconstitute *his* PECs. G18 EC reacts with CTP and UTP to form C19 EC and U20 ePEC, respectively (See also **Fig. S2.3**). Addition of an asRNA mimics formation of the *his* pause hairpin and forms the *his* hsPEC. (*B*) SDS-PAGE images of CTR crosslinks formed on complementary and bubble scaffolds of his ECs and PECs. (*C*) SPB of *his* ECs and PECs. Data represent mean ± SD (n = 3). Unpaired t test, ***p < 0.001, *p < 0.05.

2.3.6 SI3 swiveling contributes to the hairpin-stabilized pause

Although both cryoEM and our CTR data support a causal role for SI3 swiveling in pausing, a direct experimental test by preventing or promoting swiveling is needed to establish causality. To perform this test, we next used $CPR_{SI3–SI1}$ and $CPR_{SI3–RH}$ RNAPs to probe the effects on *his* pausing of strongly biasing SI3 to the swiveled or closed conformation, respectively. The SI3–SI1, swiveled crosslink was formed at G18 with high efficiency (~82%) prior to C19 radio-labeling and U20 extension (**Fig. 2.5*A***). Addition of GTP to the crosslinked PECs enabled measurement of the pause escape rate. Without crosslinking, the majority of the ePEC (94 ± 2%) rapidly escaped the weak elemental pause ($k_{ePEC,\,red} = 0.75\ s^{-1}$) whereas the majority of the hsPEC (72 ± 2%) generated after asRNA addition escaped the hairpin-stabilized pause a rate 34 times slower ($k_{hsPEC,\,red} = 0.022\ s^{-1}$; **Fig. 2.5 *B*,*C***). When $CPR_{SI3–SI1}$ RNAP was crosslinked to favor the swiveled SI3 conformation, the elemental pause escape rate was significantly slowed by a factor of >170 (k_{ePECx}=0.0046 s^{-1}), even slower than that of the uncrosslinked hsPEC. Addition of asRNA to the SI3–SI1-crosslinked U20 EC further decreased the pause escape rate by a factor of only ~1.5 compared to the 34-fold effect the uncrosslinked EC ($k_{hsPECx} = 0.0031\ s^{-1}$). These results indicate that artificially biasing SI3 toward the swiveled position greatly increases pause dwell time even in the absence of an exit channel RNA duplex. Nearly all the effect of the pause hairpin was recapitulated by the swivel-biasing crosslink, suggesting that practically all the effect of hairpin is mediated by stabilizing the swiveled SI3 conformation. We conclude that biasing SI3 to the swiveled position strongly inhibits TL folding, thus obviating the swivel-inducing effect the exit-channel RNA duplex created by asRNA addition.

To test the opposite prediction that inhibiting SI3 swiveling with the SI3–RH crosslink should also obviate the effect of an exit-channel RNA duplex, we measured pause kinetics for the *his* PEC formed with $CPR_{SI3–RH}$ RNAP with and without addition of the asRNA. As predicted for the hsPEC, formation of the SI3–RH crosslink to favor the SI3 closed position greatly reduced pausing and essentially eliminated the effect of the exit channel duplex (compare pausing profiles for the crosslinked ePEC and hsPEC, **Fig. 2.5 *D*,*E***). Biasing SI3 toward the closed conformation either prevented the exit-channel duplex from inducing swiveling or prevented the duplex from forming (e.g., if swiveling is required to accommodate the duplex in the exit channel). These results support the view that induction of SI3 swiveling and consequent inhibition of TL folding is the mechanism by which exit-channel RNA duplexes increase pause dwell time.

In contrast for the *his* ePEC, formation of the SI3–RH crosslink to favor the SI3 closed position modestly increased pausing, contrary to the expectation that swiveling aids elemental pausing (**Fig. 2.5*D*,*E***). However, this increase in pausing is consistent with stimulation of pausing at some sites observed when crosslinked $CPR_{SI3–RH}$ RNAP transcribed over a long DNA template (**Fig. S2.5*B***). We conjecture that stabilizing SI3 in the closed conformation may stabilized the pretranslocated register of the ePEC and could increase pausing at sites where the translocation is naturally slow.

To summarize, the pausing behavior of PECs with SI3 biased toward either the closed or swiveled position strongly supports the view that swiveling stabilized by an exit-channel duplex or simply in response to RNA and DNA sequences is a direct cause of transcriptional pausing, likely by inhibiting the transition from the half-translocated (RNA-only) to fully translocated state in the NAC. Stabilization of the TH by biasing SI3 toward the closed state also appears capable of stimulating pausing.

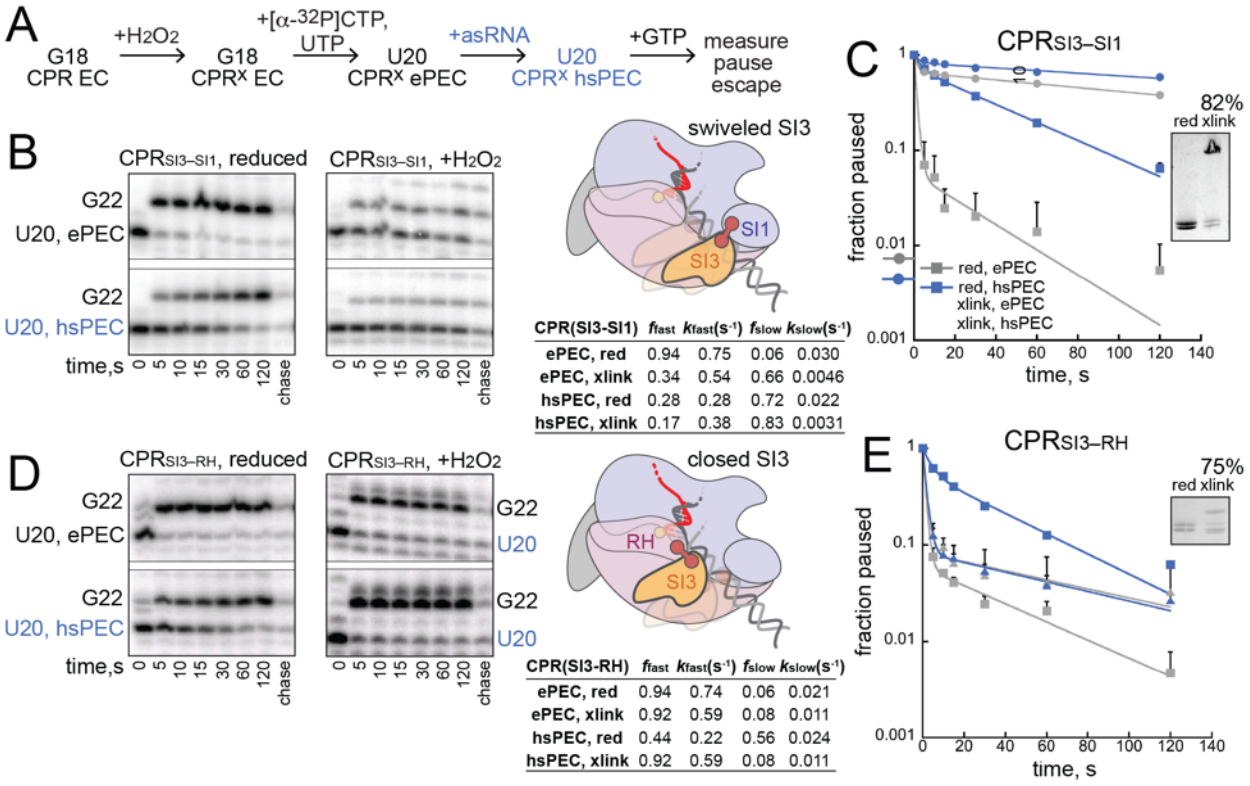
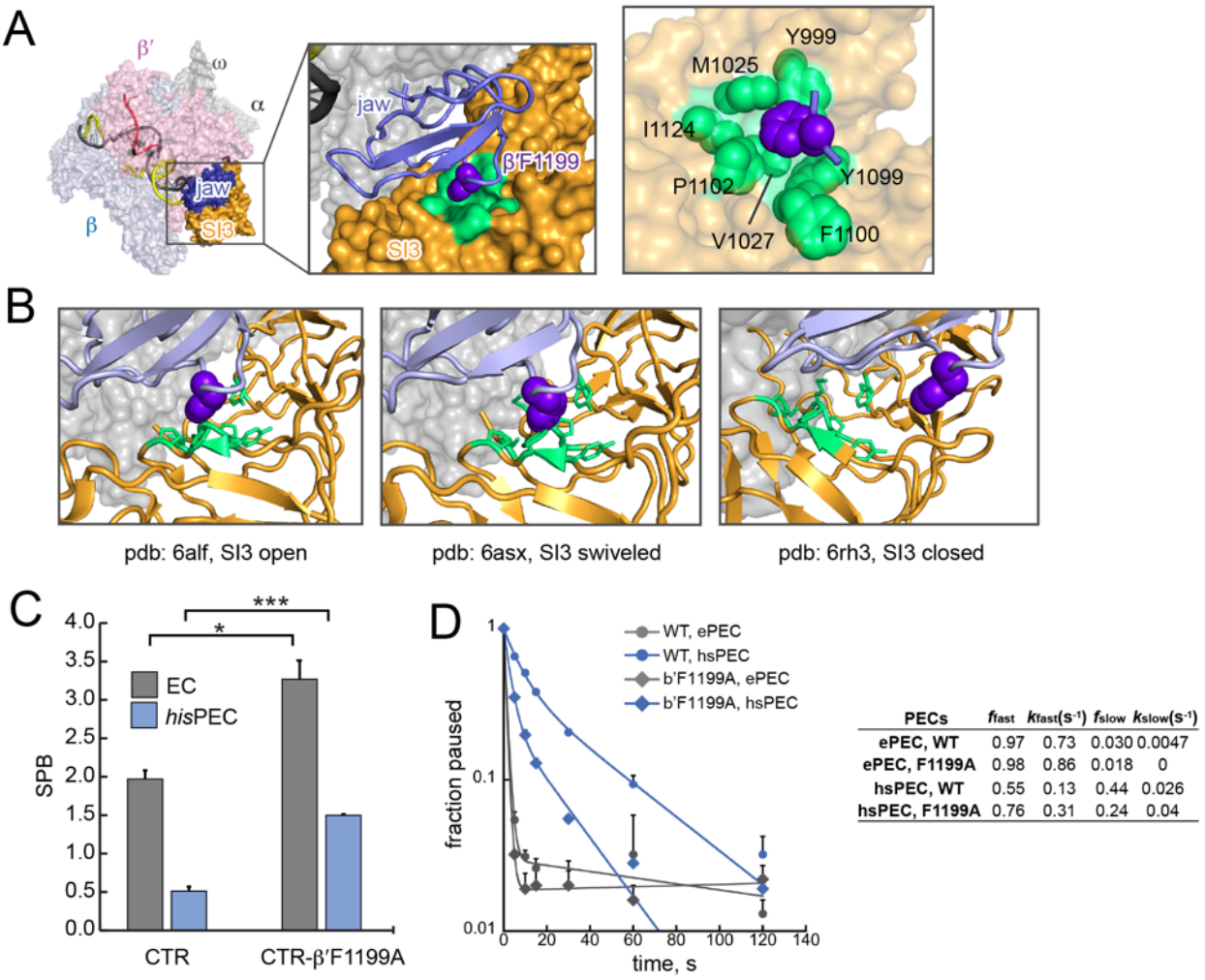


Figure 2.5. Biasing SI3 to swiveled or closed positions with disulfides establishes that SI3 swiveling is sufficient and necessary for hairpin-stabilized pausing.

(*A*) Reaction schematic to test effects of SI3 disulfides on pausing. (*B*),(*D*) RNAs formed during pause assays, schematics of SI3 disulfides tested, and pause kinetics. f_{fast} and f_{slow} are the fractions of pauses in faster and slower states when fit to a two-exponential equation and k_{fast} and k_{slow} are the corresponding rate constants (***Materials and Methods***). A complete set of kinetic fitting parameters with errors is in **Table S2.1**. (*C*),(*E*) Fraction pause RNAs as a function of time under different conditions.

2.3.7 An SI3 Phe pocket captures the RNAP jaw to tune TL folding, pausing, and catalysis

Previous biochemical studies suggest that the β′ jaw domain participates in pausing via interdependent effects with SI3 (5, 30, 31), but the basis of jaw–SI3 interaction has not been defined. We examined the SI3–jaw interface in cryoEM structures of ECs and the *his*PEC (23, 24, 32) and observed a hydrophobic alcove on the SI3 surface (a Phe pocket) that appeared to accommodate the Phe side chain of a jaw residue (β′F1199; **Fig. 2.6*A***). However, when SI3 is in the closed position, this SI3–jaw interaction is disrupted and F1199 no longer resides in the pocket (e.g., when NTP is bound by an EC; **Fig. 2.6*B***) (24). To test whether the jaw F1199–SI3 Phe pocket interaction inhibits movement of SI3 to the closed position, we generated RNAPs bearing a β′F1199A substitution and used them for CTR and pausing assays. We found that SI3 became more biased toward the closed position in both ECs and hsPECs when F1199 was changed to Ala, indicating that the F1199–Phe pocket interaction stabilizes jaw–SI3 interaction and that its disruption decreases the thermodynamic stabilities of SI3 locations in which the interaction occurs (i.e., open and swiveled SI3; **Fig. 2.6*C***). Pause assays using β′F1199A RNAP revealed a reduction in hairpin-stabilized pausing, consistent with the model that SI3 swiveling, stabilized by F1199–Phe pocket interaction, is required for hairpin-stabilized pause (**Fig. 2.6*D***). These results suggest that the jaw provides a docking site for SI3 via a Phe pocket interaction that modulates SI3 movements during both the NAC and transcriptional pausing.



PECs	f_{fast}	$k_{fast}(s^{-1})$	f_{slow}	$k_{slow}(s^{-1})$
ePEC, WT	0.97	0.73	0.030	0.0047
ePEC, F1199A	0.98	0.86	0.018	0
hsPEC, WT	0.55	0.13	0.44	0.026
hsPEC, F1199A	0.76	0.31	0.24	0.04

Figure 2.6. Capture of β′F1199 in the RNAP jaw by a Phe pocket in SI3 modulates SI3 movement.

(*A*) Phe pocket interaction in SI3 open position (pdb: 6alf) (32). Light green, the hydrophobic pocket in SI3. Purple, β′F1199. (*B*) Phe pocket interaction and disruption in open (pdb: 6alf), swiveled (6asx), and closed (6rh3) ECs or PEC (23, 24, 32). (*C*) CTR assay of wild-type and β′F1199A RNAPs bound to EC and *his*PEC scaffolds. Unpaired t test, ***$p < 0.001$, *$p < 0.05$. (*D*) Pause RNA as a function of time after mixing with GTP in reducing conditions and derived pausing kinetic parameters for CTR RNAP and CTR-β′F1199A RNAP. β′F1199A decreased hairpin-stabilized pausing by a factor of ~2.5. See **Fig. 2.5** legend for kinetic parameter definitions.

2.4 Discussion

We report new insights into the function of SI3, an unusual modulator of RNAP catalytic function, by exploiting sulfhydryls placed at *E. coli* RNAP locations that alternate proximity when SI3 changes location and H_2O_2–catalase as a superior disulfide-bond-formation system to actuate disulfide reporters of RNAP conformation: (*i*) SI3 fluctuates among a range of locations from closed to swiveled in most ECs with a predominate location determined by interactions in the EC; (*ii*) SI3 ordinarily cycles between preferentially closed and open positions in every round of nucleotide addition accompanying TL folding–unfolding; (*iii*) SI3 swiveling is a direct cause of prolonged transcriptional pausing; and (*iv*) the open and swiveled conformations of SI3 are stabilized by capture of a Phe residue in the RNAP jaw domain by a Phe-pocket formed at the interface between the two SBHM domains in SI3. These findings have important implications for the function and evolution of SI3-like domains in bacterial RNAPs.

2.4.1 SI3 positional cycling resets RNAP for timely regulation during transcript elongation

Structural analyses capture SI3 in the open position or as disordered except when active-site bound NTP stabilizes SI3 in a closed position (24, 32, 33, 54) or pausing favors a swiveled position (22-24), but it has been unclear if SI3 actually cycles between closed and open locations in concert with TL folding and unfolding during rapid nucleotide addition. The location of the open SI3 on the jaw domain stabilized by the Phe-pocket interaction appears structurally incompatible with TH formation because the linkers connecting SI3 to the TL are too short to accommodate the TH position without SI3 movement. The opposite is not necessarily true, however, raising the possibility that SI3 could remain closed during rapid TL–TH cycling and nucleotide addition and possibly open only at pause sites or in response to other regulatory events. Our findings contradict this view, since stabilizing the closed SI3 strongly inhibits rapid nucleotide addition (**Fig. 2.3*D***). Instead, they support a view in which SI3 fluctuates among many possible locations including closed, open, and swiveled in most ECs, with strong biases favoring the closed state when substrate binding stabilizes TH formation, the open state after PPi release and translocation, and the swiveled state in paused ECs (**Fig. 2.4*D*; Fig. 2.7**).

This picture of a fluctuating SI3 domain that cycles between predominately closed or open positions in concert with the nucleotide addition cycle suggests that SI3 could play an important regulatory function by resetting the RNAP secondary channel and possibly the overall the state of the EC in every round of nucleotide addition.

Resetting the secondary channel may be important because multiple secondary channel-binding regulators compete to occupy the secondary channel when SI3 is in the open position, including GreA, GreB, TraR, Rnk, and DksA in *E. coli* and additional regulators in other bacteria (55). When SI3 shifts to the closed position, steric clash would preclude secondary-channel binding by these regulators. NTPs also must enter the active site through the secondary channel but appear unable to pass through when the channel is occupied by a secondary channel-binding factor. By cycling between open and closed locations in every round of nucleotide addition, SI3 effectively resets the secondary channel to a naïve state when it moves from the closed position that excludes the regulators to the open position that allows both NTPs and regulators to compete based only new state of the EC after the last round of nucleotide addition. This resetting function of SI3 could aid efficient regulatory responses of the EC based on its current state.

SI3 cycling also could contribute to resetting the entire EC in every round of nucleotide addition. A long-standing discussion has considered whether RNAP can occupy persistent conformational states over multiple rounds of nucleotide addition; such states might, for instance, be resistant to termination or possess altered catalytic properties independent of bound regulators (56, 57). Although not definitive, cycling of SI3 in every round of nucleotide addition appears inconsistent with the existence of persistent RNAP conformational states since SI3 resets to a state able to transiently occupy multiple locations in every round of nucleotide addition.

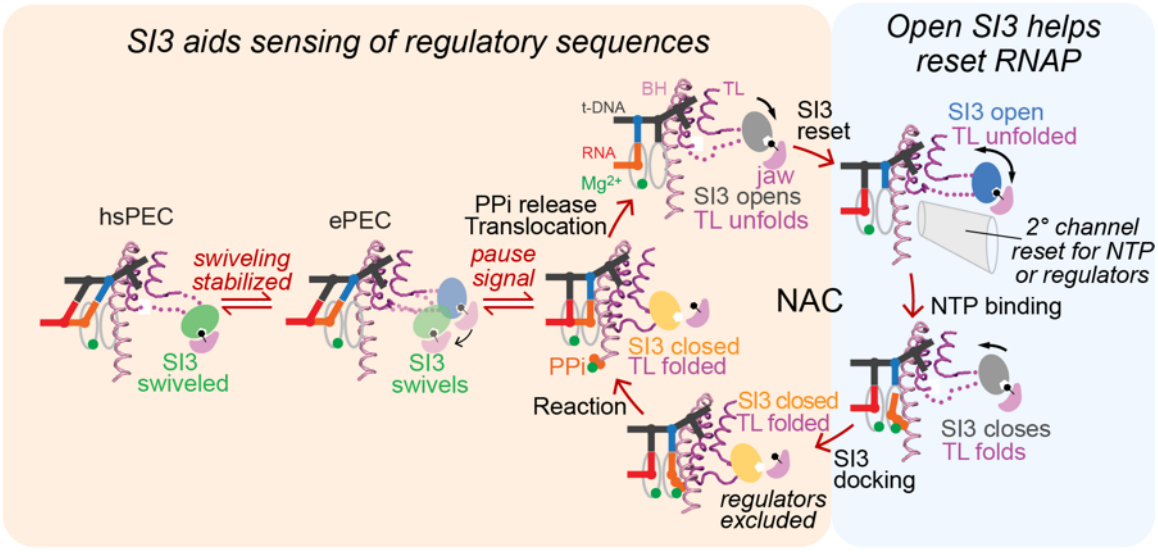


Figure 2.7. SI3 states during RNA chain elongation and pausing aid transcription.

In an active posttranslocated EC with an unfolded TL (NAC, top left), SI3 docks on the jaw domain but thermal fluctuation enables SI3 to sample among different positions. In this state, the secondary channel (2° channel) is open and can admit NTPs or regulators. Binding of cognate NTP then stabilizes the folded TH with SI3 shifted in the closed position. In this location, SI3 will clear the secondary channel of regulators and, together with the TH, close the active site, which promotes the nucleotide addition reaction. After nucleotide addition, the EC either undergoes translocation with PPi release, TL unfolding, and SI3 movement back to the open position or undergoes a transition to the elemental paused state. In response to RNA–DNA sequences, the ePEC can either isomerize back to the elongation pathway (back to the NAC) or convert to either a backtracked PEC or a hairpin-stabilized PEC in which swiveling becomes stabilizes for the pause hairpin.

2.4.2 SI3 augments the regulatory capacity of RNAP by enhancing transcriptional pausing

Our results establish that biasing SI3 location toward swiveling with a disulfide bond directly enhances transcriptional pausing, presumably by inhibiting TL folding, stabilizing the half-translocated state observed in the *his* ePEC and hairpin-stabilized PEC, or both (**Figs. 2.1 and 2.5**). Intriguingly, we also observed that some pauses may be increased when SI3 is biased toward the closed conformation by a disulfide bond (**Fig. S2.5*B***). Possibly, this increased pausing by the closed CPR RNAP occurs at sites where PECs can be stabilized in the pretranslocated state. Strikingly, however, the closed CPR RNAP was unaffected by pause hairpin formation at the *his* pause site, both confirming the role of swiveling in pause hairpin action and showing that SI3 affects different classes of pause mechanisms in different ways.

These results suggest that an important function of SI3 could be to increase the potential for sequence-dependent regulation of transcript elongation by sensitizing RNAP to different modes of pausing. The presence in RNAP of a TL insertion able to assume different conformational states increases the conformational complexity of RNAP. By way of its insertion directly into the TL, the conformation of SI3 can affect catalysis. By virtue of being surface exposed, it is possible, although as yet undemonstrated, that solution conditions could affect SI3 conformational states. Thus, by fluctuating among various positions, SI3 may increase the ways that DNA and RNA sequence can bias RNAP to meta-stable pause sites stabilized in part by SI3 contacts or clashes in particular RNAP conformations. In this way, SI3 increases the regulatory capacity of RNAP to respond to transcriptional pausing signals.

2.4.3 SI3-like insertions are widespread but the Phe-pocket is specific to γ-proteobacteria

Both resetting secondary-channel trafficking and increasing regulatory capacity may explain in part how SI3, and TL insertions more generally, were acquired during evolution of bacterial RNAPs. However, several interesting questions about SI3 evolution remain unanswered. All known TL insertions consist of two or more repeats of SBHM domains (8, 9), but their distribution among bacterial lineages begs the question of whether some insertions in different lineages arose independently or if all TL insertions are evolutionarily related (**Fig. S2.6*A***). TL insertions are present in most bacterial lineages but the lineages in which they are absent vs. present do not cluster together (58), raising the possibility these insertions arose independently. Most notably, insertions (of five SBHM domains) are present in cyanobacteria even though they are absent in most evolutionarily nearby lineages. The SBHM fold is

present elsewhere in the RNAP large subunits, including as a universally conserved feature called the flap or wall in bacterial and eukaryotic RNAPs, respectively (8). It is possible that unusual genetic rearrangements that arise over time in large, rapidly growing bacterial populations may have led to duplications of the SBHM fold and independent selection of these TL insertions in RNAP in different lineages.

Regardless of evolutionary history, another question is whether TL insertions are related to another feature of transcriptional regulation that differs in diverse bacteria. Although our understanding of the diversity of regulatory mechanisms in diverse bacteria remains primitive and in great need of expanded study, two distinct themes in bacterial regulation of transcript elongation have recently emerged. First, the universal elongation factor NusG displays markedly different effects in different lineages, stimulating pausing in some like firmicutes and actinobacteria (based *B. subtilis* and *M. tuberculosis*) whereas it suppresses pausing in proteobacteria (based on *E. coli*) (59-61). Second, transcription and translation are coupled in some bacterial lineages like *E. coli* but not coupled in others like *B. subtilis*, with related large differences in the role of Rho-dependent termination in mRNA quality control (62). However, the distribution of TL insertions among bacterial lineages appears to be unrelated to the patterns of either NusG pause effect or transcription–translation coupling among different bacterial lineages (59, 62). Thus, selection for TL insertions appears to have operated independently of these other major themes in the bacterial regulation of transcript elongation.

One possibility is that a major selective pressure for generation or retention of TL insertions is simply to increase the evolvability of RNAP to adapt to diverse environments and stresses (63). In other words, the presence of a TL insertion increases the ways in which the regulatory responses of RNAP can be altered by mutation in response to different selective pressures. Even without mutation, a TL insertion may RNAP easier to regulate by increasing the available conformations that impact catalysis. Especially in bacteria whose range of environmental niches and adaptive responses is huge, an ability to change regulatory responses simply by changing the properties and interactions of a TL insertion may be significant.

One line of evidence supporting selection for evolvability of TL insertions comes from our discovery of the role of the Phe pocket in modulating the function of SI3 in *E. coli* (**Fig. 2.6**). Although the sequences that comprise the Phe pocket and the jaw Phe residue itself are well conserved among γ-proteobacteria, these residues are not conserved in other bacterial lineages in which TL insertions are present (**Fig. S2.6*B***). Indeed, the jaw Phe residue is

more commonly His in other TL insertion-bearing lineages. This observation suggests the crucial Phe pocket evolved to modulate SI3 function in γ-proteobacteria, whereas other interactions of TL insertions, as yet uncharacterized, may predominate in other lineages. Thus, our current findings highlight two important questions for future study: (*i*) what specific regulatory features are conferred on SI3 by the Phe-pocket interaction; and (*ii*) to what extent do locations and interactions similar to those found for SI3 operate in other TL insertions versus lineage-specific evolution of TL-insertion locations and interactions?

2.5 Materials and methods

Reagents

DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT; Coralville, IA) and purified by 15% denaturing polyacrylamide gel electrophoresis (PAGE) before use. [γ-^{32}P]ATP, [α-^{32}P]CTP and [α-^{32}P]GTP were obtained from PerkinElmer Life Sciences; rNTPs, from Promega (Madison, WI, USA); and 3′deoxy ATP (3′dATP) and CMPCPP, from Jena Bioscience. H_2O_2 30% (w/w) solution and standard buffers and reagents were from Sigma-Aldrich or Thermo-Fisher.

RNAP purification

RNAP expression plasmids were transformed into BL21(DE3) and a single colony was inoculated into 5 mL LB+50 µg kanamycin/mL and grown overnight growth at 37 °C. The saturated cell culture was then added to 1 L fresh LB+50 µg kanamycin/mL and grown at 37 °C with adequate aeration by orbital shaking in a Fernbach flask until apparent OD_{600} reached 0.6. Protein expression was induced by adding IPTG (Gold Biotechnology) to 1 mM and cell growth was continued for 3 hours. The cells were harvested and homogenized by sonication in 30 mL Lysis Buffer (50 mM Tris-HCl pH 7.9, 5% v/v glycerol, 233 mM NaCl, 2 mM EDTA, 10 mM β-mercaptoethanol, 10 mM DTT, 100 µg/mL PMSF, and 1 tablet of protease inhibitor cocktail (Roche cOmplete, Mini, EDTA-free). After removing cell debris by centrifugation (11000 × *g* , 15 min, 4 ºC). DNA binding proteins including target RNAPs were precipitated by addition of polyethylenimine (PEI, Sigma-Aldrich) to 0.6% (w/v) final. After centrifugation (11000 × *g* , 15 min, 4 ºC), the protein pellet was resuspended in 25 mL PEI Wash Buffer (10 mM Tris-HCl pH 7.9, 5% v/v glycerol, 0.1 mM EDTA, 5 µM $ZnCl_2$, 500 mM NaCl) to remove non-target proteins. After centrifugation (11000 × *g* , 15 min, 4 ºC), RNAP was eluted from the pellet into 25 ml PEI Elution Buffer (10 mM Tris-HCl pH 7.9, 5% v/v glycerol, 0.1 mM EDTA, 5 µM $ZnCl_2$, 1 M NaCl). The crude extract of RNAPs was subjected to sequential FPLC purifications of Ni^{2+} column (HisTrap FF 5 ml, GE Healthcare Life Sciences) and heparin column (Heparin FF 5 ml, GE Healthcare Life Sciences) and the purified RNAP was dialyzed into RNAP Storage Buffer (10 mM Tris-HCl, 25% v/v glycerol, 100 mM NaCl, 100 µM EDTA, 1 mM $MgCl_2$, 20 µM $ZnCl_2$, 10 mM DTT) to final concentrations of 5–10 mg/mL. The high concentration of DTT (10 mM) was included to inhibit disulfide bond formation by engineered cysteine residues.

CTR crosslinking assay, gel quantitation, and SPB calculation

The DNA–RNA bubble scaffold used for EC reconstitution was assembled by incubation purified oligos (20 μM non-template strand DNA, 15 μM template strand DNA, 10 μM RNA, 20 mM Tris-OAc pH 7.7, 5 mM $Mg(OAc)_2$, 40 mM KOAc) in a thermal cycler (95 °C for 2 min, 75 °C for 2 min; 45-to-25 °C gradient over 25 min). To reconstitute CTR ECs, the CTR core RNAP was incubated with the scaffold at 1:2 ratio (2.5 μM CTR-RNAP, 5 μM scaffold) in 10 μL of Elongation Buffer (25 mM HEPES-KOH pH 8.0, 130 mM KCl, 5 mM $MgCl_2$, 0.15 mM EDTA, 5% v/v glycerol, 25 μg acetylated BSA/ml) for 20 min at 37 °C. For reconstitution of EC with a complementary scaffold, only the RNA and template strand DNA were annealed (15 μM template strand DNA, 10 μM RNA) using the same solution and thermal conditions as for the bubble scaffold. The CTR core enzyme was incubated with the scaffold for 10 min at 37 °C, then non-template strand DNA was added to 20 μM and incubation was continued for 10 min. To mimic an RNA hairpin structure at the RNA exit channel, asRNA base pairing to nascent RNA was added to 25 μM and incubation was continued for 5 min.

For assays that involved nucleotide addition to reconstituted ECs prior to disulfide formation (Figs. 3, 4, and 5), NTPs were added to 50 μM each and the reactions were incubation at 37 °C for 5 min. Nucleotide incorporation was tested for completion by measuring the change in size of 5′ end ^{32}P-labeled RNA by denaturing PAGE (Fig. S3).

For PAGE analysis, ~0.3 uL samples were mixed with NuPAGE 4X LDS Sample Loading Buffer (Novex, reducing agents omitted) and electrophoresed through a thin layer 4-15% gradient polyacrylamide gel (GE Healthcare) using a PhastSystem Electrophoresis unit (originally from Pharmacia). Gels were stained with Imperial Protein Stain (Thermo Scientific), destained with water and imaged with a CCD camera (Protein Simple).

For comparison of crosslink efficiencies among oxidants, cystamine or diamide were used at 10 mM or 5 mM, respectively, and were quenched with 12.5 mM iodoacetamide. $CuSO_4$ was used at 5 μM and was not quenched prior to PAGE.

We used the Fiji-a software package based on ImageJ to quantify the percentage of crosslinking. The intensity of two crosslinked bands ($I_{\beta\text{-}\beta'}$ for SI3–SI1crosslink, $I_{\beta'\text{-}\beta'}$ SI3–RH crosslink) and the uncrosslinked β and β′ bands ($I_{unxlink}$) were measured. Because the SI3–SI1 crosslink was inter-subunit and the SI3–RH crosslink was intra-

subunit, the two crosslink bands were easily distinguished (since MW_{β}=151 kDa, $MW_{\beta'}$=155 kDa, we assumed Coomassie brilliant blue G-250 binds β and β′ to similar levels). Thus, we calculated the % crosslinking from the band intensities (I) in the gel images as:

$SI3–SI1\% = I_{\beta–\beta'} / (I_{\beta–\beta'} + I_{\beta'–\beta'} + I_{\beta} + I_{\beta'})$; and

$SI3–RH\% = 2\times I_{\beta'–\beta'} / (I_{\beta–\beta'} + I_{\beta'–\beta'} + I_{\beta} + I_{\beta'})$,

where the crosslinked species are indicated by subscripts containing two subunits and the intensity of the SI3–RH crosslinked species is doubled because it is approximately half the size of the SI3–SI1 crosslinked species. The SPB value was then calculated as the ratio between the two crosslink species:

$SPB = (SI3–RH\%) / (SI3–SI1\%)$.

All SPB values were determined from at least 3 independent assays and reported as mean ± SD.

Ligation-elongation assay and quantification

Pre-assembled scaffold (2 µL of 5 µM RNA, 10 µM template strand DNA) was incubated with 4 µL of CPR core enzyme (12.5 µM) in 20 µL Elongation Buffer (final RNA:RNAP ratio of 1:5; 0.5 µM RNA, 1.0 µM template strand DNA, 2.5 µM CPR RNAP) for 15 min at 37 °C. Non-template strand DNA (1.5 µL to 1.5 µM final concentration) was added and the 10 µL mixture was incubated for an additional 15 min at 37 °C. The estimated concentration of assembled EC was 500 nM.

To ligate the EC to the 2.2 kb DNA template, 270 nM EC was incubated with 90 nM purified DNA template, 1 × NEB T4 ligase buffer (including 1 mM ATP and 10 mM DTT) and 400 U T4 DNA ligase in a 10 µL reaction at 16 °C for 2 hours. The estimated concentration of ligated EC was 90 nM. To extend EC to the desired register and form the crosslinked EC, 50 nM ligated EC was radiolabeled by incubation with 2 µCi [α-^{32}P] GTP (3000 Ci/mmol) for 2 min, extended with 100 µM each ATP and GTP for 10 min to form A26 EC, and then crosslinked with 5 mM H_2O_2 in a 20 µL reaction as described above. The mixture was then diluted with Elongation Buffer to 100 µL. The estimated final concentration of EC was 10 nM. To re-start elongation, 25 µL of 2 mM each NTP was combined with with 25 µL ~10 nM EC at 37 °C. Reaction samples (5 µL) were withdrawn at 10, 20, 30, 40, 60, 120 and 300 seconds and mixed with an equal volume of Stop Buffer (8 M urea, 50 mM EDTA, 90 mM Tris-borate

buffer, pH 8.3, 0.02% bromophenol blue, 0.02% xylene cyanol). For rate measurements, triplicates of 20-s reactions were performed. RNAs in all collected samples were heated to 95 °C for 2 min before size separation by denaturing PAGE (8%, 19:1 acrylamide:bis-acrylamide, 45 mM Tris-borate, pH 8.3, 1.25 mM Na_2EDTA, 8 M urea). The RNAs were run alongside a radiolabeled, MspI-digested pBR322 DNA ladder. The gel was exposed to a Storage Phosphor Screen (GE Healthcare) and scanned with Typhoon PhosphorImager.

The gels were quantified using ImageQuant software (GE Healthcare). To estimate the averaged elongation rates, the intensities of RNAs longer than 66 nt after 20 s extension with 1 mM NTPs (a 66 nt band was generated by unligated ECs) were determined and average elongation rates were calculated from the average sizes of these RNAs in 3 independent assays.

In vitro transcription pause assays and quantification

For a typical pause assay, the annealed scaffold prepared as described above was incubated with CPR core enzyme with 1: 3 ratio (1 μM RNA, 2 μM template strand DNA, 3 μM CPR RNAP) in 60 μL Elongation Buffer for 15 min at 37 °C. For complementary scaffolds, non-template DNA was added to 5 μM and incubation continued for an additional 15 min. Further binding by CPR RNAP was blocked by addition of heparin to 0.1 mg/mL and 5 min incubation. To induce disulfide bond formation, 5 mM H_2O_2 was added and the mixture was incubated for 15 min at 37 °C. Samples with and without H_2O_2 addition were taken for PAGE analysis to measure crosslink efficiency in each experiment. Excess H_2O_2 was eliminated by addition of the mixture catalase to 0.1 U/mL followed by 5 min incubation at room temperature. The mixture was then diluted with 4 volumes of Elongation Buffer and radiolabeled by incubation with 2 μCi [α-^{32}P]CTP (3000 Ci/mmol). To form paused ECs, CTP and UTP (2 μM each final) were added sequentially followed by 5 min incubation at 37 °C. The hairpin-stabilized paused complex was formed by addition of 1 μM asRNA at this point to reactions containing 200 nM ePEC and incubation was continued for 10 min. The PECs were then incubated at 37 °C for 2 min before initiating the pause assay by combination of equal volumes of 200 nM PEC solution and 20 μM GTP in Elongation buffer (100 nM ePEC/hsPEC and 10 μM GTP final). Samples (5 μL) were removed at 5, 10, 15, 30, 60 and 120 s and mixed with an equal volume of Stop Buffer. Before addition of GTP, a zero time point sample was generated by combination of 5 uL 200 nM PEC, 5 μL Elongation Buffer and 10 μL Stop Buffer. A final portion (4.45 μL) of PECs remaining at end of the time course was chased by addition of 0.55 μl 15 mM GTP and incubation for 1 min at 37 °C followed by addition of 5 μL Stop

Buffer to assess the activity of the PECs. RNAs were analyzed by denaturing PAGE (15% 19:1 acrylamide: bis-acrylamide, 45 mM Tris-borate, pH 8.3, 1.25 mM Na_2EDTA, 8 M urea). Gels were exposed to a Storage Phosphor Screen (GE Healthcare) and scanned with Typhoon PhosphorImager. Deviations from this protocol, where appropriate, are noted in the figure legends.

Pause assays were quantified using ImageQuant software (GE Healthcare). For all time point samples, the paused and the escaped RNA intensities were measured separately and the shorter transcripts corresponding to degraded RNA were not included in analysis. To correct for any readthrough transcripts formed before pause reaction started, these RNA intensities at time point zero were subtracted from each time point. In the similar way, the paused RNA intensity of the chased sample was subtracted from each time point to correct the inactive ECs/PECs. The corrected pause RNA as a function of time was used to determine pausing kinetics fitting to bi-exponential decay functions (23, 64).

Protein sequence alignment

Seventy-two bacterial species (**Table S2.2**) were chosen from 38 representative phyla and the sequences of their b′ subunits were obtained from the National Center for Biotechnology Information protein database (NCBI, National Library of Medicine, Bethesda, MD; https://www.ncbi.nlm.nih.gov/). The b′ subunit sequences were checked for the existence of trigger loop sequence insertion. Thirty sequences containing trigger loop insertions were aligned with MUSCLE algorithm (65) and the alignment were transformed into sequence logos with WebLogo (66).

Table of oligonucleotides and plasmids

Oligonucleotides (5′ to 3′)	**Assay**	**Stock #**
EC bubble scaffold, non-template strand DNA GGTCAGTACGTCCTGTCGATCTTCGGAAGAGAATTCAGATCTTCCAGTGG	CTR Crosslink	#8847
EC bubble scaffold, template strand DNA CCACTGGAAGATCTGAATTCTCTTCCCCTCTAGCTCAGGACGTACTGACC	CTR Crosslink	#8848
EC bubble scaffold, RNA	CTR Crosslink	#8855

UUUUUUUGAGCUAGAGG		
EC complementary scaffold, non-template strand DNA GGTCAGTAGGGCATCCGGAGAGGTACACGGCGAATACCCGATCTTCCAGT	CTR Crosslink	#13942
EC complementary scaffold, template strand DNA ACTGGAAGATCGGGTATTCGCCGTGTACCTCTCCGGATGCCCTACTGACC	CTR Crosslink	#13941
EC complementary scaffold, RNA GCAUUCAAAGCGGAGAGGU	CTR Crosslink	#12291
EC complementary scaffold, template strand DNA pCAAGGGACTGGTCTGAATCGGGAATACGGAACGGAAGATCTGAGTTCTCTTC CCCTCTAGCTCAGGACGTACTGACC	Ligated-scaffold transcription	#10160
EC complementary scaffold, non-template strand DNA GGTCAGTACGTCCTGAGCTAGAGGGGAAGAGAACTCAGATCTTCCGTTCCGTA TTCCCGATTCAGACCAGTCC	Ligated-scaffold transcription	#10161
EC complementary scaffold, RNA UUUUUUUGAGCUAGAGG	Ligated-scaffold transcription	#8855
consensus ePEC complementary scaffold, non-template strand DNA GGTCAGTACGTCCGGCATAGTTGCGCCCGTAAATTCAGATCTTCCAGTGG	CTR Crosslink/ pause assay	#9563
consensus ePEC complementary scaffold, template strand DNA CCAGTCATGCAGGCCGTATCAACGCGGGCATTTAAGTCTAGAAGGTCACC	CTR Crosslink/ pause assay	#8334
consensus ePEC complementary scaffold, RNA UUUUUUGGCAUAGUUG	CTR crosslink	#8401
consensus ePEC complementary scaffold, RNA	pause assay	#8342

UUUUUUGGCAUAGUU		
his/ePEC complementary scaffold, non-template strand DNA, complementary GGTCAGTACGTCCTGCGATGTGTGCTGGAAGAGAATTCAGATCTTCCAGT	CTR crosslink	#14074
his/ePEC complementary scaffold, non-template strand DNA, bubbled GGTCAGTACGTCCTATGTCGATCTCTGGAAGAGAATTCAGATCTTCCAGT	CTR crosslink	#14736
his/ePEC complementary scaffold, template strand DNA ACTGGAAGATCTGAATTCTCTTCCAGCACACATCGCAGGACGTACTGACC	CTR crosslink	#14073
his/ePEC complementary scaffold, RNA GUCAUCCGGCGAUGUGUG	CTR crosslink/ pause assay	#12644
his/ePEC complementary scaffold, non-template strand DNA GGTCAGTACGTCCTGATGTGTGCTGGAAGAGATTCAGAGTCTTCCAGTGG	pause assay	#11614
his/ePEC complementary scaffold, template strand DNA CCACTGGAAGACTCTGAATCTCTTCCAGCACACATCAGGACGTACTGACC	pause assay	#11613
his/ePEC complementary scaffold, antisense RNA rCrCrGrGrArUrGrA	CTR crosslink/ pause assay	#6598
Plasmids	**Source**	**Stock #**
pRM756 *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus	(5)	#2956
pRM843	(67)	#5143

*Eco*RNAP (α2ββ′ω) overexpression plasmid containing HMK-Strep tag at rpoC C-terminus and his10-ppx tag at β N-terminus		
pYB101 CPR(SI3–RH): *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus, plus β′D1051C and β′G671C substitutions	This work	#6351
pRM1212 CPR(SI3–SI1): *Eco*RNAP (α2ββ′ω) overexpression plasmid containing HMK-Strep tag at β′ C-terminus and his10-ppx tag at β N-terminus, plus βR267C and β′D1051C substitutions	This work	#6012
pRM1213 CTR: *Eco*RNAP (α2ββ′ω) overexpression plasmid containing HMK-Strep tag at β′ C-terminus and his10-ppx tag at β N-terminus, plus βR267C, β′D1051C and β′G671C substitutions	This work	#6013
pRL785 Encodes β subunit of *E. coli* RNAP. Used as PCR template for preparation of dsDNA for ligation-elongation experiment.	(10, 50)	#1785
pYB106 *Eco*RNAP (α2ββ′ω) overexpression plasmid containing HMK-Strep tag at rpoC C-terminus, his10-ppx tag at β N-terminus and βR267C, β′D1051C, β′G671C, β′E1168S substitutions	This work	#6356
pYB107 *Eco*RNAP (α2ββ′ω) overexpression plasmid containing HMK-Strep tag at β′ C-terminus, his10-ppx tag at β N-terminus and βR267C, β′D1051C, β′G671C, β′F1199A substitutions	This work	#6357
pYB108	This work	#6358

*Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus and the β′F1199A substitution		

2.6 Supplementary figures and tables

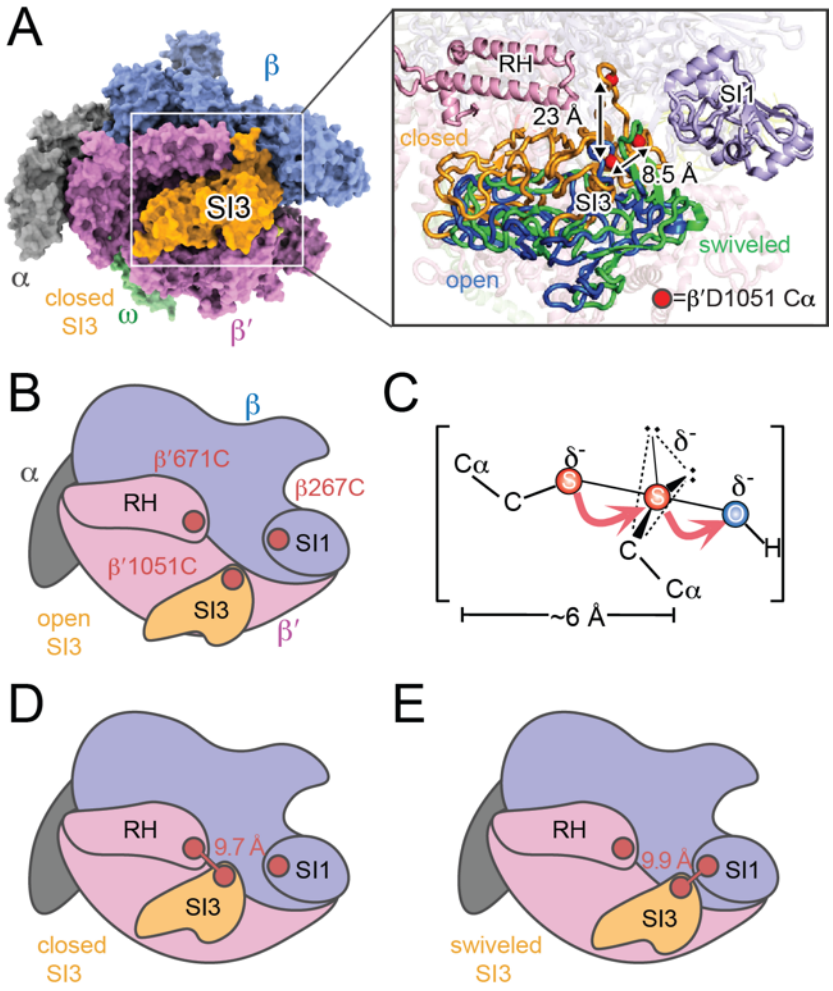


Figure S2.1. Alternative locations of SI3 in ECs and PECs and CTR Cys residues.

(*A*) Locations of SI3 in ECs, RNAP is shown in the TH conformation with SI3 in closed position (pdb 4yln) (33). Inset, superposition of three different SI3 locations: closed (orange, pdb 4yln); open (blue, pdb 6alf), and swiveled (green, pdb 6asx) (23, 32, 33). (*B*) Schematic representation of CTR. (*C*) Reaction geometry of disulfide formation with H_2O_2. (*D*), (*E*) Cα–Cα distances of residues changed to Cys to generate the CTR (red spheres): β′SI3–β′RH (closed) and β′SI3–βSI1 (swiveled).

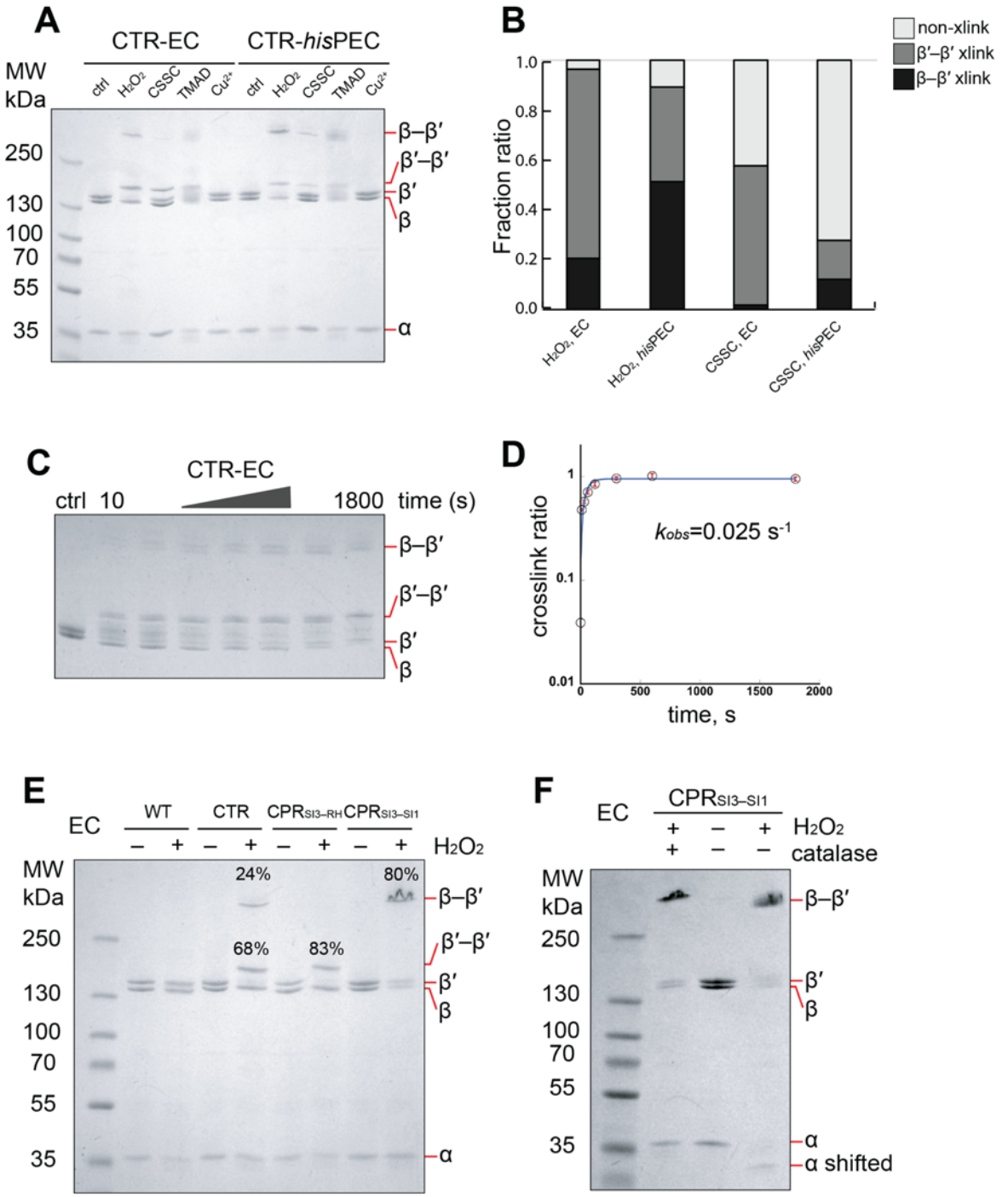
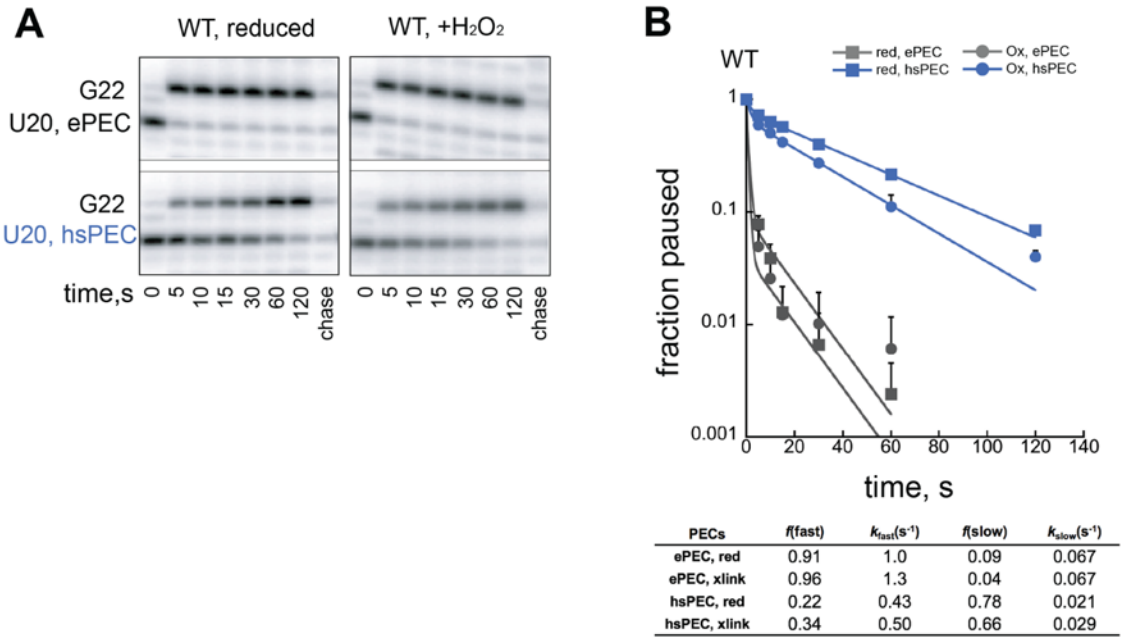

A
CTR-EC
CTR-hisPEC
MW
kDa
ctrl
H2O2
CSSC
TMAD
Cu2+
250
130
100
70
55
35
β–β′
β′–β′
β′
β
α
B
non-xlink
β′–β′ xlink
β–β′ xlink
Fraction ratio
1.0
0.8
0.6
0.4
0.2
0.0
H2O2, EC
H2O2, hisPEC
CSSC, EC
CSSC, hisPEC
C
CTR-EC
ctrl 10
1800 time (s)
β–β′
β′–β′
β′
β
D
crosslink ratio
1
0.1
0.01
kobs=0.025 s-1
0
500
1000
1500
2000
time, s
E
EC
WT
CTR
CPRSI3–RH
CPRSI3–SI1
– + – + – + – +
H2O2
MW
kDa
24%
80%
68%
83%
250
130
100
70
55
35
β–β′
β′–β′
β′
β
α
F
EC
CPRSI3–SI1
+ – +
+ – –
H2O2
catalase
MW
kDa
250
130
100
70
55
35
β–β′
β′
β
α
α shifted

Figure S2.2. Optimization of the CTR crosslink assay using H_2O_2.

(A) CTR crosslink assay of in EC and *his*PEC with different oxidants. CSSC, cystamine; TMAD, diamide. See scaffold sequences in Figs. 2.3 and 2.4. (B) Fractions of crosslinked and uncrosslinked species produced by cystamine or H_2O_2. (C) Time course of H_2O_2-mediated crosslinking in CTR RNAP reconstituted on an EC scaffold. (D) Plot of crosslink ratio as a function of time for CTR RNAP. Fit to a single exponential gives a pseudo-first-order rate of 0.025 s^{-1}. (E) Comparison of crosslink species formed by CTR, $CPR_{SI3–RH}$, and $CPR_{SI3\text{-}SI1}$ RNAPs. (F) Removal of excess H_2O_2 by catalase treatment does not affect crosslink ratio but eliminates a shift in α subunit electrophoretic mobility.



PECs	*f*(fast)	k_{fast}(s^{-1})	*f*(slow)	k_{slow}(s^{-1})
ePEC, red	0.91	1.0	0.09	0.067
ePEC, xlink	0.96	1.3	0.04	0.067
hsPEC, red	0.22	0.43	0.78	0.021
hsPEC, xlink	0.34	0.50	0.66	0.029

Figure S2.3. H_2O_2 minimally affects pause kinetics of WT RNAP at the *his* elemental and hairpin-stabilized pause site.

(*A*) Pause escape by *his* ePEC and hsPEC in reducing and oxidizing conditions. (*B*) Pause kinetics of *his* ePEC and hsPEC in reducing and oxidizing conditions.

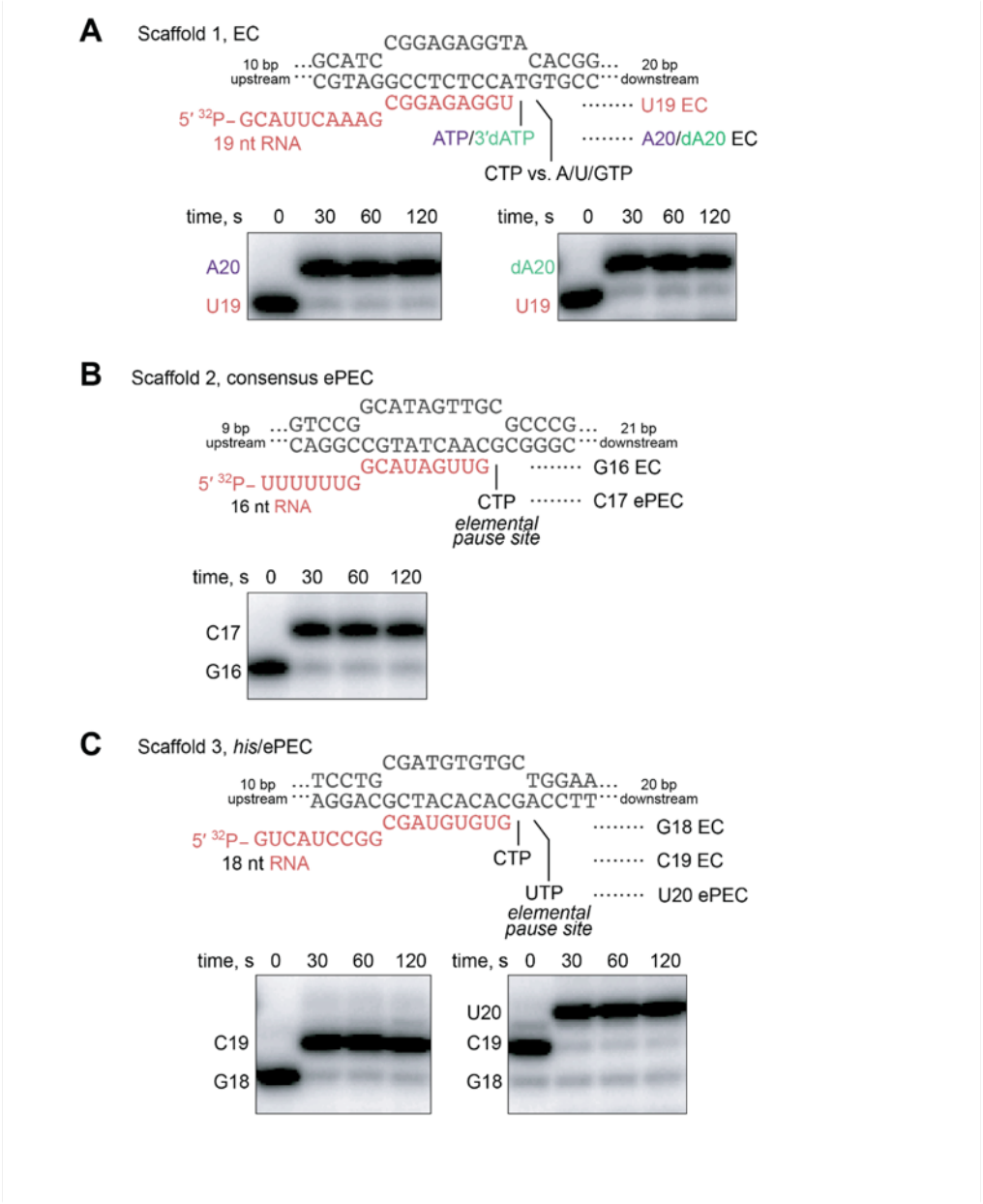


Figure S2.4. RNA extension assays for RNA–DNA scaffolds used in CTR assays in Figs. 2.2, 2.3, and 2.4.

The 5′ ends of the RNAs were radiolabeled prior to reconstitution ECs and then reacted with 100 μM cognate NTPs. Samples were analyzed by denaturing PAGE (***Materials and Methods***) to verify that the desired nucleotide incorporation was complete in the CTR assays.

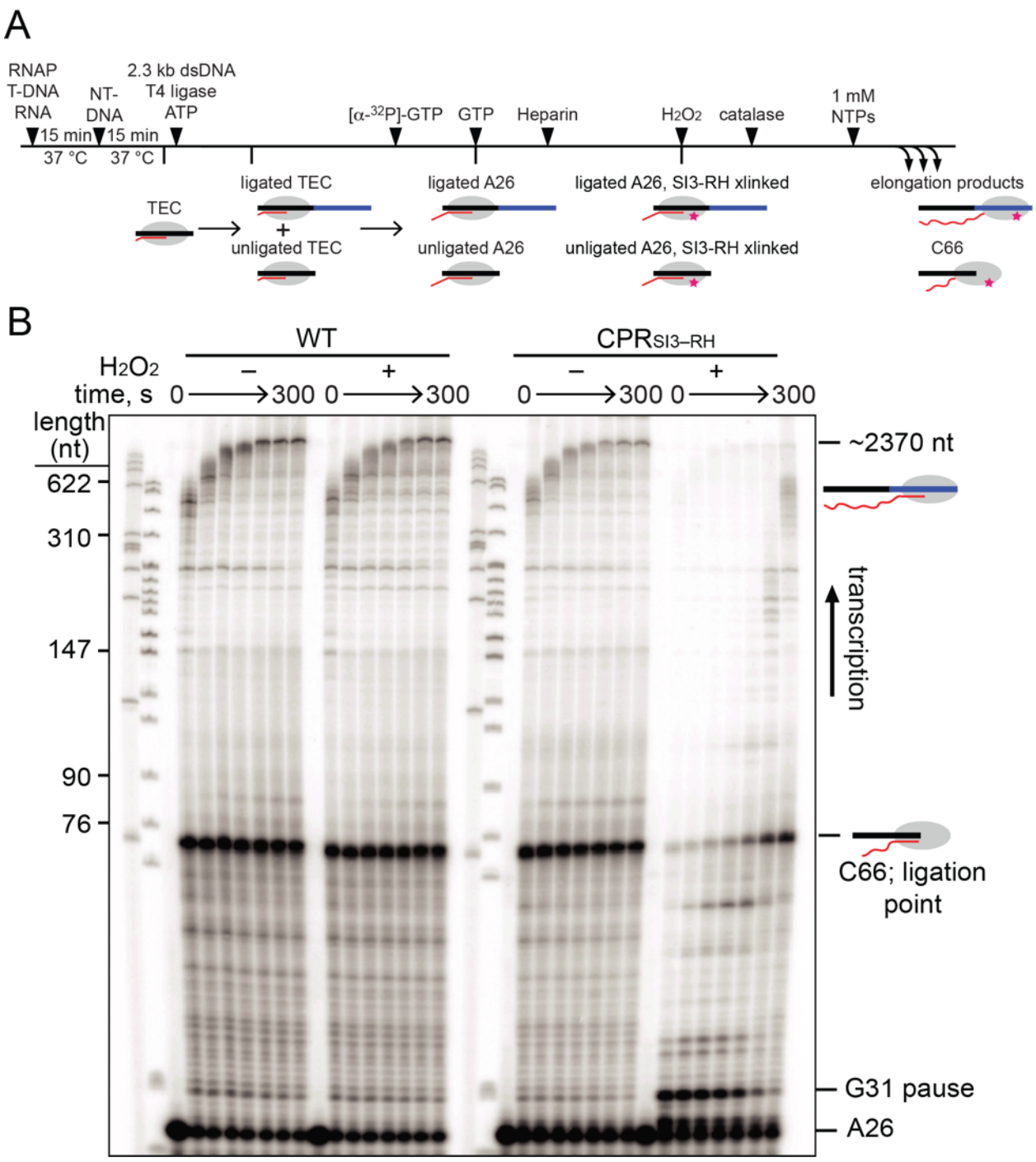


Figure S2.5. Ligated-scaffold transcription assay.

(*A*) Schematic of ligated-scaffold transcription assay. (*B*) Phosphorimage of denaturing gel separating RNAs formed during the ligated-scaffold transcription assay by WT RNAP and $CPR_{SI3–RH}$ in either reducing and oxidizing conditions (***Materials and Methods***).

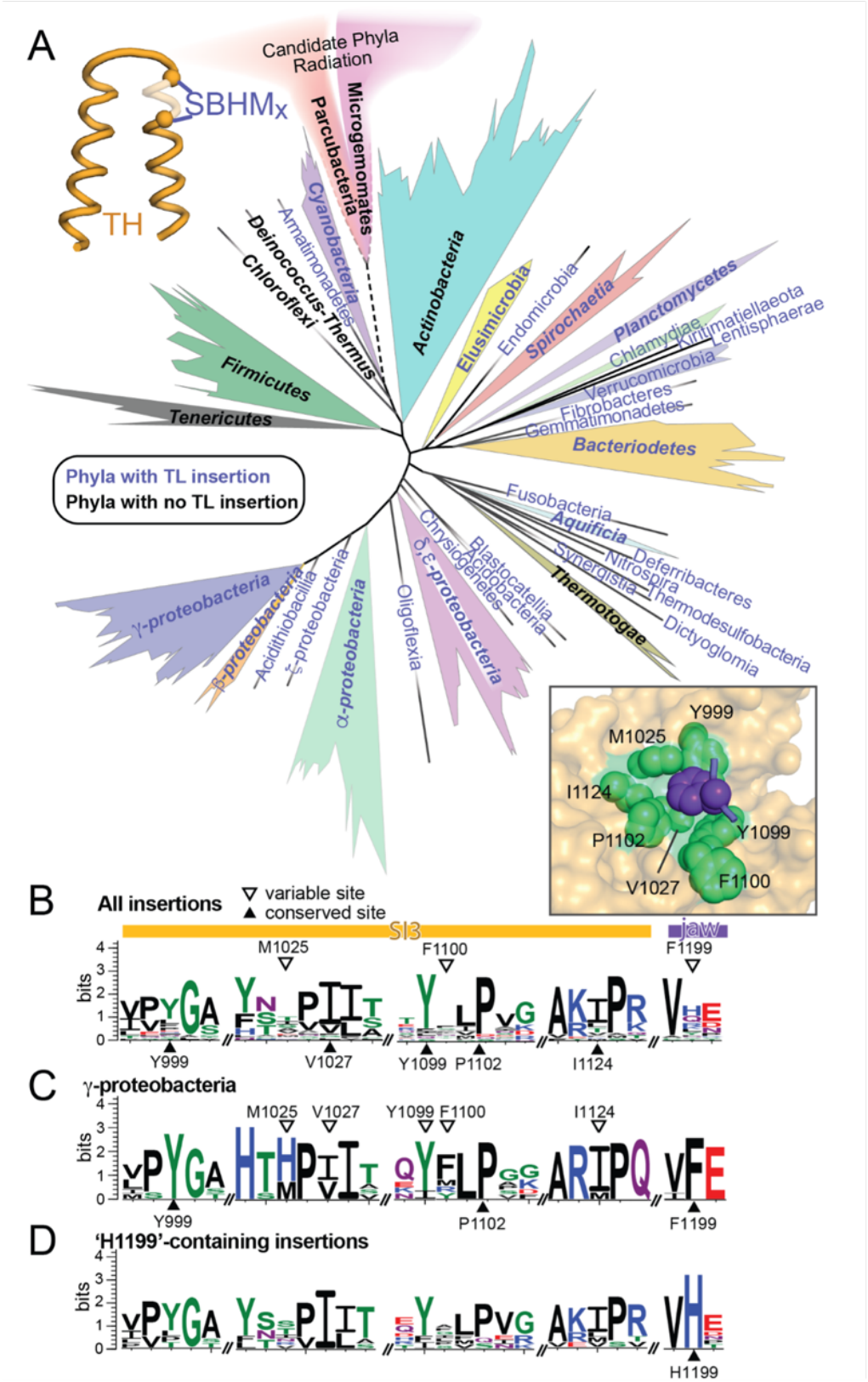

A
SBHMx
TH
Candidate Phyla Radiation
Parcubacteria
Microgemomates
Actinobacteria
Cyanobacteria
Amaltimonadetes
Deinococcus-Thermus
Chloroflexi
Elusimicrobia
Endomicrobia
Spirochaetia
Planctomycetes
Chlamydiae
Kiritimatiellaeota
Lentisphaerae
Verrucomicrobia
Fibrobacteres
Gemmatimonadetes
Bacteriodetes
Firmicutes
Tenericutes
Phyla with TL insertion
Phyla with no TL insertion
Fusobacteria
Aquificia
Deferribacteres
Nitrospira
Synergistia
Thermodesulfobacteria
Dictyoglomia
Thermotogae
Chrysiogenetes
Blastocatellia
Acidobacteria
δ,ε-proteobacteria
Oligoflexia
α-proteobacteria
γ-proteobacteria
β-proteobacteria
Acidithiobacillia
ξ-proteobacteria
Y999
M1025
I1124
P1102
Y1099
V1027
F1100
B
All insertions
variable site
conserved site
SI3
jaw
M1025
F1100
F1199
bits
Y999
V1027
Y1099
P1102
I1124
C
γ-proteobacteria
M1025
V1027
Y1099
F1100
I1124
Y999
P1102
F1199
D
'H1199'-containing insertions
H1199

Figure S2.6. Evolution of TL insertions in RNAP.

(*A*) An evolutionary tree depicting bacterial phylogeny based on 49 universally present genes (58) with lineages in which the TL contains an insertion of two or more SBHM domains indicated in blue. The area covered by each lineage is an approximate reflection of the number of known species in it. Conservation of residues involved in the Phe-pocket interaction: (*B*) in all TL insertion-containing bacteria; (*C*) in γ-proteobacteria; or (*D*) in TL insertion-containing bacteria in which the jaw Phe residue is replaced by His.

Table S2.1. Kinetic fitting parameters for pause assays shown in Figs. 2.5 and 2.6.

Fig. 2.5C

CPR(SI3-SI1)	f_{fast}	error	$k_{fast}(s^{-1})$	error	f_{slow}	error	$k_{slow}(s^{-1})$	error
ePEC, red	0.94	0.020	0.75	0.13	0.056	0.017	0.030	0.016
ePEC, xlink	0.34	0.0091	0.54	0.078	0.66	0.0059	0.0046	0.00018
hsPEC, red	0.28	0.028	0.28	0.060	0.72	0.026	0.022	0.0012
hsPEC, xlink	0.17	0.047	0.38	0.35	0.83	0.031	0.0031	0.00066

Fig. 2.5D

CPR(SI3-RH)	f_{fast}	error	$k_{fast}(s^{-1})$	error	f_{slow}	error	$k_{slow}(s^{-1})$	error
ePEC, red	0.94	0.0086	0.74	0.064	0.058	0.0070	0.022	0.0052
ePEC, xlink	0.92	0.020	0.59	0.076	0.083	0.014	0.011	0.0048
hsPEC, red	0.44	0.092	0.22	0.080	0.56	0.090	0.024	0.0050
hsPEC, xlink	0.92	0.0077	0.59	0.031	0.083	0.0056	0.011	0.0020

Fig. 2.6D

PECs	f_{fast}	error	$k_{fast}(s^{-1})$	error	f_{slow}	error	$k_{slow}(s^{-1})$	error
ePEC, WT	0.97	0.0090	0.73	0.070	0.030	0.0056	0.0047	0.0039
ePEC, F1199A	0.98	0.0029	0.86	0.042	0.018	0.016	0	0.0013
hsPEC, WT	0.55	0.12	0.13	0.031	0.44	0.13	0.026	0.0065
hsPEC, F1199A	0.76	0.066	0.31	0.043	0.24	0.065	0.044	0.013

Fig. S2.2B

WT PEC	f_{fast}	error	$k_{fast}(s^{-1})$	error	f_{slow}	error	$k_{slow}(s^{-1})$	error
ePEC, red	0.91	0.079	1.0	1.7	0.089	0.077	0.067	0.064
ePEC, xlink	0.96	0.071	1.3	6.4	0.040	0.070	0.067	0.13
hsPEC, red	0.22	0.017	0.43	0.11	0.78	0.014	0.021	0.00073
hsPEC, xlink	0.34	0.029	0.50	0.16	0.66	0.026	0.029	0.0019

Table S2.2. Bacterial species used to screen for and align trigger-loop sequence insertions.

Phylum	Genus Species	Accession	β', aa	SI3?	SI3, aa
Thermus	*Thermus thermophilus*	Q8RQE8	1524	No	0
	Deinococcus geothermalis	Q1J0P7	1537	No	0
	Deinococcus radiodurans	Q9RVW0	1546	No	0
Thermotogae	*Petrotoga mobilis*	A9BF32	1643	No	0
	Thermotoga maritima	P36252	1690	No	0
	Thermosipho melanesiensis	A6LKB7	1649	No	0
Aquificae	*Aquifex aeolicus*	O67763	1547	Yes	206
	Balnearium lithotrophicum	WP_142934069.1	1471	Yes	168
	Hydrogenothermus marinus	WP_121922216.1	1602	Yes	207
Actinobacteria	*Mycobacterium tuberculosis*	P9WGY7	1316	No	0
	Streptomyces avermitilis	Q82DQ4	1299	No	0
	Bifidobacterium adolescentis	A1A316	1337	No	0
Chloroflexi	*Dehalococcoides mccartyi*	Q3Z8V3	1295	No	0
	Herpetosiphon aurantiacus	A9B6J1	1543	No	0
	Chloroflexus aurantiacus	A9WH11	1503	No	0
Cyanobacteria	*Gloeobacter violaceus*	Q7NDF7	1262	Yes	633
	Synechococcus elongatus	Q5MZ21	1318	Yes	638
	Trichodesmium erythraeum	Q110H3	1415	Yes	684
Firmicutes	*Clostridioides difficile*	A0A125V165	1161	No	0
	Bacillus subtilis	P37871	1199	No	0
	Mycoplasma pneumoniae	A0A0H3DMP7	1290	No	0
Tenericutes	*Acholeplasma equifetale*	WP_026399639.1	1367	No	0
	Entomoplasma ellychniae	PPE04721.1	1252	No	0
	Haloplasma contractile	ERJ12394.1	1204	No	0
Planctomycetes	*Rhodopirellula baltica*	Q7URW4	1429	Yes	173
	Chlamydia trachomatis	Q3KM48	1396	Yes	190
	Chlamydia abortus	Q5L5I4	1393	Yes	190
Spirochaetes	*Leptospira interrogans*	Q8F0S3	1404	Yes	223
	Borreliella garinii	B7XT60	1377	Yes	224
	Treponema pedis	S5ZS96	1424	Yes	228
Bacteroidetes	*Chlorobium phaeobacteroides*	A1BD24	1492	Yes	190
	Bacteroides fragilis	Q5L898	1427	Yes	180
	Flavobacterium johnsoniae	A5FIJ4	1436	Yes	180
δ-proteobacteria	*Sorangium cellulosum*	A9GRB4	1427	Yes	185
	Geobacter sulfurreducens	I7EPD8	1395	Yes	182
	Desulfovibrio vulgaris	Q727C6	1385	Yes	182
ε-proteobacteria	*Wolinella succinogenes*	Q7MA56	2883	Yes	332
	Sulfurovum lithotrophicum	WP_046551950.1	1509	Yes	333
	Campylobacter hominis	WP_041570540.1	1509	Yes	331
α-proteobacteria	*Ehrlichia ruminantium*	Q5HC04	1411	Yes	186
	Zymomonas mobilis	A0A0H3FXE3	1391	Yes	178

	Caulobacter vibrioides	Q9AAU1	1396	Yes	183
β-proteobacteria	*Acidovorax delafieldii*	C5T742	1408	Yes	188
	Herminiimonas arsenicoxydans	A4G9U4	1414	Yes	189
	Bordetella petrii	A9IJ22	1412	Yes	188
γ-proteobacteria	*Legionella pneumophila*	Q5ZYP9	1401	Yes	188
	Aliivibrio fischeri	Q5E239	1401	Yes	188
	Escherichia coli	P0A8T7	1407	Yes	188
Zetaproteo	*Mariprofundus ferrooxydans*	WP_018295044.1	1399	Yes	189
Oligoflexia	*Bdellovibrio bacteriovorus*	WP_063244102.1	1376	Yes	183
Chrysiogenetes	*Chrysiogenes arsenatis*	WP_084417671.1	1347	Yes	175
Acidobacteriia	*Acidobacterium capsulatum*	C1F3Y4	1389	Yes	173
Blastocatellia	*Pyrinomonas methylaliphatogenes*	WP_041973492.1	1389	Yes	173
Dictyoglomia	*Dictyoglomus thermophilum*	WP_012547665	1429	Yes	258
Synergistia	*Aminomonas paucivorans*	WP_006301288.1	1654	Yes	314
Thermodesulfobacteria	*Thermodesulfobacterium commune*	WP_038061948.1	1353	Yes	183
Nitrospira	*Nitrospira defluvii*	CBK41054.1	1396	Yes	197
Deferribacteres	*Caldithrix abyssi*	APF17647.1	1429	Yes	175
Fusobacteriia	*Fusobacterium necrophorum*	RXZ69643.1	1322	Yes	165
Gemmatimonadetes	*Gemmatimonas aurantiaca*	BAH37898.1	1436	Yes	182
Fibrobacteres	*Fibrobacter succinogenes*	WP_173384178.1	1480	Yes	178
Verrucomicrobia	*Verrucomicrobium spinosum*	WP_009964681.1	1376	Yes	174
Lentisphaerae	*Lentisphaera araneosa*	WP_007279810.1	1423	Yes	174
Kiritimatiellaeota	*Kiritimatiella glycovorans*	AKJ64962.1	1383	Yes	174
Chlamydiae	*Chlamydia abortus*	ASD30784.1	1393	Yes	190
Endomicrobia	*Endomicrobium proavitum*	WP_052571414.1	1580	Yes	184
Elusimicrobia	*Elusimicrobium minutum*	WP_012415591	1385	Yes	173
Armatimonadetes	*Armatimonas rosea*	WP_184203436.1	1473	Yes	52
Acidithiobacillia	*Acidithiobacillus caldus*	AUW32002.1	1382	Yes	189
Candidate phyla radiation	Parcubacteria	KND51703.1	1253	No	0
Candidate phyla radiation	Microgenomates	RJR29364.1	1257	No	0

2.7 References

1. S. A. Seibold *et al.*, Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *Biochim Biophys Acta* **1799**, 575-587 (2010).

2. D. G. Vassylyev *et al.*, Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163-168 (2007).

3. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941-954 (2006).

4. G. A. Belogurov, I. Artsimovitch, The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J Mol Biol* **431**, 3975-4006 (2019).

5. T. A. Windgassen *et al.*, Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res* **42**, 12707-12721 (2014).

6. J. Zhang, M. Palangat, R. Landick, Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* **17**, 99-104 (2010).

7. M. Chlenov *et al.*, Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. *J Mol Biol* **353**, 138-154 (2005).

8. L. M. Iyer, E. V. Koonin, L. Aravind, Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* **3**, 1 (2003).

9. W. J. Lane, S. A. Darst, Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J Mol Biol* **395**, 671-685 (2010).

10. I. Artsimovitch, V. Svetlov, K. S. Murakami, R. Landick, Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* **278**, 12344-12355 (2003).

11. R. Landick, The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**, 1062-1066 (2006).

12. M. H. Larson *et al.*, A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).

13. J. Y. Kang, T. V. Mishanina, R. Landick, S. A. Darst, Mechanisms of Transcriptional Pausing in Bacteria. *J Mol Biol* **431**, 4007-4029 (2019).

14. I. Jonkers, J. T. Lis, Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 167-177 (2015).

15. A. Mayer, H. M. Landry, L. S. Churchman, Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr Opin Cell Biol* **46**, 72-80 (2017).

16. J. Zhang, R. Landick, A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci* **41**, 293-310 (2016).

17. R. Landick, J. Carey, C. Yanofsky, Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc Natl Acad Sci U S A* **82**, 4663-4667 (1985).

18. S. Proshkin, A. R. Rahmouni, A. Mironov, E. Nudler, Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504-508 (2010).

19. I. Gusarov, E. Nudler, The mechanism of intrinsic transcription termination. *Mol Cell* **3**, 495-504 (1999).

20. N. J. Proudfoot, Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).

21. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *eLife* **8**, e40981 (2019).

22. X. Guo *et al.*, Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e814 (2018).

23. J. Y. Kang *et al.*, RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* **69**, 802-815 e801 (2018).

24. M. Abdelkareem *et al.*, Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell* **75**, 298-309 e294 (2019).

25. I. Artsimovitch, R. Landick, Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**, 7090-7095 (2000).

26. N. Komissarova, M. Kashlev, Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A* **94**, 1755-1760 (1997).

27. E. Lerner *et al.*, Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **113**, E6562-E6571 (2016).

28. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**, 33-41 (1997).

29. C. L. Chan, R. Landick, The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J Biol Chem* **264**, 20796-20804 (1989).

30. T. M. Conrad *et al.*, RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proc Natl Acad Sci U S A* **107**, 20500-20505 (2010).

31. J. Ederth, I. Artsimovitch, L. A. Isaksson, R. Landick, The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J Biol Chem* **277**, 37456-37463 (2002).

32. J. Y. Kang *et al.*, Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *eLife* **6**, e25478 (2017).

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

34. A. Mazumder, M. Lin, A. N. Kapanidis, R. H. Ebright, Closing and opening of the RNA polymerase trigger loop. *Proc Natl Acad Sci U S A* **117**, 15642-15649 (2020).

35. J. Y. Kang *et al.*, Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e1614 (2018).

36. M. J. Bellecourt, A. Ray-Soni, A. Harwig, R. A. Mooney, R. Landick, RNA Polymerase Clamp Movement Aids Dissociation from DNA but Is Not Required for RNA Release at Intrinsic Terminators. *J Mol Biol* **431**, 696-713 (2019).

37. D. Nayak, M. Voss, T. Windgassen, R. A. Mooney, R. Landick, Cys-pair reporters detect a constrained trigger loop in a paused RNA polymerase. *Mol Cell* **50**, 882-893 (2013).

38. S. Sekine, Y. Murayama, V. Svetlov, E. Nudler, S. Yokoyama, The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol Cell* **57**, 408-421 (2015).

39. A. P. Wiita, S. R. Ainavarapu, H. H. Huang, J. M. Fernandez, Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc Natl Acad Sci U S A* **103**, 7222-7227 (2006).

40. R. M. Sterritt, J. N. Lester, Interactions of heavy metals with bacteria. *Sci Total Environ* **14**, 5-17 (1980).

41. D. Luo, S. W. Smith, B. D. Anderson, Kinetics and mechanism of the reaction of cysteine and hydrogen peroxide in aqueous solution. *J Pharm Sci* **94**, 304-316 (2005).

42. D. E. Heck, M. Shakarjian, H. D. Kim, J. D. Laskin, A. M. Vetrano, Mechanisms of oxidant generation by catalase. *Ann N Y Acad Sci* **1203**, 120-125 (2010).

43. B. Wang, A. V. Predeus, Z. F. Burton, M. Feig, Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophys J* **105**, 767-775 (2013).

44. I. Toulokhonov, J. Zhang, M. Palangat, R. Landick, A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell* **27**, 406-419 (2007).

45. A. M. Malinen *et al.*, Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res* **40**, 7442-7451 (2012).

46. K. D. Westover, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481-489 (2004).

47. C. D. Kaplan, K. M. Larsson, R. D. Kornberg, The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell* **30**, 547-556 (2008).

48. M. L. Kireeva *et al.*, Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell* **30**, 557-566 (2008).

49. M. H. Larson *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc Natl Acad Sci U S A* **109**, 6555-6560 (2012).

50. T. V. Mishanina, M. Z. Palo, D. Nayak, R. A. Mooney, R. Landick, Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci U S A* **114**, E5103-E5112 (2017).

51. C. L. Chan, D. Wang, R. Landick, Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J Mol Biol* **268**, 54-68 (1997).

52. D. Wang *et al.*, Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell* **81**, 341-350 (1995).

53. P. P. Hein, M. Palangat, R. Landick, RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. *Biochemistry* **50**, 7002-7014 (2011).

54. B. Liu, Y. Zuo, T. A. Steitz, Structures of E. coli sigmaS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc Natl Acad Sci U S A* **113**, 4051-4056 (2016).

55. R. L. Gourse *et al.*, Transcriptional Responses to ppGpp and DksA. *Annu Rev Microbiol* **72**, 163-184 (2018).

56. Z. Pasman, P. H. von Hippel, Active Escherichia coli transcription elongation complexes are functionally homogeneous. *J Mol Biol* **322**, 505-519 (2002).

57. K. J. Harrington, R. B. Laughlin, S. Liang, Balanced branching in transcription termination. *Proc Natl Acad Sci U S A* **98**, 5019-5024 (2001).

58. E. Khaledian, K. A. Brayton, S. L. Broschat, A systematic approach to bacterial phylogeny using order level sampling and identification of HGT using network science. *Microorganisms* **8**, e8020312 (2020).

59. A. V. Yakhnin *et al.*, NusG controls transcription pausing and RNA polymerase translocation throughout the Bacillus subtilis genome. *Proc Natl Acad Sci U S A* **117**, 21628-21636 (2020).

60. A. Czyz, R. A. Mooney, A. Iaconi, R. Landick, Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**, e00931 (2014).

61. E. Burova, S. C. Hung, V. Sagitov, B. L. Stitt, M. E. Gottesman, Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J Bacteriol* **177**, 1388-1392 (1995).

62. G. E. Johnson, J. B. Lalanne, M. L. Peters, G. W. Li, Functionally uncoupled transcription-translation in *Bacillus subtilis*. *Nature* **585**, 124-128 (2020).

63. J. L. Payne, A. Wagner, The causes of evolvability and their evolution. *Nature reviews. Genetics* **20**, 24-38 (2019).

64. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *Elife* **8** (2019).

65. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

66. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).

67. P. P. Hein *et al.*, RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* **21**, 794-802 (2014).

Chapter 3: *E. coli* SI3 controls global transcriptional pausing *in vivo*

Adapted from:

Yu Bao and Robert Landick. (2023) *E. coli* RNA polymerase SI3 domain controls global transcriptional pausing.

Manuscript in preparation for submission to:

Nucleic Acids Research

3.1 Abstract

Transcriptional pausing aids gene regulation by cellular RNA polymerases (RNAPs). In many bacteria, a surface-exposed domain inserted into the catalytic trigger loop (TL) of RNAP, called SI3 in *Escherichia coli,* modulates pausing and is essential for growth. Here we describe a viable *E. coli* strain lacking SI3 enabled by a suppressor TL substitution (β′Ala941→Thr; ΔSI3*). ΔSI3* increased transcription rate *in vitro* relative to ΔSI3, possibly explaining its viability, but retained both positive and negative effects of ΔSI3 on pausing. ΔSI3* inhibited pauses stabilized by nascent RNA structures (pause hairpins; PHs) but enhanced other pauses. Using NET-seq, we found that ΔSI3*-enhanced pauses resemble the consensus elemental pause sequence whereas sequences at ΔSI3*-suppressed pauses, which exhibited greater association with PHs, were more divergent. ΔSI3*-suppressed pauses also were associated with apparent pausing one nt upstream from the consensus sequence, often generating tandem pause sites. These '–2 pauses' were stimulated by pyrophosphate in vitro and by addition of apyrase to degrade residual NTPs during NET-seq sample processing. We propose that some pauses are readily reversible by pyrophosphorolysis or single-nucleotide cleavage. Our results document multiple ways that SI3 modulates pausing *in vivo* and may explain differences in consensus pause sequences found in some NET-seq studies.

3.2 Introduction

Transcriptional pausing is a regulatory mechanism shared by all cellular DNA-dependent RNA polymerases (RNAPs) (1-3). Upon encountering pause signals during transcript elongation, RNAPs can enter a paused state that delays RNA synthesis by factors of ~20–1000 (**Figure 3.1**) (4, 5). A consensus elemental pause (ce-pause) signal has been identified with a sequence motif whose most conserved elements are 5′-$G_{-11}G_{-10}t_{-3}g_{-2}Y_{-1}G_{+1}$ (–1 corresponds to 3′ end position of nascent RNA) (4, 6). Multiple sequence features, including the downstream fork junction, upstream fork junction, DNA-RNA hybrid, and downstream DNA, additively contribute to the strength of an elemental pause (5). The ce-pause sequence was identified in *E. coli*. Although it affects diverse RNAPs from bacterial to human (4), the consensus sequence has been difficult to discern in NET-seq studies of some organisms; pauses sometimes appear to occur just before instead of just after the consensus Y_{-1} (7-9).

Pausing provides time windows for dissociable transcription factors to bind RNAP, for nascent RNA to fold, for RNAP to backtrack, or for elemental paused elongation complexes (ePECs) to rearrange into other types of transcriptional pauses (1, 10, 11). Hairpin-stabilized PECs (hs-PECs) form when the nascent RNA folds into a stem–loop secondary structure (pause hairpin, PH) within the RNA-exit channel of RNAP (12). Hs-pauses guide formation of alternative RNA structures, couple RNAP to translating ribosomes, and aid both intrinsic transcription termination and attenuation mechanisms (13, 14). Cryo-EM analyses of ePECs and hs-PECs have revealed a global conformational change in RNAP called swiveling that inhibits catalysis (15-17). Swiveling, which is stabilized by PHs, involves rotation by a few degrees of interconnected modules in the PEC, including the clamp, dock, shelf, jaw, and lineage-specific sequence insertion 3 (SI3) in *E. coli* (15).

SI3 in *E. coli* contains two sandwich-barrel hybrid motifs (SBHMs) (18). SBHMs are inserted in the catalytic trigger loop (TL) of RNAP of many bacteria phyla except Firmicutes, Actinobacteria, Deinococcus-Thermus, Thermotogae, and Chloroflexi (18-20) (**Figure S3.1**). SI3's location on the surface of RNAP changes in coordination with folding and unfolding of the TL into which SI3 is inserted. Upon TL folding, which greatly promotes catalysis of nucleotide addition (Figure 1), SI3 breaks contact with the RNAP jaw module and contacts the RNAP rim helices (RHs).

Multiple lines of evidence suggest SI3 modulates transcriptional pausing in different ways (15, 20-22). SI3 is proposed to be a key connector between RNAP swiveling and TL conformational change in a PECs; SI3–jaw rotates

as part of the swivel module and thus stabilizes the unfolded TL (15, 20, 22). Conversely, SI3 may stabilize the folded TL in pretranslocated PECs, which form on strong ce-pause sequences (16). Deletion of SI3 or disruption of its connection with other swivel-module parts greatly reduces hs-pause strength *in vitro* (20, 22, 23). However, deletion of SI3 in viable *E. coli* cells has been unsuccessful to date, possibly due to SI3 participation in multiple transcription stages (23-25).

Here we describe the construction of a viable SI3-deleted *E. coli* strain and its use to quantify changes transcriptional pausing *in vivo* compared to wild type using NET-seq. Consistent with different roles of SI3 in pausing, deletion of SI3 increases pausing at some sites and decreased pausing at others. The ΔSI3-enhanced pause sequences strongly resembled the ce-pause sequence motif. The ΔSI3-suppressed pauses, some of which are known hs-pauses, were associated with a greater likelihood of RNA structure formation at PH locations and with what appears to be pausing one nucleotide upstream in the ce-pause sequence (–2 pauses). Our experimental results support multiple roles of SI3 in modulating pausing globally and document how –2 pauses may arise because some pauses are sensitive to pyrophosphorolysis including during NET-seq sample processing.

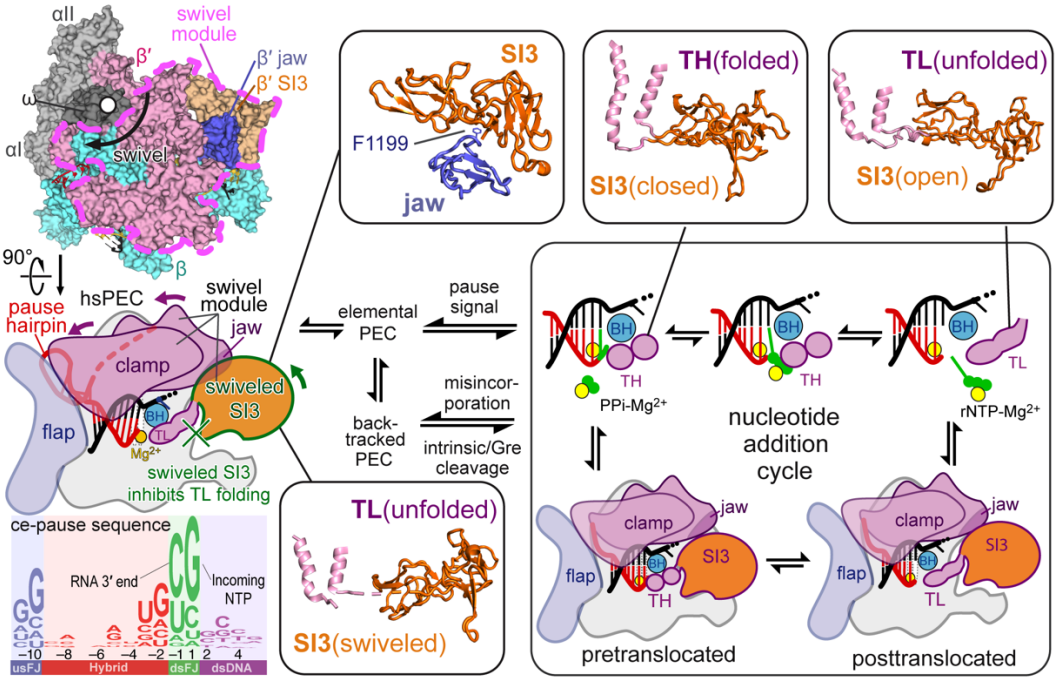


Figure 3.1. The positional changes of SI3 contributes to transcriptional elongation and pausing.

The box on the right shows that SI3 occupies two major positions during a nucleotide addition cycle as closed and open positions (also known as in and out positions) when TL is folded (pdb: 6rh3) and unfolded (pdb: 6rin), respectively. The folding process of TL is aided by the cognate NTP binding to the active center. The formula in the middle summarizes the conversion of different EC states when encountering pause signals. The left panel shows the swiveling of hairpin-stabilized PEC from the front and the top views (derived from pdb: 6asx). The swiveling of SI3 through swiveling module (pink contour in the top view) is mediated by jaw-SI3 specific interaction (upper box in the middle panel). The SI3 swiveling in hs-pause stabilizes TL in the unfolded conformation, thus enhancing the pause.

3.3 Results

3.3.1 A viable *E. coli* strain lacking SI3

Deletion of SI3 from E. coli RNAP (ΔSI3; β′Δ[R943–G1130]) decreases the dwell time of the *his* hs-pause ~3 fold *in vitro* (21, 22) but is inviable *in vivo* (23). To study the effects of ΔSI3 RNAP on pause kinetics *in vivo*, we sought to delete SI3 from the *E. coli* genome using Scarless Cas9-Assisted Recombineering (no-SCAR) (26). This method combines λ-red recombination with CRISPR-guided selection, allowing gene editing without selection markers. As a positive control, a viable strain with an SI3 truncation, del27 (β′Δ[T1045–L1053]) (27) was constructed in parallel (**Figure 3.2A**). We easily obtained positive colonies of the del27 strain (25 out of 46 screened colonies), which were verified by PCR–sequencing. In contrast, we isolated only a single colony (1 out of 177 screened colonies) carrying the full-length SI3 deletion. The recovered deletion also encoded a single amino-acid substitution, β′A941T, in the ΔSI3 TL (**Figure 3.2B**). For conciseness, we refer to β′A941T-ΔSI3 as ΔSI3* hereafter. To confirm that β′A941T is required for growth of ΔSI3* *E. coli*, we re-performed no-SCAR recombineering in parallel using ΔSI3 and ΔSI3* (β′A941T-ΔSI3) recombination targets. Consistent with a requirement of the suppressor substitution β′A941T for viability, 36 of 48 colonies isolated using the ΔSI3* (β′A941T-ΔSI3) target contained ΔSI3* but none of the 48 colonies screened using the ΔSI3 target contained the deletion. However, ΔSI3* strains produced smaller colonies than wild type (*E. coli* K-12 MG1655) when grown on rich medium agar plates at 37 °C (**Figure 3.2C**). In liquid culture, ΔSI3* grew exponentially at a rate similar to wild type in rich medium (determined by apparent OD_{600}) but with significantly smaller cell sizes (**Figure 3.2D** and **Figure S3.2A-C**).

To probe the effect of the β′A941T substitution alone, we exogenously expressed wild-type, ΔSI3, ΔSI3* and β′A941T *rpoC* genes from a plasmid in *E. coli* strain RL602. RL602 is unable to express native *rpoC* at >37 °C (28). Unexpectedly, β′A941T alone could not support cell growth, but produced viable colonies when combined with ΔSI3 (ΔSI3*) (**Figure 3.2E**). Consistent with previous results (23), ΔSI3 *rpoC* also could not support cell growth. We conclude that a ΔSI3 strain is viable only when accompanied by suppressor substitution like β′A941T.

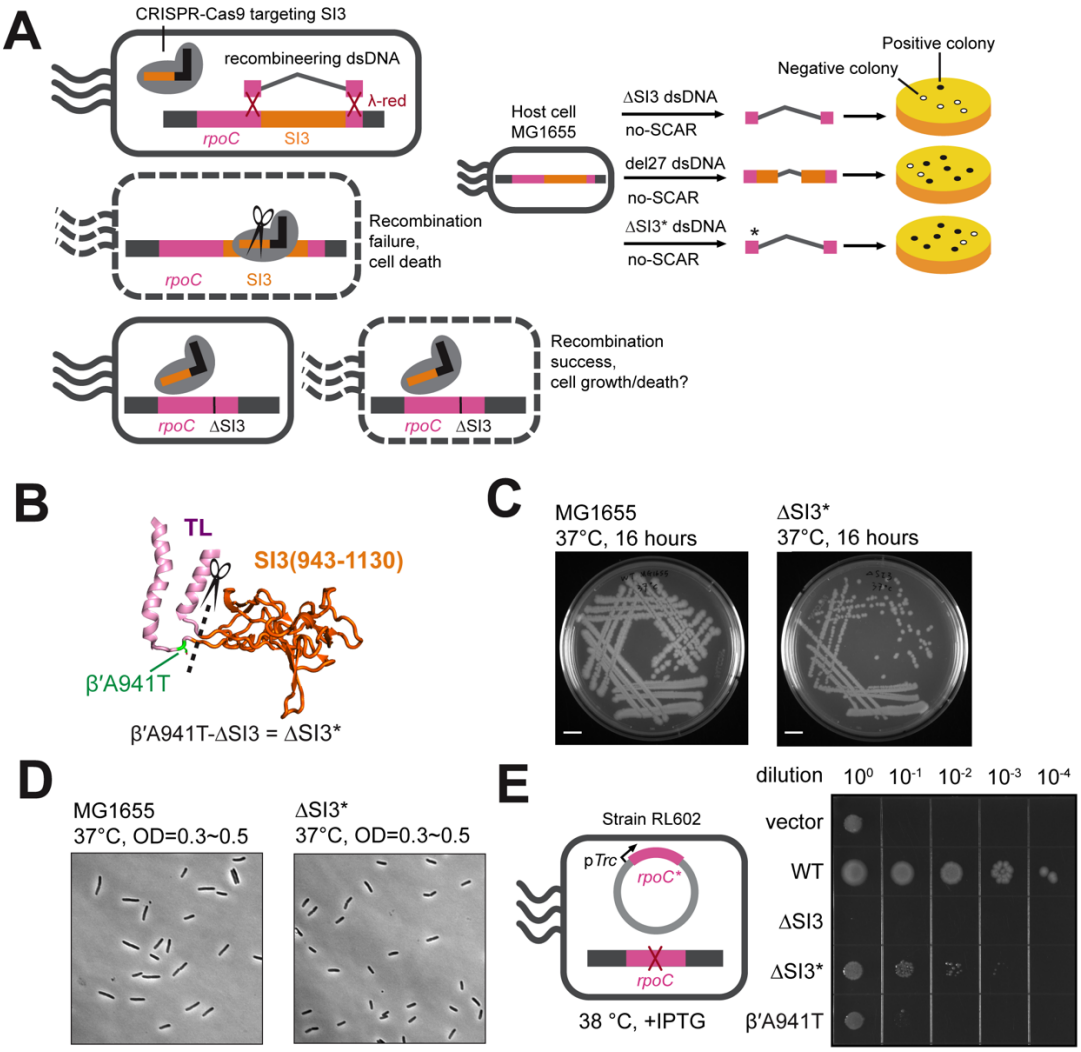


Figure 3.2. Construction of an SI3-deleted *E. coli* strain.

(A) Left panel, schematic diagram of deletion SI3 with no-SCAR method. A DNA sequence in genomic SI3 region is targeted by CRISPR-Cas9, and SI3 region is switched with a recombineering ΔSI3 DNA fragment with λ-Red system. Cells with unsuccessful recombination will be killed by CRISPR-Cas9. Cells with recombination will have uncertain outcomes. Right panel, no-SCAR method with various recombineering DNA fragments showed different results. We only got one colony with ΔSI3 which contained a natural β′A941T mutation (ΔSI3*). No-SCAR with ΔSI3* or del27 recombineering DNAs yielded large numbers of positive colonies (black dots). (B) The location of β′A941T mutation in TL. (C) Colony size comparison between WT and ΔSI3* strains constructed by no-SCAR

method. (D) Microscopy images of WT and ΔSI3* cells in log-phase, bar representing 5 μm. (E) Complementarity experiment of various *rpoC* genes. The plasmids harboring inducible *rpoC* gene varients were expressed in RL602, of which native *rpoC* expression is suppressed at 38 °C.

3.3.2 ΔSI3* RNAP exhibited changes in elongation and pausing *in vitro*

To compare the transcriptional properties of β′A941T, ΔSI3, and ΔSI3* to WT, we purified recombinant RNAPs and conducted *in vitro* transcription elongation assays with an ~2.3 kb DNA template (20, 29). Both β′A941T and ΔSI3 RNAPs exhibited decreased elongation rates (33% and 86% of WT, respectively). In contrast, ΔSI3* RNAP exhibited an elongation rate indistinguishable from WT (**Figure 3.3A**). We postulate that the impaired average elongation rate of ΔSI3 RNAP reflects aberrant TL conformational dynamics that alter the balance between folded and unfolded states. Such imbalance may contribute to lethality. β′A941T may function as a suppressor by also altering TL dynamics in a way that itself is inviable but becomes viable when combined with ΔSI3 by restoring a balance of folded and unfolded TL states. The β′A941T substitution is not evident in TL sequences from bacteria lacking SI3 (**Figure S3.1**). However, its position and the two adjacent residues ($AA_{941}S$ in *E. coli*) undergo major conformational change during TL folding–unfolding and are both bulkier and more functionalized on average in TL sequences lacking SI3 (TAK vs. VAG; **Figure S3.1**). It also is possible that ΔSI3 alters other key RNAP processes (e.g., initiation, termination, or proofreading) and that β′A941T could affect these processes.

We next asked if ΔSI3* RNAP exhibited reduced hs-pausing like ΔSI3 RNAP. We found that ΔSI3* RNAP decreased the *his* hs-pause more than ΔSI3 RNAP (~27% vs. ~50% of WT RNAP pause strength, respectively; Figure 3B, middle panel). PHs stimulate pausing when the PH stem extends to –13, –12, or –11 (the position in PH stem closest to the RNA 3′ end, which is –1) (1, 30). We used different antisense RNA oligos to mimic PHs formed at the –11, –12 and –13 positions. ΔSI3* RNAP reduced pause strength for all three by factors of ~2–3 (**Figure 3.3B**, right panel); the –12 PH mimic gave the greatest pause stimulation. We conclude that, like ΔSI3 RNAP, ΔSI3* RNAP inhibits hs-pauses.

SI3 is known to mediate effects of PHs via swiveling, so this effect of ΔSI3* is as expected. In contrast, the ce-pause appears to favor a pretranslocated rather than swiveled state (16). Thus, we next probed how ΔSI3* RNAP responds to a ce-pause signal. ΔSI3* RNAP increased ce-pause strength (1.4-fold greater than wild type) (**Figure 3.3C**), in contrast to its reduced hs-pausing. We conclude that ΔSI3* also mimics ΔSI3 in increasing ce-pausing possibly by increasing folded-TL stabilization of the pretranslocated ce-PEC (16).

As a final check of ΔSI3* *in vitro*, we verified whether it inhibits PH effects at sites distinct from the ce-pause. The *his* hs-pause exhibits elemental pausing when the PH is ablated (12). Therefore we asked how ΔSI3* RNAP affects a *his* hs-pause mutant with little or no pausing in the absence of the PH. For this purpose, we replaced the *his* pause 3′ rU with rA, changed the downstream DNA sequence, and used a –12 PH mimic oligonucleotide (asRNA) to control PH formation. Without the asRNA, this scaffold does not trigger detectable pausing (**Figure 3.3D**). Addition of an asRNA to form the –12 PH mimic caused measurable pausing by WT RNAP but not by ΔSI3* RNAP. This result confirms that ΔSI3* RNAP reacts differently to hs-pauses vs. ce-pauses and that ΔSI3* reduces pausing solely dependent on PH-induced swiveling.

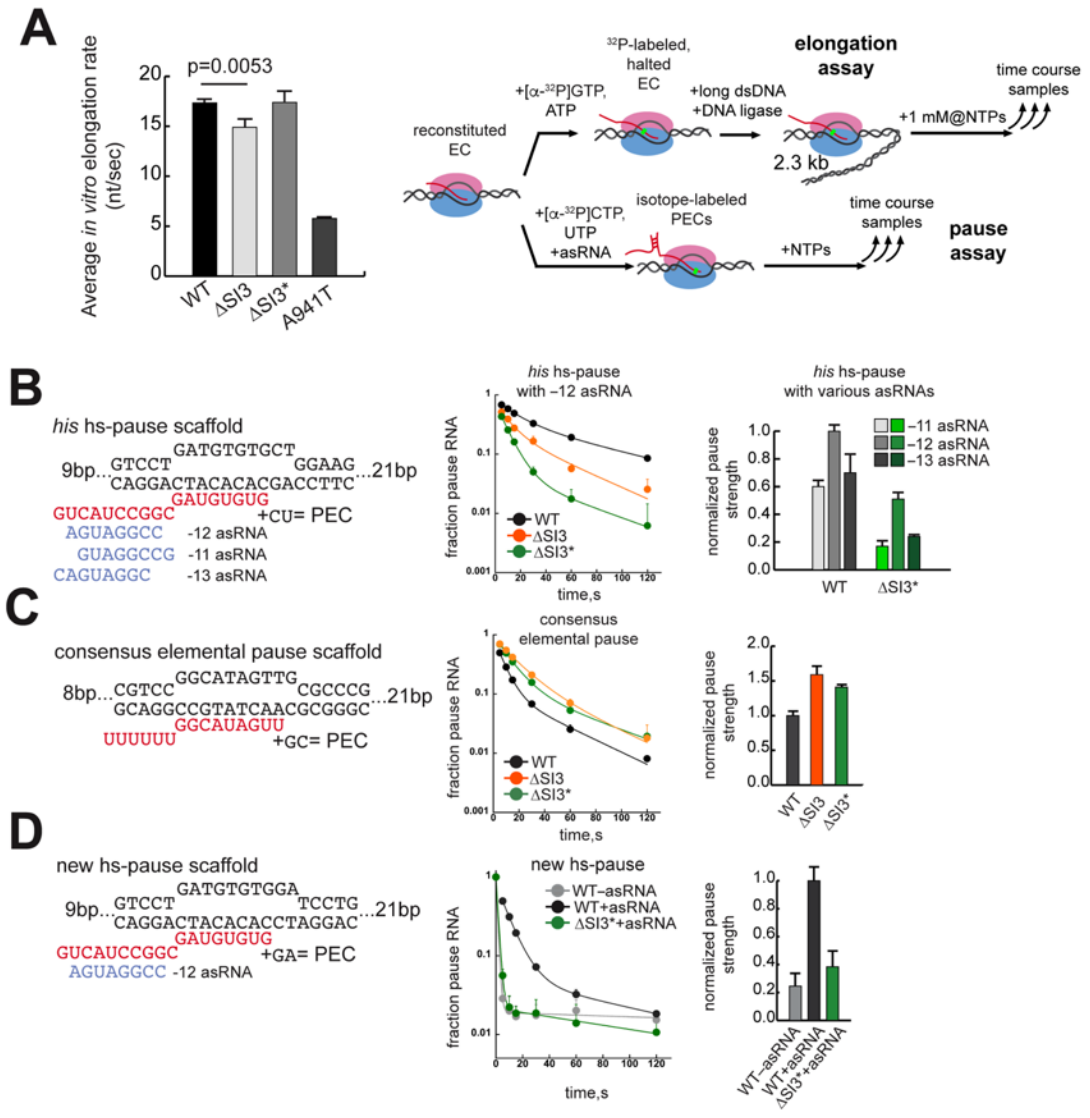


Figure 3.3. ΔSI3* RNAP has comparable average elongation rate to WT, and comparable pause kinetics to ΔSI3 RNAP.

(A) Left panel, average elongation rates of various RNAPs determined by *in vitro* elongation assay. Results represent mean ± SD (n = 3) and the p value is from two-tailed t-test. Right panel, schematic diagram of the elongation assay and the pause assay. In elongation assays, the isotope labeled halted EC is ligated to a 2.3 kb DNA template before re-initiating elongation and elongation rates is based on EC movement over a long range. In pause assays, the PEC is reconstituted on pause scaffolds and the kinetics of pause escaping was measured. (B) ΔSI3*

RNAP *his* hs-pause assay. From left to right shows the pause scaffold, the pause fraction change through time and comparison of pause strength between WT and ΔSI3* RNAP in conditions with various pause hairpin positions. Results represent mean ± SD (n = 3). (C) ΔSI3* RNAP ce-pause assay. (D) ΔSI3* RNAP pause assay with a new hs-pause scaffold different from a consensus elemental pause sequence motif.

3.3.3 ΔSI3* caused global changes in RNAP pausing profiles

To ask how ΔSI3* affects different classes of pause sites *in vivo*, in particular hs-pauses, we next used NET-seq to catalog pauses and apparent pause strengths in wild-type and ΔSI3* strains. To create strains for NET-seq, we used a different recombination method to modify and FLAG-tag native *rpoC* (31). This strain (*rpoC*ΔSI3*::3×FLAG; ΔSI3*-NETseq strain) also exhibited smaller colonies than wild-type (*rpoC*::3×FLAG; WT NET-seq strain), consistent with the phenotypes of the strains constructed using the no-SCAR method (**Figure S3.2D**).

We performed NET-seq on the WT and ΔSI3* NET-seq strains in biological triplicate using a modification of the previously described method (4) that enabled quantitative comparisons of global pause patterns between stains. To enable scaling of read counts among different samples and strains, we added a spike-in of 3×FLAG-tagged *rpoC Bacillus subtilis* cells to samples at a constant cell ratio prior to cell homogenization (3.4% spike-in for WT samples and 1.4% spike-in for ΔSI3* samples on a per cell basis; see Materials and Methods). Uniformly trimmed reads (15 nt from 3′ ends) were mapped to the *E. coli* and *B. subtilis* genomes. After normalization, the 3′-end reads for a given template position were proportional to the number of RNAPs located at the site (**Figure 3.4A**, right part). We observed much greater variation between WT and ΔSI3* signals than between biological replicates both for normalized 3′ end reads (**Figures 3.4B**). These results indicate that ΔSI3* globally altered RNAP distribution among genomic positions. These ΔSI3* effects were immediately evident at specific loci where ΔSI3* either enhanced or decreased pausing compared to WT at different positions (**Figure 3.4C**). Interestingly, these differences were not seen in rRNA genes (**Figure 3.4B**), where pausing is known to be suppressed by antitermination factors (32). These results confirm that the ΔSI3* RNAP alters transcriptional pausing globally *in vivo*.

The NET-seq results reveal numerous increases and decreases in pausing caused by ΔSI3* genome-wide. To catalog the affected pause sites statistically, we used a sliding window algorithm to assign pause sites as locations with a normalized 3′-end count more than 4 standard deviations greater than the mean of all the non-zero reads within a +99/–100-bp window ($z > 4$; **Figure 3.4E**). To ensure strong pauses did not mask detection of weak pauses, four rounds of pause calling were performed with pauses identified in each round removed for subsequent rounds. Biological replicates were scored separately and only sites significant in all replicates were retained. This pause-

calling method yielded ~43,000 pauses in non-ribosomal RNA operons for WT *E. coli* and ~53,000 for the ΔSI3* strain, comparable to published pause densities (4). We then applied a cutoff of ≥40 average, normalized counts, yielding 10,700 WT pauses in dataset and 14,230 ΔSI3* pauses (**Figure 3.4E**). Only 5259 sites were found in both the ΔSI3* and WT datasets; ΔSI3* reduced some WT pauses to z <4 and many ΔSI3* pauses were not significant for WT.

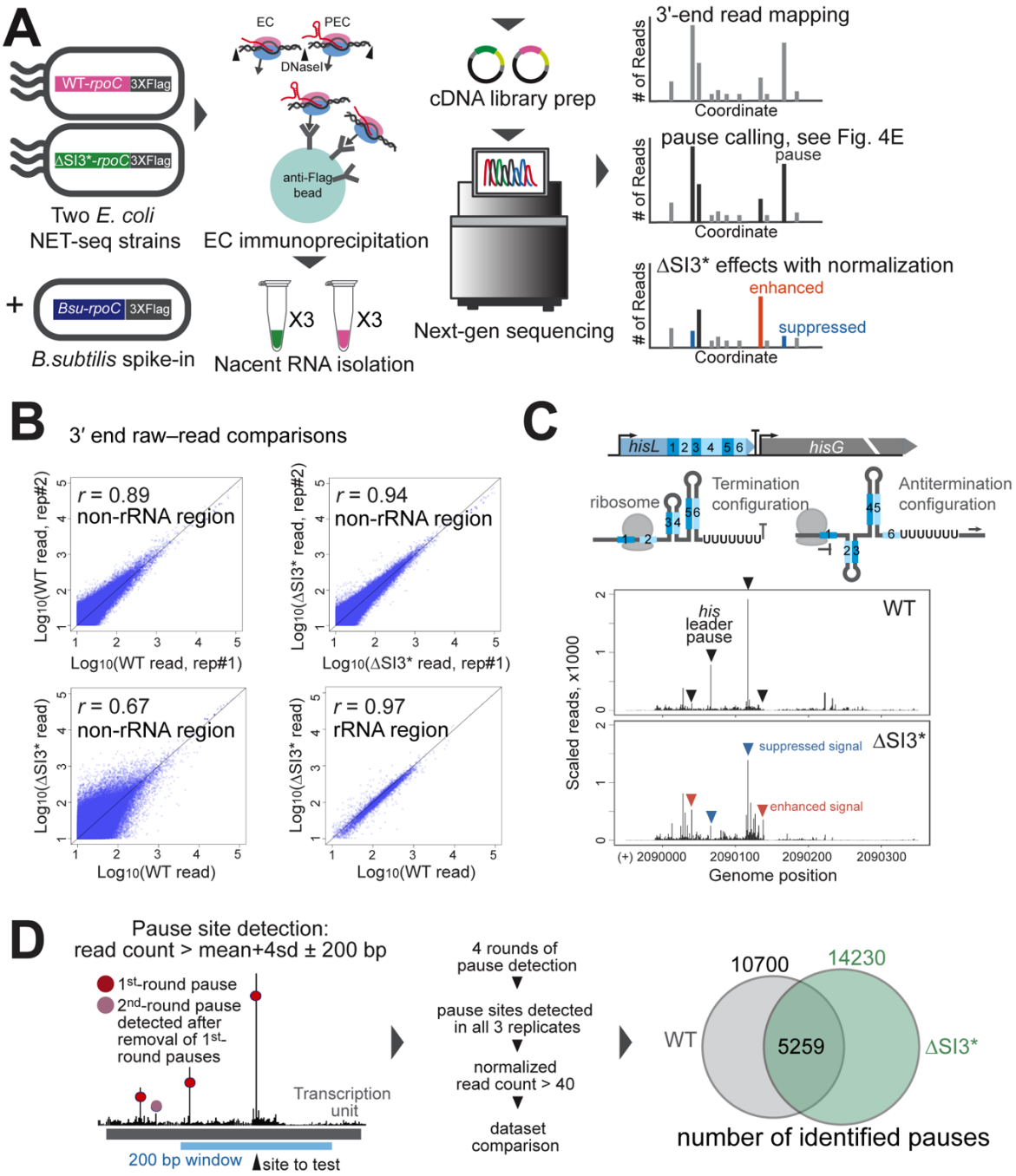


Figure 3.4. ΔSI3*-induced global changes in pause profiles probed by NET-seq.

(A) Pipeline of quantitative NET-seq library preparation and data analysis. To allow quantitative comparison of pausing between WT and ΔSI3* strains, a *B. subtilis* strain was used as spike-in (proportional to *E. coli* harvest OD,

6.2×10^8 CFU per $OD_{E.coli}$) control throughout library preparation. (B) The consistency of NET-seq reads in non-rRNA region and rRNA region between datasets represented by Pearson's coefficient (*r*). (C) Comparison of NET-seq reads between ΔSI3* and WT datasets in the *his* operon leader region. Numbers 1 to 6 represents alternatively pairing RNA segments in the attenuation control region (1:2–3:4–5:6 forming the termination configuration and 1–2:3–4:5–6 forming the antitermination configuration). (D) Schematic diagram of pause site detection and numbers of pause signals in WT and ΔSI3* datasets. Overlapping region represents common pause sites between the two datasets.

3.3.4 SI3 exerts opposite effects on two types of pauses *in vivo*

To assess quantitatively how ΔSI3* RNAP alters pause strength, we calculated a pause index for each of the 10,700 WT and 14,230 ΔSI3* pause sites. The pause index is defined as the site's normalized 3′-end counts divided by the number of all 15-nt normalized reads that include the position (**Figure 3.5A**) (33). This approach corrects for differences in NET-seq signals that arise from differences in gene expression levels. Although these pause index values are not a direct measure of pause strength, changes in the values provide a useful measure of changes in pause strengths caused by RNAP alteration in different strains. Similar to normalized NET-seq reads (**Figure 3.5B**), pause indices were much more divergent between ΔSI3* and WT strains than between replicates (**Figure 3.5B**) and this divergence was not evident in ribosomal RNA operons in which pausing is suppressed by antitermination factors (**Figure 3.5C**).

To categorize pauses significantly changed by ΔSI3*, we used a two-sample Z proportion test to assess whether ΔSI3* pause strength differed significantly from WT ($p < 0.0001$). Using the fold change threshold of ΔSI3*/WT ≤ 0.8 or ≥1.25, we found 4491 pauses significantly affected by ΔSI3* (**Figure 3.5A**). We used ΔSI3*/WT ≤ 0.8 as a threshold for suppressed pausing because 0.8 was the smallest change of pause strength among several known hs-pause sites in our dataset (**Figures 3.5E** and **S3.3A, B**). We used ΔSI3*/WT ≥1.25 as a symmetric threshold for increased pausing to detect pause sites enhanced by ΔSI3* (e.g., resembling the ce-pause signal enhanced 1.4-fold by ΔSI3* *in vitro*; see **Figure 3.3C**). Overall ΔSI3* changed ~42% of WT pauses enough to merit further investigation, with 3134 decreased and 1357 increased. We also examined the pause indices for ΔSI3* sites. Consistent with the analysis of WT sites, most ΔSI3* pauses were suppressed rather than enhanced in the WT strain (Figure 5F, G). These results confirm that SI3 modulates pausing globally both positively and negatively.

Since ΔSI3* enhanced the pause strength of the ce-pause *in vitro* (**Figure 3.3C**), we wondered if the ΔSI3*-enhanced pause sites exhibited ce-pause sequence features. Strikingly, the sequence logo generated from 1,357 ΔSI3-enhanced pause sites strongly resembled the ce-pause logo ($G_{-11}G_{-10}t_{-3}g_{-2}Y_{-1}G_{+1}$) (4) with more enrichment for –10G, –3T, –1Y, and +1G, less enrichment for–11 G and –2 G (**Figure 3.6A**, top). In addition, the pauses specific to the ΔSI3* strain (**Figure 3.6A**, the second panel from top) generated a comparable sequence logo to that of the ΔSI3-enhanced pauses (**Figure 3.6A**, top). We hypothesize that SI3 reduces the strength of strong elemental

pauses at which TL folding stabilizes a pretranslocated paused state (16) and that these pauses are favored by sequences with optimal matches to the ce-pause sequence. Hence, SI3 deletion increases pausing at these sites.

Since ΔSI3* RNAP reduces hs-pausing *in vitro*, we next asked if ΔSI3* RNAP reduced pausing *in vivo* at known hs-pause sites in amino-acid biosynthetic regions (*thrL1, hisL, pheL, ilvL, leuL, pheM* and *ivbL*) (34, 35) (**Figure S3.3A**). ΔSI3* significantly reduced pause strength at these sites (ΔSI3* pause strength 45%–80% of WT, **Figure S3.3B**). For comparison, we examined two non-hs-pause sites, the $pyrL_{54C}$ and $rfaQ_{56U}$; $pyrL_{54C}$ and $rfaQ_{56U}$ exhibited similar pause strengths in the ΔSI3* and WT strains (Figure S3B. These results suggest that deletion of SI3 also reduces hs-pause strength *in vivo*.

To ask if other ΔSI3*-suppressed pause sites are associated with PHs, we examined RNA secondary–structure predictions (36) for nascent RNA 40 nt upstream from pause sites (–50 to –11, where –1 is the pause RNA 3′ end) (**Figure 3.6B**). We screened for stem–loop RNA structures that paired to positions –11 to –14 upstream from the RNA 3′ end (–11 to –14 PHs). RNA hairpins at the –12 and –13 positions give the strongest PH effects, but –11 and –14 locations also may increase pausing at least in the presence of NusA (37, 38). RNA molecules easily form compact structures due to wobble base-pairing and base-stacking (39, 40); thus, we first determined the background level of structures predicted using our criteria by sampling 1500 random RNA sequences from each genomic strand. Over a third of random forward and reverse strand sequences yielded a predicted RNA structure (38 ± 1% and 37 ± 3%, respectively; **Figure 3.6B**). A higher fraction of the ΔSI3*-suppressed pause signals (n = 3134) yielded a putative PH (42% and 42% from two genomic strands, p = 0.021 and 0.00063, single-tailed Z proportion test). In contrast, ΔSI3*-enhanced pause signals (n = 1357) yielded only background levels of PH predictions (39% and 36%, p = 0.34 and 0.57, single-tailed Z proportion test). Although predicted –12 PHs accounted for the majority, their portion was not statistically distinct among 3 pools tested. We designated the 1321 ΔSI3-suppressed signals with possible PHs putative hs-pause signals.

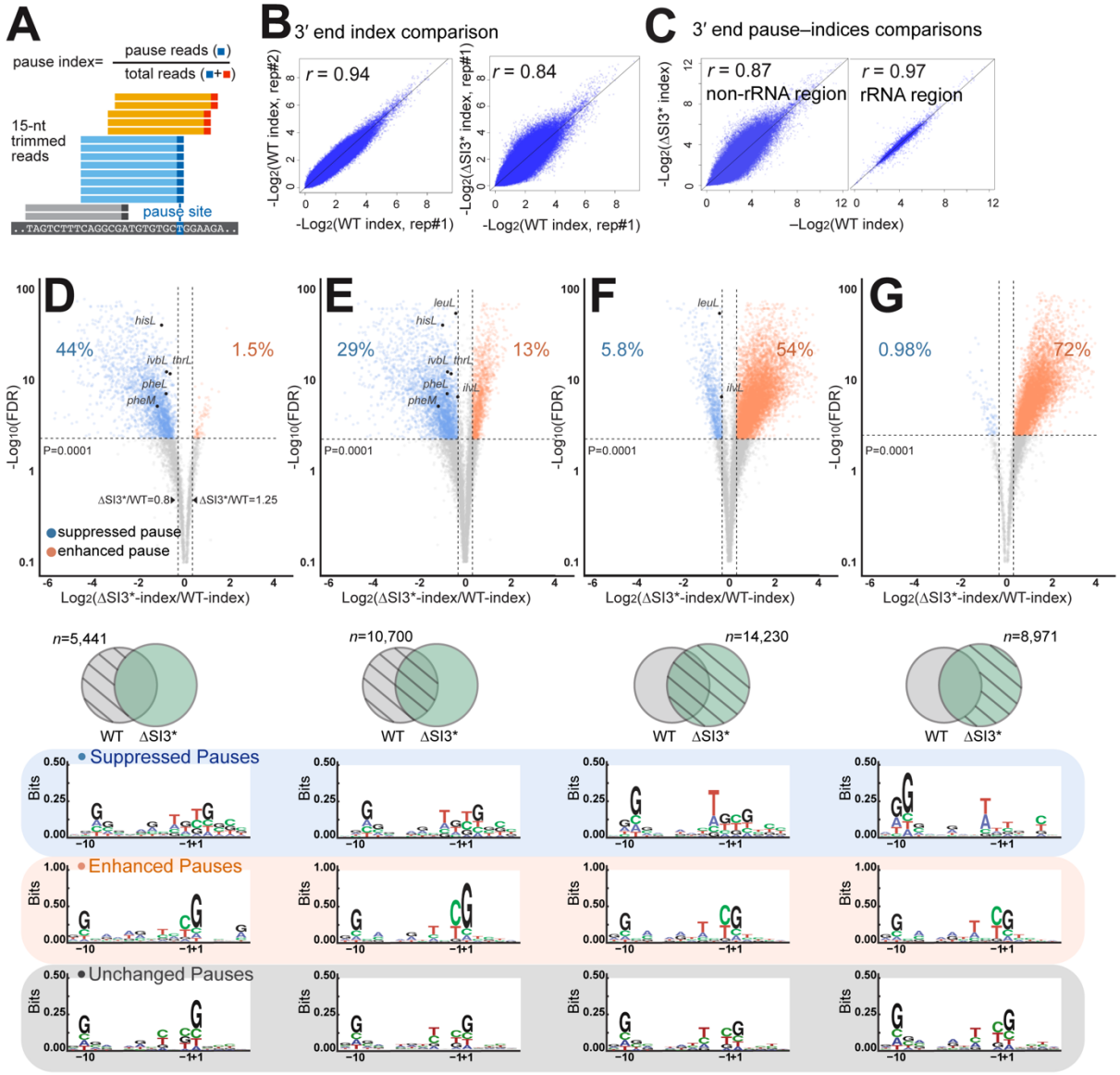


Figure 3.5. Compasion of pause strengths between WT and ΔSI3* datasets using pause indices.

(A) Quantification of the pause index. Dark colored squares represent 3′ ends of the mapped reads. (B) 3′-end pause index comparisons between WT biological replicates and between WT and ΔSI3* datasets. (C) 3′-end pause index comparisons between WT and ΔSI3* datasets in non-rRNA regions and in the rRNA region. (D–G) Top panel, volcano plots of pause index comparisons between WT and ΔSI3* datasets at WT-specific pause sites, all WT pause sites, all ΔSI3* pause sites and ΔSI3*-specific sites, respectively. Blue dots represent pause sites with statistically

significant decreased pause indices in ΔSI3* vs. WT, and orange dots represent pause sites with statistically significant increased pause indices in ΔSI3* vs. WT. Index change thresholds are marked with vertical dotted lines and the p value threshold (0.0001) is marked with the horizontal dotted line. Middle panel, pause sites for the upper volcano plots are marked with slashed regions. Bottom panel, sequence logos of the corresponding regions in the volcano plots.

3.3.5 *In vitro* validation of putative hs-pause signals

To ask whether the putative hs-pause signals indeed were stimulated by PHs, we selected three candidates to compare *in vitro* to the well-charactered *hisL* hs-pause (in *malT*, *fabH* and *yfaS* genes; **Figure 3.6C**, **Figure S3.4A**). Short (~60-bp) fragments of the native sequences (–40 to +10 relative to the pause sites) were inserted into a λP_R promoter-driven DNA template (**Figures 3.6C**). *In vitro* transcription assays revealed that the pause strength for each case was significantly decreased in using ΔSI3* RNAP relative to WT RNAP (**Figure 3.6D**). Using RNA sequencing ladders, we confirmed that each *in vitro* pause corresponded to the expected position based on NET-seq data (**Figure S3.4B**). Further, addition of NusA, which is known to enhance hs-pause by promoting swiveling and stabilizing pause hairpins (17, 41), increased pause strength for each candidate site (**Figure 3.6E**). Finally, the contributions of putative the PHs to pause stabilization were confirmed by reductions in pause strengths upon the addition of 18-nt antisense DNA oligos (asDNAs) that specifically paired to the upstream PH stems (**Figure 3.6F**). These results collectively validate our assignment of putative hs-pause signals based on ΔSI3* suppression of pause index and PH prediction as criteria.

We conclude that hs-pauses are infrequent in *E. coli* (~12% or 1,321 of the 10,700 pauses assigned by our analysis for WT RNAP appear to be hs-pauses). Additionally, we our validation of hs-pauses establishes unambiguously that SI3 modulates different classes of pauses oppositely. Pauses resembling the consensus elemental pause are inhibited by ΔSI3, presumably because SI3 destabilizes the folded-TL pretranslocated ePEC state whereas it inhibited TL folding in the hairpin-stabilized, swiveled PEC (16).

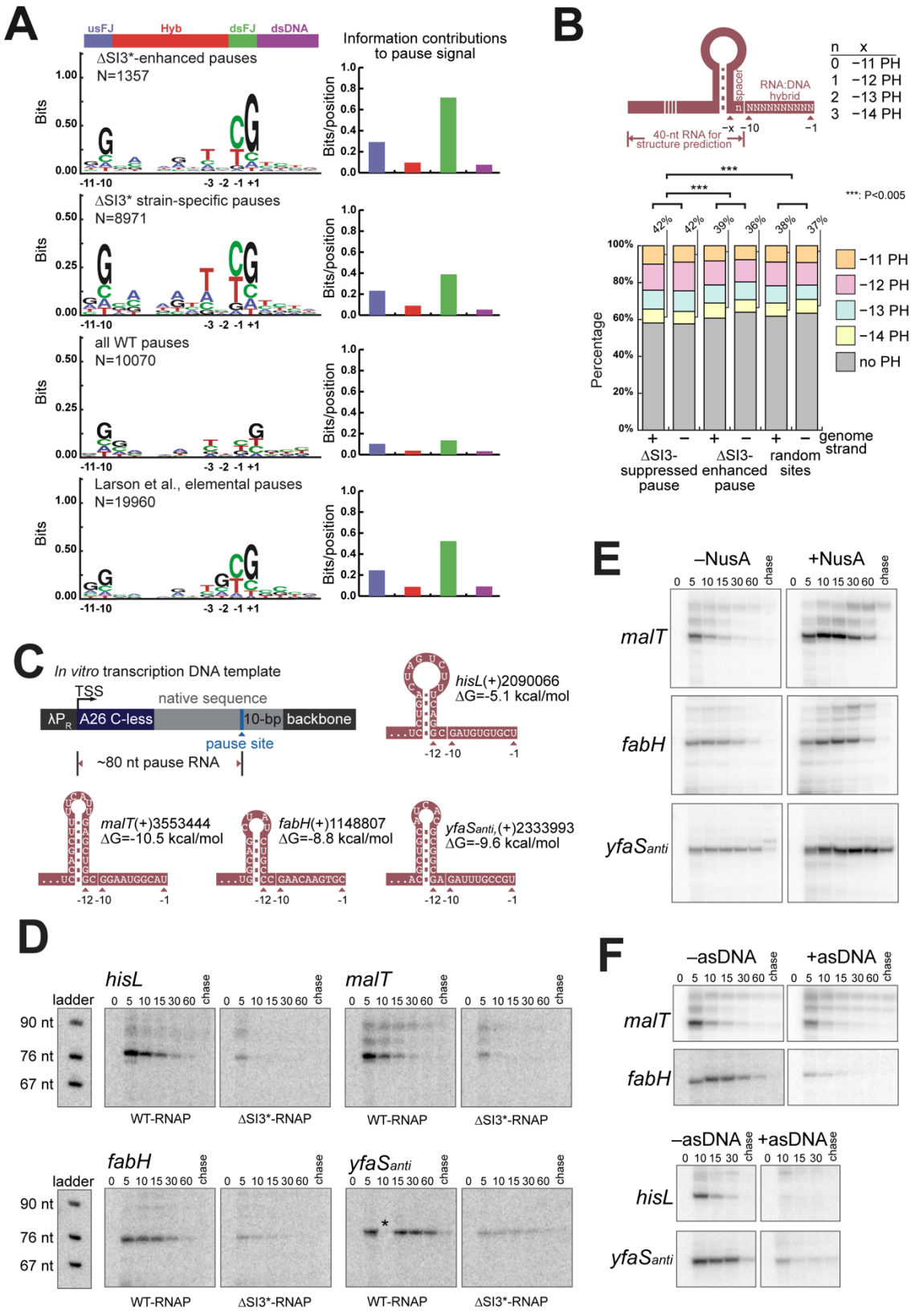

A
usFJ
Hyb
dsFJ
dsDNA
ΔSI3*-enhanced pauses
N=1357
ΔSI3* strain-specific pauses
N=8971
all WT pauses
N=10070
Larson et al., elemental pauses
N=19960
Bits
Information contributions to pause signal
Bits/position
-11 -10
-3 -2 -1 +1
B
spacer
RNA:DNA hybrid
NNNNNNNNNN
40-nt RNA for structure prediction
n x
0 −11 PH
1 −12 PH
2 −13 PH
3 −14 PH
***: P<0.005
42%
42%
39%
36%
38%
37%
Percentage
−11 PH
−12 PH
−13 PH
−14 PH
no PH
genome strand
ΔSI3-suppressed pause
ΔSI3-enhanced pause
random sites
C
In vitro transcription DNA template
TSS
native sequence
λP_R
A26 C-less
10-bp
backbone
pause site
~80 nt pause RNA
hisL(+)2090066
ΔG=-5.1 kcal/mol
malT(+)3553444
ΔG=-10.5 kcal/mol
fabH(+)1148807
ΔG=-8.8 kcal/mol
yfaSanti,(+)2333993
ΔG=-9.6 kcal/mol
D
hisL
malT
fabH
yfaSanti
ladder
90 nt
76 nt
67 nt
0 5 10 15 30 60 chase
WT-RNAP
ΔSI3*-RNAP
E
–NusA
+NusA
malT
fabH
yfaSanti
F
–asDNA
+asDNA
malT
fabH
hisL
yfaSanti

Figure 3.6. SI3 exerts opposite effects on two types of pauses *in vivo*.

(A) The sequence logo of the ΔSI3*-enhanced pauses, the ΔSI3*-specific pauses, the WT dataset pauses and the consensus elemental pause sequence from Larson et al. The four sequence elements contributing to elemental pause strength are labeled in colors and the contribution of these elements in the four logos is quantified and shown in bar graphs on the right. (B) Ratios of predicted PHs in ΔSI3*-suppressed pauses, ΔSI3*-enhanced pauses and random genomic sites. The nomenclature of PHs at various positions is defined above. (C) The design of the *in vitro* transcription template with hs-pause signals and the predicted PHs of the putative hs-pause candidates. (D) *In vitro* transcription assay of the putative hs-pause candidates with WT and ΔSI3* RNAPs. The asterisk (*) in $yfaS_{anti}$ gel marks a missing sample. (E) *In vitro* transcription assay of the putative hs-pause candidates with NusA. (F) *In vitro* transcription assay of the putative hs-pause candidates with PH-disrupting asDNA.

3.3.6 ΔSI3-enhanced pause and putative hs-pause signals are enriched in early ORF regions

We next investigated the distribution of ΔSI3*-enhanced and putative hs-pause signals across the genome. Overall, both types of pauses exhibited genomic distributions that are similar to each other and to pauses generally (4). The majority (63–69%) were located in sense transcripts of protein-coding sequences of genes [including the 5′ leader peptide coding regions of transcription units (TUs)], followed by the 5′ untranslated regions of protein-coding TUs (5′ UTRs, 10–12%), (**Figure 3.7A**). We identified a small number of putative hs-pause sites in 3′ UTRs, which could be associated with intrinsic terminators. Some of each pause class also occurred in antisense TUs. When normalized for the number of bp in these regions, 5′ UTRs and 3′ UTRs were enriched in both types of pauses relative protein-coding regions (Pearson's Chi-squared test, $p < 0.005$) (**Figure 3.7B**). This enrichment is consistent with the roles of pausing and tunable RNA structures in regulating transcription elongation/termination and translation in these two regions **(4, 42-44)**.

We also examined how SI3*-modulated pause sites were distributed within ORFs. ΔSI3*-enhanced pauses and putative hs-pauses were found at significantly higher levels in the early ORF region (first 250 bp), accounting for 56% and 60% of the totals (single-tailed Z proportion test, $p = 0.003$ and 10^{-5}, respectively). Both ΔSI3*-enhanced pauses and putative hs-pauses were located closer to the start codons than all pauses in aggregate (two sample Kolmogorov-Smirnov test, single-tailed, $p = 0.001$ and 4×10^{-5}) (**Figure 3.7C**). A consensus elemental pause sequence is known to be enriched coinciding with translation start sites ($AU_{-1}G_{+1}$ pauses) (4). These 'AUG' pauses were not enriched in the ΔSI3*-enhanced and putative hs-pauses relative to their frequency in all pauses (19 of 709 ΔSI3*-enhanced pauses and 4 of 844 putative hs-pauses were AUG pauses). Consistent with prior analysis of pauses in protein-coding regions (4), we conclude that ΔSI3*-modulated pauses are more prominent before ribosomes load on nascent mRNA and may aid in establishing transcription–translation coupling.

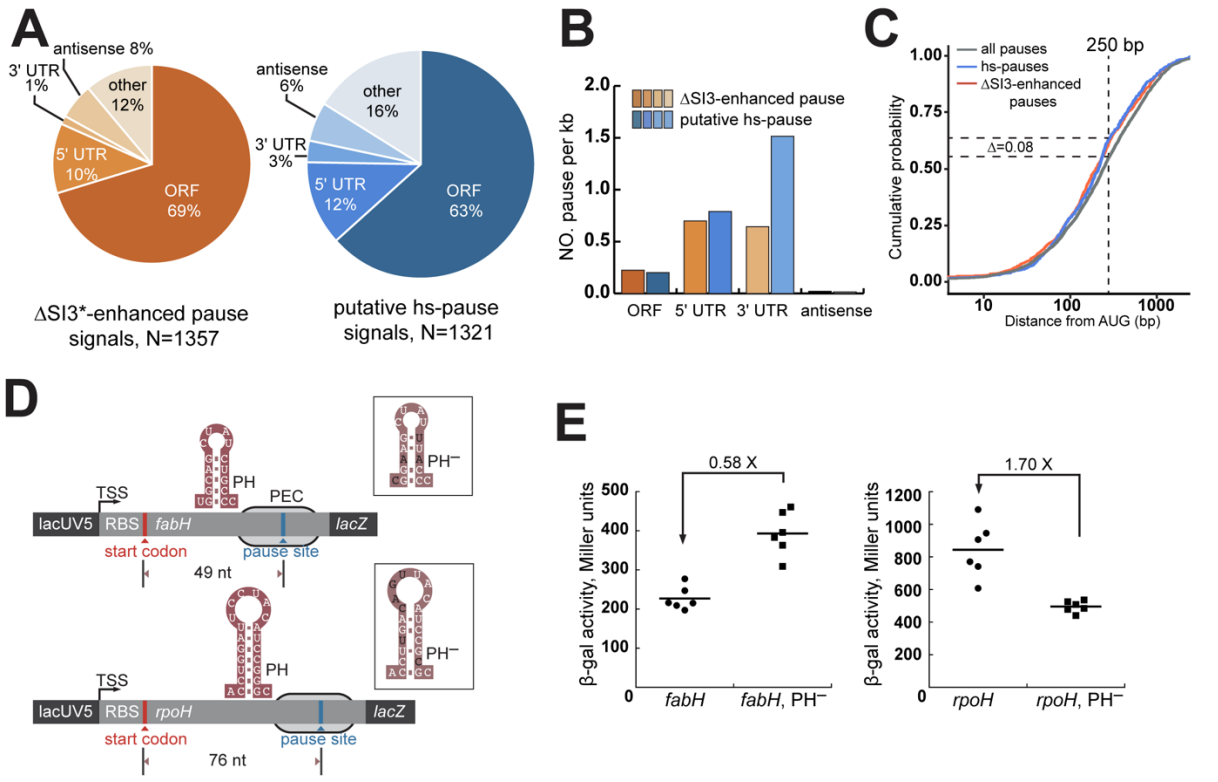


Figure 3.7. Genome distributions of ΔSI3*-enhanced pauses and putative hs-pauses, and the *in vivo* test of two putative hs-pause signals.

(A) Distribution of ΔSI3*-enhanced pauses and putative hs-pauses in different genome locations. (B) Densities of pauses in various genome regions. (C) Cumulative probability of the length distribution of the total pauses from WT dataset (grey), ΔSI3*-enhanced pauses (orange) and putative hs-pauses (blue). The approximate cumulated probability difference between all pauses and putative hs-pauses is marked by horizontal dotted lines. (D) Designs of the *lacZ* reporters fused with two N-terminal peptide sequences containing the *fabH* and *rpoH* hs-pause signals and their PH disrupted mutants (PH⁻). In PH⁻ constructs, the altered bases are colored in black. (E) Expression comparisons of the *lacZ* reporters between hs-pause candidates and their mutants. Plot represents 6 biological replicates and their mean values.

3.3.7 Hs-pause signals in early ORFs divergently affect gene expression through PH action

To ask if hs pauses detected early in ORFs indeed depending on a PH for function *in vivo*, we tested two such pauses in detail using translational fusions of the ORFs to a *lacZ* reporter preceded by the native ribosome binding site (RBS) from the candidate ORF and the lacUV5 promoter. These hs pauses were located at +59 nt of *fabH* and +76 nt of *rpoH*, respectively (**Figures 3.7D** and **S3.5A**). For each, we generated PH^{-} variants by introduced base substitutions in the PH stem regions without changing the encoded protein sequences or forming alternative PHs (**Figures 3.7D** and **S3.5B**). RNA secondary structure predictions verified that the PHs were disfavored but that the RBSs remained single-stranded in the mutant sequences (**Figure S3.5C**). *In vitro* transcription assays verified that both hs-pauses were significantly weakened in the PH^{-} variants (**Figure S3.5D**).

Strikingly, *fabH* and *rpoH* hs-pauses had opposite effects on gene expression. The mutants decreased *lacZ* expression ~42% for the *fabH* pause but increased it ~73% for the *rpoH* pause (**Figure 3.8E**). These divergent effects suggest the ORF-internal hs-pauses may play diverse roles in regulation but that these pause functions indeed depend of PH action. Inhibition of ρ with bicyclomycin at levels with minimal effects on cell growth (20 μg/mL) (45) reduced the effect of the *rpoH* PH (to ~24%) but did not alter the *fabH*-PH effect (**Figure S3.5E**). We speculate that the *rpoH* hs-pause may aid ribosome recruitment and transcription–translation coupling and thus decrease ρ-dependent termination (46) whereas the opposite effect of the *fabH* PH will require further study to understand.

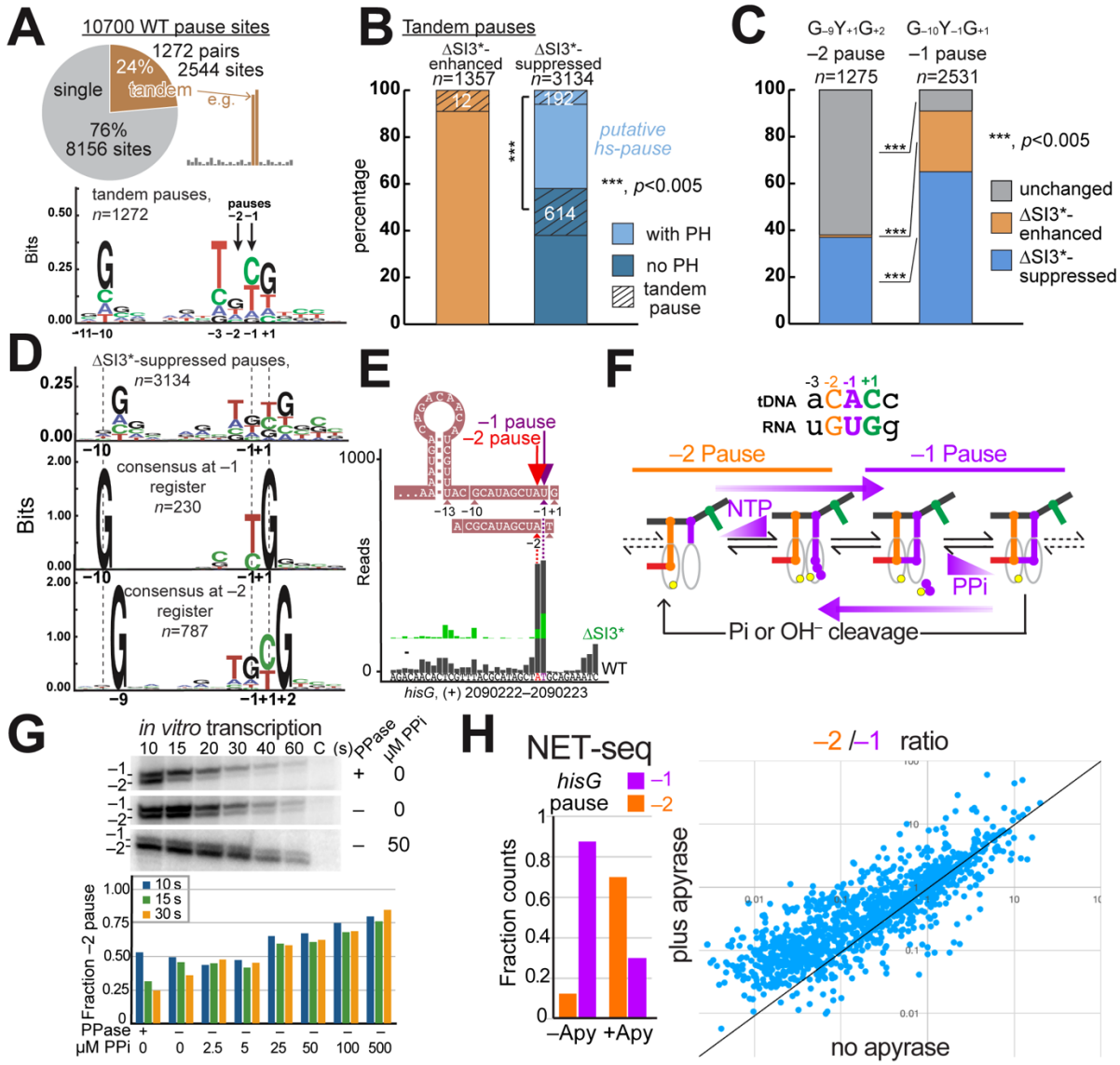


Figure 3.8. Hs-pause signals in early ORF alters gene expression.

Tandem pauses in NET-seq. (A) Top, ratios of tandem pauses in all WT pauses. Bottom, the sequence logo of all –1 pauses in tandem pause pairs. (B) Ratios of tandem pauses in ΔSI3*-enhanced pauses and ΔSI3*-suppressed pauses. P value is from Z-proportion test. (C) Numbers of all pauses halted at –1 and –2 positions relative to the consensus elemental sequence (pause at –1) and their pause index changes in ΔSI3* dataset compared to WT. P value is from Z-proportion test. (D) Sequence logos of ΔSI3*-suppressed pauses, –1 and –2 pauses. (E) The tandem pauses in

hisG. (F) A putative mechanism of transformation between –1 and –2 pauses. (G) *In vitro* transcription of *hisG* tandem pauses at various concentrations of PPi. Bottom bar graph shows the relative amount of –1 and –2 pauses sampling at 3 time points upon transcription restart at various PPi concentrations. (H) The changes of tandem pause registers in NET-seq without/with apyrase treatment. Bar graph on the left shows the 3′ end number changes of *hisG* tandem pauses in the two NET-seq datasets. Scattering plot on the right shows the relative ratio changes of tandem pause registers between the two NET-seq datasets. The pause sites here are from WT dataset defined previously as described in **Figure 3.8A**.

3.3.8 Some pause sites, including some hs-pauses, generate tandem –2 and –1 RNA 3′ ends

While examining the ΔSI3*-modulated pauses in detail, we observed an unexpected correlation in which many enriched RNA 3′ ends occurred at directly adjacent positions (*i.e.*, tandem pauses; **Figure 3.8A**). These apparent tandem pauses occurred more frequently among the class of ΔSI3*-suppressed pauses including hs-pauses. They occurred with highest probability among ΔSI3*-suppressed pauses lacking PHs (**Figure 3.8B**). Further analysis revealed that some NET-seq pauses were scored with 3′ ends at the –2 register (i.e., pause before pyrimidine) relative to the consensus sequence without the accompanying classical –1 pause (i.e., pause after pyrimidine). Considering both tandem pauses and the isolated –2 pauses, around 40% of the –2 (shifted-register) pauses were nearly exclusively ΔSI3*-suppressed whereas 65% of the –1 pauses were suppressed and 26% of the –1 pauses were enhanced by ΔSI3* (**Figures 3.8C**). The existence of the apparent –2 pauses could partially explain why the sequence logo for ΔSI3*-suppressed pauses was more degenerate than observed previously for pauses in general and for ΔSI3*-enhanced pauses (**Figure 3.8D**).

We also noticed that the frequency of tandem pauses and –2 pauses appeared to vary in different samples of the same strain and growth conditions prepared for NET-seq analysis. Of note, the DNA and RNA contacts of –2 pauses in the post-translocated register are the same as the contacts of –1 pauses in the pre-translocated register (**Figure 3.8F**). These observations led us to hypothesize that the apparent –2 pauses might arise by conversion of –1 pauses to –2 pauses during NET-seq sample processing. Current NET-seq protocols include an incubation in the presence of Mg^{2+} to digest DNA. Thus, the RNAP active center could be active during this step for nucleotide addition, pyrophosphorolysis, and intrinsic transcript cleave by OH^- or phosphate (47).

To test this hypothesis, we first examined a tandem pause from the early *hisG* gene using *in vitro* transcription (**Figure 3.8E, G**). Like the *fabH* and *rpoH* hs-pauses (Figure S5D), the *hisG* pause is hairpin-stabilized (–13 PH for –1 or –12 PH for –2, **Figure 3.8E**). *In vitro*, apparent pausing at both –2 and –1 was readily apparent (**Figure 3.8G**). However, the –2 pause band was suppressed by addition of pyrophosphatase and was enhanced by even low concentrations of added pyrophosphate. These results establish that the canonical *hisG* hs-pause is highly susceptible to reversal of nucleotide addition (pyrophosphorolysis) and that the apparent –2 pauses might result from

pyrophosphorolysis of –1 pauses. They do not, however, fully exclude the possibility that weak pausing might occur directly in the –2 register.

To ask if reversal of –1 to –2 pause RNA 3′ ends might occur during NET-seq sample processing, we adopted a different approach by adding apyrase to destroy NTPs in the samples prior to exposure to Mg^{2+}. This approach had the advantage that it also tests whether nucleotide addition occurred during sample processing. If Mg^{2+} enables interconversion of –2 and –1 PECs during NET-seq sample processing by addition of UMP to –2 PEC, pyrophosphorolysis of –1 PEC or 1-nt intrinsic cleavage of –1 PEC by OH^- or Pi, then destruction of NTPs by apyrase is predicted to increase the observed –2/–1 RNA 3′-end ratio at apparent tandem pause sites. Strikingly, apyrase both shifted the –2/–1 ratio of 3′ ends at the *hisG* pause and shifted the –2/–1 ratio toward –2 for tandem pauses generally (**Figure 3.8H**). This result suggests that both nucleotide addition to convert –2 to –1 pause registers and likely pyrophosphorolysis or 1-nt intrinsic cleavage by OH or Pi can occur during NET-seq sample preparation (**Figure 3.8F**).

3.4 Discussion

Our results lead to several important insights into the function of SI3 as a modulator of RNAP and pausing as well as into the accuracy of the NET-seq method. By deleting all of SI3 enabled by a suppressor substitution, we established that *E. coli* can grow relatively well without this important domain. However, ΔSI3* RNAP not only responds less effectively to RNA PHs it also also pauses more strongly at consensus elemental pause sequences. By mapping hs-pauses based on the effect of ΔSI3* and RNA structure predictions, we found that ~12% of pauses sites in E. coli are associated with PHs but that the distribution of hs-pauses is not dramatically different than pauses in general. Finally, the effects of ΔSI3* uncovered a class of apparent tandem pause sites that document difficulties in accurately capturing pause RNA 3′ ends by NET-seq and thus explain some discrepancies in prior descriptions of consensus pause sequences.

3.4.1 SI3 and TL residues shape TL dynamics

Although small deletions within *E. coli* SI3 are known to be viable and to alter pausing (27), a complete deletion of SI3 was thought to be inviable (23). β′A941T near the point of SI3 insertion in the TL enabled construction of a viable SI3 deletion and increased the rate of nucleotide addition by ΔSI3-RNAP (**Figures 3.2** and **3.3A**). However, β′A941T alone was inviable and decreased RNAP elongation rate (**Figure 3.3A**). These opposite effects of β′A941T in the presence and absence of SI3 highlight the complex dynamics of TL folding and unfolding that is easily perturbed by small changes (48-50). The added bulk and H-bonding capacity of Thr (51, 52) may enable interactions with SI3 itself that slow TL folding but interactions with non-SI3 residues that aid TL folding when SI3 is absent. Alternatively, β′A941T may shift the folded–unfolded TL balance in the same direction in both cases, but this effect could be compensatory for ΔSI3 and over-shift the balance when SI3 is present.

Examination of TL sequences from bacteria with and without SI3 does not provide obvious insight into how β′A941T enables a viable SI3 deletion (**Figure S3.1**). The helix2a region, where SI3 is inserted, is less conserved between bacteria with or without native SI3, likely due to adaptive evolution after SI3 insertion event. However the Ala-to-Thr change itself is not evident in TLs lacking SI3 and an Ala-to-Thr substitution in yeast RNAPII, which contains Ala at the corresponding residue and lacks SI3, is phenotypically silent (50). TLs in archaeal RNAPs and

eukaryotic RNAPII contain one less amino acid in TH2a but the region corresponding to β′A941 and two adjacent residues is not generally conserved other than it typically containing smaller amino acids (e.g., G, A, S, V). It may be relevant that bacterial RNAPs lacking SI3 often contain Lys adjacent to the A941 position. Like Lys, Thr at A941 would increase the potential for side chain interactions that could be explain its ΔSI3-suppressor phenotype. SI3 interacts with both peripheral domains and dissociable factors (20, 22, 53). Thus, increasing the potential for interactions when SI3 is absent might be compensatory. A more systematic study of diverse substitutions in this region with and without SI3 would be required to explain the β′A941T phenotype more completely.

3.4.2 SI3 influences transcriptional pausing positively and negatively via multiple intermediates

An emerging view of pausing emphasizes the roles of multiple interconverting paused states that range from pretranslocated PECs with a folded TL to swiveled PECs with a half-translocated RNA:DNA hybrid (**Figure 3.9**) (15). Different pause sequences may preferentially stabilize one or more state among of this ensemble of paused states. Collectively the ensemble stabilizes the PEC against formation of a post-translocated state competent for NTP binding and continuation of transcription.

A major role of SI3 in pausing is stabilizing the unfolded TL when SI3 swivels as part of the RNAP swivel module (consisting mainly of the clamp, dock, jaw, shelf, and SI3) (1, 15-17). Additionally, PECs can be stabilized in the pretranslocated register on sequences close to the consensus elemental pause (16). Our NET-seq results and *in vitro* validation substantiate these roles of SI3 in oppositely affecting different paused intermediates. SI3 promotes pausing at sites for which swiveling plays a dominant role, including hs-pauses, but inhibits pausing at sites where the TL-folded, pretranslocated PEC plays a dominant role (**Figure 3.9**).

Because SI3 mediates PH inhibition of TL folding in the swiveled PEC , we could identify hs-pauses based on the effect of ΔSI3* combined with RNA structure predictions. This analysis suggested that *E. coli* encodes a relatively large number (~1300) of hs-pauses with different sequence contexts and PH structures genome-wide (**Figures 3.6** and **3.7**). We observed a slight bias in locations of hs-pauses (and strong elemental pauses inhibited by SI3) toward earlier regions of protein coding genes, but overall hs-pauses occur throughout transcribed regions. We did not include a free energy cut-off for secondary RNA secondary prediction in identifying hs-pauses because NusA may chaperone formation of weak hairpins to promote pausing. Additionally, ternary interactions not captured

in our prediction may stabilize PHs (54). Further analysis of the ΔSI3-suppressed pause sites using improved algorithms for RNA structure prediction that can detect complex RNA structures involving more nascent RNA may reveal additional hs-pause sites that tune transcription via novel mechanisms (55, 56).

It also is notable that a significant fraction of pause sites are stabilized in ΔSI3*, presumably because the TL-folded, pretranslocated state is a significant pause intermediate at these sites. Although the consensus elemental pause sequence is known to exhibit such stabilization (16), few if any *E. coli* pause sites exactly match the consensus sequence. Our results suggest the TL-folded, pretranslocated state, which is stabilized by SI3 deletion, is a kinetically significant intermediate for a large number of pauses even when they match only the most highly enriched positions of the consensus pause sequence (e.g., –10, –1, and +1).

Finally, we note that our NET-seq analysis also verifies the predicted inhibition of pausing during ribosomal RNA gene transcription. During the exponential phase, up to two-thirds of active RNA polymerases (RNAPs) are found in the rRNA region, which highlights the significance of rRNA transcription in bacterial growth **(57, 58)**. To enable uninterrupted and rapid rRNA transcription, a transcription antitermination complex (TAC) is formed, which helps prevent transcriptional pausing and premature termination **(32)**. Our NET-seq analysis of wild-type and pause-altering RNAPs provides direct evidence supporting the effectiveness of TAC in rRNA transcription (**Figure 3.4B**, **Figure 3.5C**).

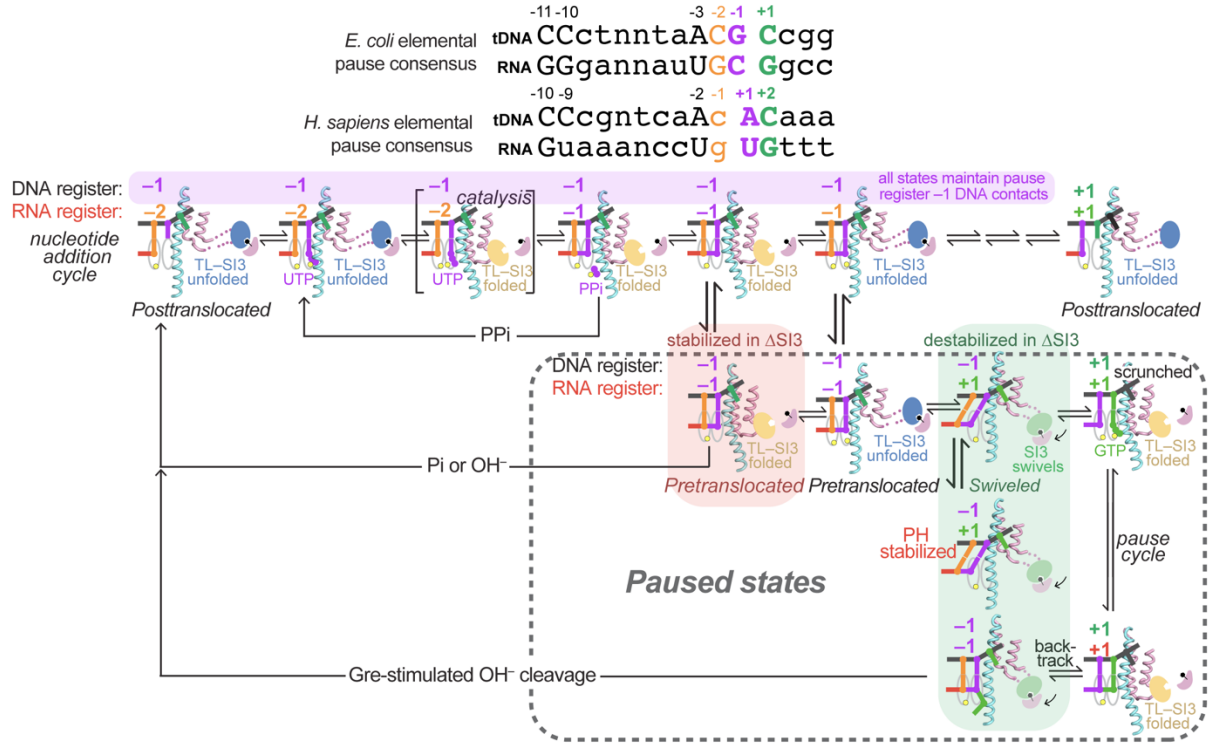


Figure 3.9. SI3 in interconverting PECs and transcription cleavage.

Top, the consensus elemental pause sequence logos from *E. coli* and *H. sapiens*. These two logos are similar in sequence elements but the positions of the paused 3′ ends (labeled by –1) differ by 1 nt. Our results attribute the discrepancy to the possible rNTP addition or PPi/Pi/OH^--induced cleavage during the NET-seq library preparation. Middle, interconversion of EC/PEC states from the –2 species at the post-translocated register to the –1 species at the post-translocated register. PPi acts as the factor contributing to the 3′-end cleavage of the –1 pause register. Bottom box, the effects of ΔSI3 on different types of pauses. In our results, ΔSI3 stabilized a type of pause resembling the consensus elemental pause, by stabilizing the pre-translocated register in this pause. Meanwhile, ΔSI3 destabilized hs-pause by counteracting swiveling. The globular area connected to TL is SI3 domain and its positions are indicated by different colors (blue: open position, yellow: closed position, green: swiveled position).

3.4.3 NET-seq may not report precise pause sites and elongation dynamics accurately

Diverse bacterial RNAPs and mammalian RNAPII respond similarly to the *E. coli* consensus and anti-consensus elemental pause sequences *in vitro* (4), although some variations in *in vitro* pausing do occur (59). However, the consensus pause sequences derived from different NET-seq experiments vary even for the same species and sometimes indicate pausing before instead of after the U or C typically found at pause RNA 3′ ends (6-9). Our results suggest a possible resolution to these discrepancies by documenting that pause RNA 3′ ends can shift during NET-seq sample work-up either by reaction with NTPs or possibly by pyrophosphorolysis or intrinsic transcript cleavage (**Figure 3.9**). Although we have not extensively characterized these changes, the very low K_{NTPs} for nucleotide addition at some template positions (60), the high rates of 1-nt cleavage observed for reaction with inorganic phosphate (47), and the high sensitivity of some pause RNA 3′ ends to pyrophosphorolysis reported here all suggest a need to develop rapid-quench, Mg^{2+}-free methods for NET-seq analysis. Specifically, we suggest that the observation of pause consensus sequences for mammalian and yeast RNAPII similar to the *E. coli* consensus but shifted 1 nt upstream (8) may have resulted from cleavage reactions during sample processing (**Figure 3.9**). In the same vein, the extreme discrepancy in reported consensus sequences for *B. subtilis* RNAP (4, 9) may reflect the consequences of RNA 3′-end changes that occur when NTPs, PPi, and Pi remain in reactive states during NET-seq sample preparation.

Overall, our findings both reveal the complex ways the SI3 modulates pausing and suggest a need to improve NET-seq methods to accurately capture nascent RNA 3′ ends in living cells. SI3 can increase pausing by favoring swiveling at some sites and decrease pausing at other sites by disfavoring stabilization of the pretranslocated, TL-folded state. New NET-seq methods should allow us to define the interconversion of these states in vivo with far greater accuracy and should be a key goal for mechanistic work on transcription elongation and pausing for the immediate future.

3.5 Materials and methods

Reagents

Plasmids, oligonucleotides, cell strains are listed in **Table of oligonucleotides, plasmids and strains**. Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA, USA) and RNA oligonucleotides and NET-seq cDNA library primers were purified with polyacrylamide gel electrophoresis (PAGE). [α-^{32}P]GTP, [α-^{32}P]CTP and [γ-^{32}P]ATP were from PerkinElmer Life Sciences; rNTPs were from Promega (Madison, WI, USA).

Proteins

His-tagged wild-type, ΔSI3 and ΔSI3* RNAPs were purified as described previously (20). Briefly, plasmids harboring *rpoA*, *rpoB*, *rpoC*-His10 and *rpoZ* were transformed into BL21(DE3) cells. Overnight cultures (5 mL) were inoculated into 1 L LB medium with kanamycin (50 μg/mL) and the cells were grown at 37 °C, 220 rpm until apparent OD_{600} reached 0.4. Then, 1 mM IPTG was introduced to induce protein expression for 3 hours before harvesting. The cells were resuspended in 50 mL Lysis Buffer (50 mM Tris-HCl, pH 7.9, 5% (v/v) glycerol, 200 mM NaCl, 2 mM EDTA, 10 mM β-ME, 10 mM DTT, 100 μg PMSF/mL, and 1 tablet of protease inhibitor cocktail (Roche). Cells were lysed by sonication and clarified by centrifugation at 11,000 × *g*, 4 °C for 15 min. A crude precipitation of RNAPs were obtained by adding PEI to 0.6% (v/v) to the supernatant followed by centrifugation at 11,000 × *g*, 4 °C for 15 min. Contaminants were extracted from the pellet by crushing with a tissue homogenizer and washing with 25 mL PEI Wash Buffer (10 mM Tris-HCl, pH 7.9, 5% (v/v) glycerol, 0.1 mM EDTA, 5 μM $ZnCl_2$, 500 mM NaCl, 10 mM DTT). Crude RNAPs were extracted from the pellet by crushing with a tissue homogenizer and washing with 25 mL PEI Elution Buffer (10 mM Tris-HCl, pH 7.9, 5% (v/v) glycerol, 0.1 mM EDTA, 5 μM $ZnCl_2$, 1 M NaCl, 10 mM DTT). Proteins were precipitated by adding 0.37 g $(NH_4)_2SO_4$ per milliliter of eluted solution. Precipitated RNAPs were resuspended in 35 mL His-column Buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 5 mM imidazole and 5 mM β-ME) and loaded into HisTrap FF 5 mL column. Proteins were washed with His-column Buffer and eluted with the same solution containing 30 mL gradient increasing imidazole (5-500 mM imidazole) in His-column Buffer. Samples containing RNAPs were pooled and adjusted to 200 mM final NaCl. The

protein was then loaded into HiTrap Heparin HP 5 mL column to remove contaminants including residual nucleic acids. The protein was washed with Wash Buffer (10 mM Tris-HCl, pH 7.9, 5% (v/v) glycerol, 25 mM EDTA, 5 μM $ZnCl_2$, 200 mM NaCl, 10 mM DTT) and eluted with Elution Buffer (10 mM Tris-HCl, pH 7.9, 5% (v/v) glycerol, 0.1 mM EDTA, 5 μM $ZnCl_2$, 500 mM NaCl, 10 mM DTT). Fractions containing pure RNAPs were pooled and dialyzed in RNAP Storage Buffer (10 mM Tris-HCl, pH 7.9, 25% (v/v) glycerol, 0.1 mM EDTA, 20 μM $ZnCl_2$, 1 mM $MgCl_2$, 100 mM NaCl, 10 mM DTT) and stored at –80 °C.

Native *rpoC* modification with no-SCAR

We followed the published protocols (26, 61) to delete SI3 from *E. coli* genome. A plasmid with a Tet-inducible Cas9 protein (pCas9cr4) and a plasmid (pYB301) harboring an Ara-inducible λ-Red system plus an SI3-targeting sgRNA (target sequence: TTACGCGTCAGACCGACGAA) were sequentially transformed into an *E. coli* K-12 MG1655 (62). The cells were cultured at 30 °C in SOB medium (2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) containing 30 μg/mL chloramphenicol and 50 μg/mL spectinomycin until apparent OD_{600} reached 0.4. λ-Red was induced by adding 0.2% (w/w) L-arabinose and incubating at 30 °C for 15 min. The cells were chilled quickly on ice and made competent for electroporation using reported methods (26). A 202-bp double-stranded DNA (dsDNA) recombination template of a ΔSI3-*rpoC* was PCR-amplified from an existing plasmid (pRM759) and gel-purified. The del27 dsDNA template was cloned from pRM839 and the ΔSI3* dsDNA template was cloned from pYB205. The competent cells were transformed with 1 μg of dsDNA template, plated on LB plates containing 30 μg/mL chloramphenicol, 50 μg/mL spectinomycin and 200 ng/mL anhydrotetracycline, and grown at 30 °C. The SI3 region in the genome was screened using colony PCR and primers flanking the SI3 insertion site and verified by Sanger sequencing. The two plasmids were cured as described (26).

E. coli cell microscopic imaging and cell size quantification

ΔSI3* and WT cells were cultured in MOPS rich defined medium (RDM) with 0.2% (w/v) glucose to early log-phase (apparent OD_{600} ~ 0.4). Cells (1 μL) were spotted on a homemade 15 mm×15 mm×2 mm 1% agarose pad to immobilize the cells and covered with glass coverslips. Cells were imaged on an Olympus IX-83 inverted microscope with a 60X phase contrast objective. Raw images were adjusted with brightness/contrast in ImageJ software (63) and the cell area was analyzed with "Analyze particle" module. The cell area distribution histograms were fitted with sigma distribution curves.

rpoC complementation experiment

To test if modified *rpoC* could support cell growth, plasmids harboring full-length wildtype. ΔSI3, ΔSI3*, or β′A941T *rpoC* expressed from a *trc* promoter were transformed into strain RL602 (28) in which expression of native *rpoC* is inhibited at >37 °C. Cells were cultured at 37 °C until log-phase, diluted sequentially, spotted on LB agar plate containing 0.5 mM IPTG, and incubated at 39 °C for 16 hours to score cell growth.

NET-seq strain construction

We generated an *rpoC*(ΔSI3*) strain with a C-terminal 3X FLAG tag and adjacent kanamycin-resistance marker by plasmid-mediated recombination as described previously (4, 31). The plasmid pYB403 was transformed into strain RL324 and double-crossover recombinants were selected on LB-Kan plates (50 μg/mL kanamycin). Colonies were restruck on LB-Kan plates for several passages to ensure plasmid loss, which was confirmed by loss of ampicillin resistance. Replacement of *rpoC* with the rpoCΔSI3*::3X FLAG was confirmed by PCR-Sanger sequencing. P1vir lysate grown on this strain was then used to transduce RL3000 to give strain RL4002.

In vitro transcription assay

In vitro transcription assays to measure elongation and pausing were performed as described previously (20), with minor modifications. For elongation assays, pre-assembled nucleic-acid scaffold [5 μM RNA, 10 μM template strand DNA (t-DNA)] was incubated with *E.coli* RNAP core enzyme in Elongation Buffer (20 mM Tris-acetate, pH 8.0, 40 mM KOAc, 5 mM $Mg(OAc)_2$, 1 mM DTT) with 0.5 μM RNA, 1.0 μM t-DNA, 1.5 μM RNAP for 15 min at 37 °C. Non-template strand DNA (nt-DNA) was added to 1.5 μM and the mixture was incubated for an additional

15 min at 37 °C. The estimated concentration of assembled EC was 500 nM. To incorporate ^{32}P, ECs were diluted to 150 nM with Elongation Buffer and 50 µL was incubated with 10 µCi [α-^{32}P] GTP (3,000 Ci/mmol) for 5 min at 37 °C. Labeled ECs were then extended to form A26 halted ECs by incubating with ATP and GTP (83 µM each final) for 5 min at 37 °C.

To ligate ECs to the 2.2 kb DNA template (PCR-amplified from plasmid pRL785), 5 nM A26 ECs were incubated with 15 nM purified DNA template and T4 DNA ligase (NEB; 25 unit/µL final) in T4 DNA ligase buffer (NEB) at 16 °C for 2 h. Further reconstitution was then blocked by incubation with heparin (0.1 mg/mL final) for 5 min at 37 °C. To re-start elongation, an equal volume of template-ligated EC was mixed with 4 NTPs (1 mM each). Reaction samples were withdrawn at 20 s and mixed with an equal volume of Stop Buffer (8 M urea, 50 mM EDTA, 90 mM Tris-borate buffer, pH 8.3, 0.02% bromophenol blue, 0.02% xylene cyanol). RNAs in all samples were resolved by denaturing PAGE (8% polyacrylamide, 19:1 acrylamide:bis-acrylamide, 45 mM Tris-borate, pH 8.3, 1.25 mM Na_2EDTA, 8 M urea), exposed to a Storage Phosphor Screen (GE Healthcare) and scanned with Typhoon PhosphorImager (GE Healthcare). An ^{32}P-labeled pBR322 MspI digested DNA ladder (NEB) was run on the gel for size reference.

The gels were quantified using ImageQuant software (GE Healthcare). To estimate the averaged elongation rates, the intensities of RNAs longer than 66 nt (the 66 nt band was generated by un-ligated ECs) after 20 s extension with 1 mM NTPs were determined and average elongation rates were calculated from the average sizes of these RNAs in 3 independent assays.

For scaffold-based pause assays, pre-assembled nucleic-acid scaffolds (5 µM RNA, 10 µM template strand t-DNA) were incubated with *E.coli* RNAP core enzyme in Transcription Buffer (10 mM Hepes, pH 7.9, 50 mM potassium glutamate, 10 mM magnesium glutamate, 0.1 mM EDTA, 1 mM DTT and 5 µg acetylated BSA/mL) with 0.5 µM RNA, 1.0 µM t-DNA, 1.5 µM RNAP for 15 min at 37 °C. Nt-DNA was added to 1.5 µM and the mixture was incubated for an additional 15 min at 37 °C. Further reconstitution was then blocked by incubation with heparin (0.1 mg/mL final) for 5 min at 37 °C. ECs were incorporation labeled using appropriate combinations of [α-^{32}P]NTP and unlabeled NTPs as specified in the figure legends. For example, the *his* hs-pause assay was performed by incubating 150 nM reconstituted EC (in 100–500 µL) with 10 µCi [α-^{32}P] CTP (3,000 Ci/mmol) for 3 min at 37 °C.

The labeled ECs were then extended to the pause site by adding CTP and UTP to 2 μM each, and a pause hairpin mimic was then generated by incubation with of antisense RNA oligo (2 μM final) for 5 min at 37 °C. Pause escape was initiated by addition of GTP to 10 μM and samples from various time points were quenched with Stop Buffer prior to analysisby 15% PAGE. To analyze pause kinetics, the pause fraction change through time was fit with bi-exponential decay equation:

$$y = F_{fast} * e^{-k_{fast}t} + F_{slow} * e^{-k_{slow}t}$$

F_{fast} and F_{slow} represent the fraction of RNA species entering the slow and fast escape phase and k_{fast} and k_{slow} describe their escape rates. The pause strength (τ) was calculated as:

$$\tau = F_{fast}/k_{fast} + F_{slow}/k_{slow}$$

For promoter-based pause assays, the ~400 bp transcription templates were PCR-amplified from pre-constructed plasmids (see Supplementary Table) followed by spermine precipitation to recover DNA free of primers (64).Transcription holoenzyme was formed by mixing 5 μM *E. coli* σ^{70} with 1 μM RNAP core enzyme in Transcription Buffer (see above) for 20 min at 37 °C. The core enzyme was incubated with other transcription components [50 nM holoenzyme, 40 nM DNA template, 0.5 unit/μL RNase inhibitor (Promega), 150 μM ApU dinucleotide, 10 μCi [α-^{32}P] GTP (3,000 Ci/mmol), 5 μM ATP, 5 μM UTP and 2 μM GTP] to form ^{32}P-labeled A26 halted ECs for 10 min at 37 °C. Further reconstitution was then blocked by incubation with heparin (0.1 mg/mL final) for 5 min at 37 °C. If used, NusA was added at 200 nM and incubated for an additional 5 min at 37 °C. To restart transcription, A26 halted ECs were mixed with 100 μM each NTP, incubated at 37 °C, and samples were removed at specified times and combined with an equal volume of Stop buffer. Higher NTPs (1 mM each) were added after the time course and incubated for an additional 1 min at 37 °C to generate a chase sample. To prevent RNA 2° structure interference by A26 sequence, 2 μM of the antisense DNA oligo (5′-ATACAACCTCCTTACTACAT, #15363) that specifically anneals with A26 RNA was included in the NTP mix. In assays to disrupt candidate hs-pauses with asDNA oligos, asDNA oligos were used at 2 μM final concentration. RNA species were resolved by 12% PAGE and imaged and quantified as described above.

NET-seq library preparation

Cell culture and harvest: A single colony of an *E. coli* NET-seq strain was inoculated into MOPS rich defined medium with 0.2% (w/v) glucose (RDM) (65) and 50 μg/mL kanamycin for overnight growth at 37 °C. A 2.5-mL portion of the overnight culture was inoculated into 0.5 L RDM with 50 μg kanamycin/mL (freshly prepared and sterilized with 0.22 μM filter) in 2L Erlenmeyer flasks. The cultures were grown on an orbital shaking at 200 rpm and 37 °C until cell density reached early log-phase (apparent OD_{600} ~0.4) and then rapidly filtered at 37 °C through a 0.22 μM cellulose filter. Cells were scrapped from the filter with a spatula and plunged into liquid nitrogen. Frozen cells were stored at –80 °C before processing. Biological triplicate samples were collected for each NET-seq strain. *B. subtilis* cells bearing a 3X FLAG tag on *rpoC* (RL3116) for use as a spike-in quantitation control were grown in parallel using the same procedure except with 7.5 μg kanamycin/mL at 37 °C. *B. subtilis* cells were recovered by centrifugation at 4,000 × *g* for 10 min, resuspended in 10 mL ice cold sterile PBS (137 mM NaCl, 2.7 mM KCl, 2 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH = 7.4) containing 20% (v/v) glycerol and 0.5 pill of protease inhibitor cocktail at 12.4 × 10^9 CFU/mL, and flash frozen in liquid nitrogen.

Nascent RNA extraction: Frozen cells harvested from 500 mL culture were were combined 0.5 mL *B. subtilis* spike-in cells per apparent OD_{600} unit of *E. coli* (~6.2 ×10^8 CFU spike-in cells per $OD_{E.coli}$, approximately 3.4% for WT samples and 1.4% for ΔSI3* samples) and 400 μL frozen lysis buffer (20 mM Tris-HCl, pH 8.0, 0.4% Triton-X 100, 0.1% NP-40, 100 mM NH4Cl, 10 mM $MnCl_2$, 250 Unit RNaseIn (Promega), 1X protease inhibitor cocktail, 0.4 mg/mL puromycin). The mixture was pulverized in a Retsch MM400 Mixer Mill (canisters pre-cooled in liquid nitrogen) 6 times at 15 Hz for 3 min. The lysate was resuspended on ice in 5 mL of lysis buffer by gentle pipetting. RQ1 DNase I (110 Unit total, Promega) was added and incubated for 15 min on ice. The reaction was quenched with EDTA (25 mM final) to releases ribosome from the transcripts. The lysate was then clarified at 4 °C by centrifugation at 25,000 ×g for 10 min and incubated with 0.5 mL anti-FLAG M2 affinity resin (Sigma-Aldrich) for 3 h at 4 °C. The resin was washed with 10 mL Wash Buffer (20 mM Tris-HCl, pH 8.0, 0.4% Triton-X 100, 0.1% NP-40, 300 mM KCl, 1 mM EDTA) 4 times at 4 °C and the ECs were then eluted with 600 μL Elution Buffer [20 mM Tris-HCl, pH 8.0, 0.4% Triton-X 100, 0.1% NP-40, 2 mg/mL FLAG peptide (Sigma-Aldrich)]. Nascent RNA

was extracted from the eluate with miRNeasy Mini Kit (Qiagen) following manufacturer's instructions. The nascent RNA sample was further purified by EtOH precipitation, dissolved in DEPC-treated water, and stored at –80 °C.

Preparation of cDNA libraries: NET-seq library preparation was performed following published procedures (66) with minor changes. The adaptor oligo (#14047, 6 μM) was 5′ adenylated by incubating with 80 μM ATP, 6 μM *Mth* RNA ligase (NEB), 25% (v/v) PEG 8000 in 1X NEB Buffer 1 at 65 °C for 4 hours, then 85 °C for 5 min. Nascent RNA (16.7 ng/μL, 1.5 μg per sample) was ligated to the adaptor by incubating with 1X T4 RNA ligase buffer (NEB), 25% PEG 8000, 6.7 U/μL T4 RNA ligase 2, truncated (NEB), 2 μM adaptor, 1.3 U/μL RNaseIn at 37 °C for 3 h. The ligation was stopped by adding EDTA to 16 mM. The adaptor-ligated RNA was alkaline fragmented by incubating with 6 mM Na_2CO_3, 44 mM $NaHCO_3$, 1 mM EDTA at 95 °C for ~35 min (time varied depending on the freshness of the alkaline solution) and the RNA was precipitated with equal volume of isopropanol and 0.3 M NaOAc (pH = 5.5) using 15 μg Glycogen-Blue (Thermo Fisher) as a co-precipitant. The fragmented, adaptor-ligated RNA was electrophoresed through a 15% polyacrylamide TBE urea (90 mM Tris base, 90 mM boric acid, 2 mM EDTA, 8M urea) gel (Novex, Thermo Fisher) and the 50–100 nt RNAs were excised and extracted as described previously (66). The RNA was reverse transcribed into cDNA in 1X First Strand Buffer (Thermo Fisher), 0.5 mM dNTPs, 0.3 μM RT adaptor (#14637), 0.6 Unit/μL SuperaseIn, 5 mM DTT, Superscript III (Thermo Fisher; 10 U/μL) at 50 °C for 60 min. RNA template was removed from the cDNA by incubation in 0.1 M NaOH at 95 °C for 20 min followed by neutralization with 0.1 M HCl. The cDNA was electrophoresed though a 10% polyacrylamide TBE urea gel (Novex, Thermo Fisher) and the 100–200-nt DNAs were excised and extracted as described previously (66). The cDNA library was circularized by incubation with 1X Reaction Buffer (Lucigen), 50 μM ATP, 2.5 μM $MnCl_2$, CircLigase (Lucigen; 5 U/μL) at 60 °C for 1 h and then at 80 °C for 10 min. The cDNA libraries were amplified with Q5 DNA polymerase (NEB) using PCR conditions recommended by the manufacturer and dual-indexed primers for the NovaSeq6000 (Illumina) sequencing platform, and then subjected to 150-nt paired-end sequencing.

NET-seq data analysis

NET-seq data analysis was performed using open-source programs written in Python (67) and custom scripts written in R 4.1.0 (68). Adaptor sequence in raw NET-seq reads was removed with Cutadapt (69) and aligned to *E. coli*

MG1655 reference genome (NCBI Reference Sequence: NC_000913.3) with Bowtie (70). Unwanted reads in ribosomal RNA regions and t-RNA regions were removed with SAMtools (71) and converted to BED format with BEDTools (72). Full length NET-seq reads were then trimmed to 15 nt from 3′ ends. The total read numbers and 3′ end read numbers in genome positions were quantified with BEDTools. To scale reads among strains and replicates, 3′ end read numbers (R_{raw}) were normalized according to the *E. coli* cell genomic DNA amount in cell harvested ($M_{E.\ coli}$), the *B. subtilis* spike-in cell number ($N_{spike\text{-}in}$) and the reads ($R_{spike\text{-}in}$) specifically aligned to *B. subtilis* genome (NCBI Reference Sequence: NC_000964.3) using the following formula.

$$\text{scaled 3′ end read } R_{scaled} = 2.5 \times 10^8 \times R_{raw} \times N_{spike\text{-}in} / (M_{E.\ coli} \times R_{spike\text{-}in})$$

To determine the relationship between apparent OD_{600} and DNA amount per 1 mL cell for the two NET-seq strains, the cells were cultured in the same way as the NET-seq procedure, harvested at two different apparent OD_{600} values in mid-log phase by rapidly mixing 0.5 mL culture with equal volume of ice-cold methanol to stop cell growth. The cells were spun down and lysed with TES solution (10 mM Tris-HCl, pH=8, 10 mM EDTA, 2% SDS) by incubating at 65 °C for 5 min. The lysate was extracted with equal volume of phenol:chloroform:IAA (Invitrogen) and the supernatant was sampled to determine genomic DNA concentration with Qubit dsDNA BR assay (Invitrogen). The linear relationships between apparent OD_{600} value and DNA amount per 1 mL culture (m, μg) were as follows:

$$\text{(WT NET-seq strain)}\ \ \mu g\ DNA/mL = 7.24 \times OD + 3.98,$$

$$\text{(ΔSI3* NET-seq strain)}\ \ \mu g\ DNA/mL = 9.83 \times OD + 9.34.$$

$M_{E.\ coli}$ values of NET-seq samples were calculated based on the strain, the apparent OD_{600} value of each sample at harvest and the culture volume harvested.

We identified pause sites using a four-round sliding window strategy in which a position was designated a pause if its read count was greater than 4 standard deviations higher than the mean for for non-zero counts in window from –100 to +99 ($z > 4$). For rounds 2–4, pause positions identified in the preceding round or rounds were removed and the sliding-window *z*-score calculation was repeated to identify additional pause sites that might have been masked in the preceding round by a nearby strong pause signal. Pause sites that were not identified in all three

biological replicates were removed from the final data sets. Additionally, putative pauses corresponding to tRNA or sRNA 3′ ends or known mRNA 3′ ends (73) were removed.

We estimated pause strength as the read number corresponding to a pause site RNA 3′ end divided by the number of 15-nt reads covering the location. We estimated changes in pause strength between WT and ΔSI3* strains by calculating the ratio of pause strength numbers for the two strains at a given pause position.

For pause hairpin prediction, RNA sequences from –50 to –11 positions (–1 corresponding to the 3′ end) were used for minimum free energy secondary structure prediction using the Vienna RNA package 2.0 (36). Stem loop structures with at least 4 base pairs in the stem region (allowing at most 1 mismatch inside the stem) were designated as putative pause hairpins (PH).

β-galactosidase activity assay

Plasmids with *lacZ* fusion reporters were transformed into VH1000 *E. coli* cells (Δ(*lacI–lacZ*). Fresh colonies were inoculated into MOPS RDM with glucose (0.2 % w/v) for overnight growth. On the next day, 10 µL of overnight culture was mixed with 190 µL RDM in a 96-well plate and shaken at 600 rpm at 37 °C for 2~3 hours until reaching mid-exponential phase. The cells were resuspended using a multichannel pipettor before recording apparent OD_{600} values on a plate reader (Tecan M1000). 10 µL A portion (10 µL) of cells was permeabilized by thoroughly mixing with 110 µL lysis buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, 0.005% w/v SDS) and 4 µL chloroform at room temperature. The ß-galactosidase reaction was initiated by addition of 24 µL ortho-Nitrophenyl-β-galactoside (ONPG, 4 mg/mL) and stopped by addition of 60 µL 1M Na_2CO_3 at the reaction time (*t*, min) recorded. Absorbance at 420 nm (A_{420}) and 550 nm (A_{550}) was measured using a Tecan plate reader (M1000). ß-galactosidase activity was calculated using the following formula.

$$U \text{ (Miller units)} = 1000 \times (A_{420} - 1.75 \times A_{550})/(OD_{600} \times t \times 0.081).$$

When used, bicyclomycin was added to cultures to 20 µg/mL final concentration.

Table of oligonucleotides, plasmids and strains

Oligonucleotides (5′ to 3′)	Assay	Stock #
EC scaffold, template strand DNA pCAAGGGACTGGTCTGAATCGGGAATACGGAACGGAAGATCTGAGTTCTC TTCCCCTCTAGCTCAGGACGTACTGACC	Elongation	#10160
EC scaffold, non-template strand DNA GGTCAGTACGTCCTGAGCTAGAGGGGAAGAGAACTCAGATCTTCCGTTCC GTATTCCCGATTCAGACCAGTCC	Elongation	#10161
EC scaffold, RNA UUUUUUUGAGCUAGAGG	Elongation	#8855
his hs-PEC scaffold, template strand DNA CCACTGGAAGACTCTGAATCTCTTCCAGCACACATCAGGACGTACTGACC	Pause	#11613
his hs-PEC scaffold, non-template strand DNA GGTCAGTACGTCCTGATGTGTGCTGGAAGAGATTCAGAGTCTTCCAGTGG	Pause	#11614
his hs-PEC scaffold, RNA GUCAUCCGGCGAUGUGUG	Pause	#12644
his hs-PEC scaffold, antisense RNA for –12 pause hairpin CCGGAUGA	Pause	#6598
his hs-PEC scaffold, antisense RNA for –11 pause hairpin GCCGGAUG	Pause	#6599

his hs-PEC scaffold, antisense RNA for –13 pause hairpin CGGAUGAC	Pause	#12645
consensus ePEC scaffold, template strand DNA CCAGTCATGCAGGCCGTATCAACGCGGGCATTTAAGTCTAGAAGGTCACC	Pause	#8334
consensus ePEC scaffold, non-template strand DNA GGTCAGTACGTCCGGCATAGTTGCGCCCGTAAATTCAGATCTTCCAGTGG	Pause	#9563
consensus ePEC scaffold, RNA UUUUUUGGCAUAGUU	Pause	#8401
new hs-PEC scaffold, template strand DNA CCACTGGAAGACTCTGAATCTCTAAGTCCACACATCGGGACGTACTGACC	Pause	#14948
new hs-PEC scaffold, non-template strand DNA GGTCAGTACGTCCCGATGTGTGGATCCAGAGATTCAGAGTCTTCCAGTGG	Pause	#14947
new hs-PEC scaffold, RNA GUCAUCCGGCGAUGUGUG	Pause	#12644
new hs-PEC scaffold, antisense RNA for –12 pause hairpin CCGGAUGA	Pause	#6598
Forward primer to amplify transcription template CGTTAAATCTATCACCGCAAGGG	PCR	#3071
Reverse primer to amplify transcription template CAGTTCCCTACTCTCGCATG	PCR	#645
Antisense DNA oligo to bind A26 sequence	Pause	#15363

ATACAACCTCCTTACTACAT		
Antisense DNA oligo to disrupt *hisL* pause hairpin TAGTCAGGATGATGGTGA	Pause	#15357
Antisense DNA oligo to disrupt *malT* pause hairpin GAAAAGCTGGGCGAAGAG	Pause	#15418
Antisense DNA oligo to disrupt *yfaS*$_{anti}$ pause hairpin GCAGCGTACTTTCTCTTC	Pause	#15362
Antisense DNA oligo to disrupt *fabH* pause hairpin GCTGCCAGTACCAATAAT	Pause	#15625
Plasmids	Source	Stock #
pCas9cr4 plasmid encoding Cas9	(26)	#4987
pYB301 plasmid encoding λ-red system and sgRNA targeting SI3 sequence (5′- TTACGCGTCAGACCGACGAA)	This work	#6376
pRM759 plasmid encoding β′ΔSI3, used for amplifying dsDNA for no-SCAR	This work, constructed by R. A. Mooney	#2859

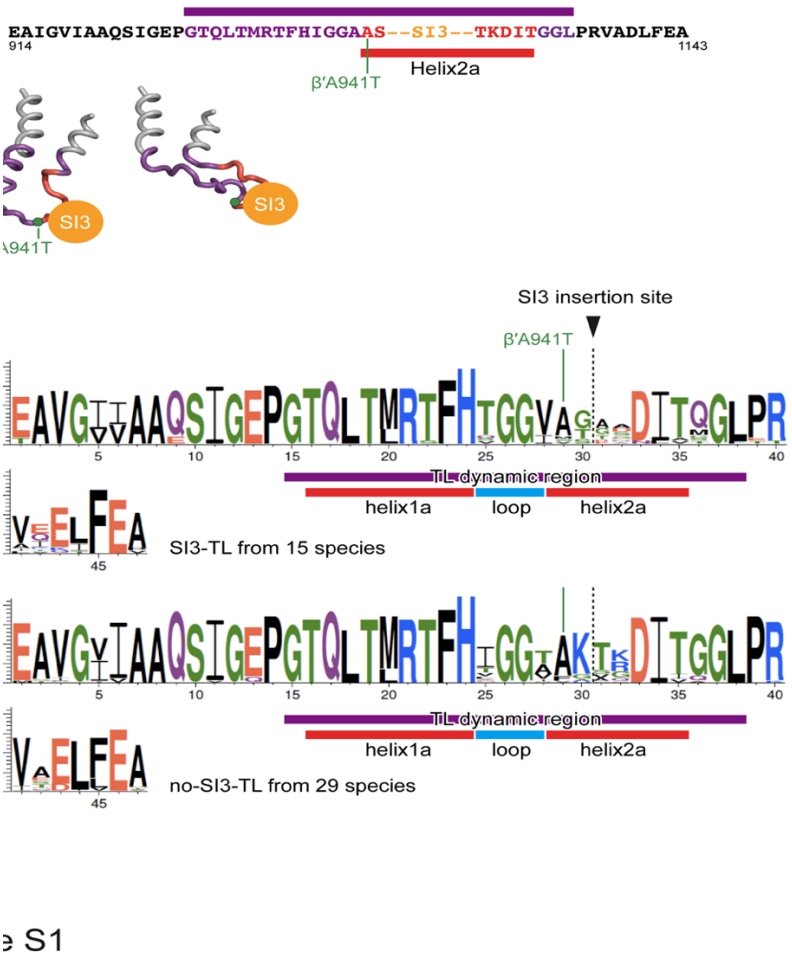
pRM839 plasmid encoding β′del_27, used for amplifying dsDNA for no-SCAR	This work, constructed by R. A. Mooney	#5139
pYB403 plasmid containing β′::3XFLAG, upstream and downstream genome sequences	This work	#6383
pRM756 *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus	(22)	#2956
pYB103 ΔSI3 *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus	This work	#6353
pYB104 ΔSI3* *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus	This work	#6354
pYB105 β′A941T *Eco*RNAP (α2ββ′ω) overexpression plasmid containing his10-ppx tag at β′ C-terminus	This work	#6355
pRL662 pTrc controlled β′ expression plasmid	(74)	#1662
pYB211 β′ deleted from pRL662	This work	#6371

pYB212 pTrc controlled β′ΔSI3 expression plasmid	This work	#6372
pYB213 pTrc controlled β′ΔSI3* expression plasmid	This work	#6373
pYB214 pTrc controlled β′A941T expression plasmid	This work	#6374
pYB417 *hisL* pause sequence (2090017…2090076) inserted after λPR+A26 C-less cassette	This work	#6397
pYB419 *malT* pause sequence (3553395…3553454) inserted after λPR+A26 C-less cassette	This work	#6399
pYB421 $yfaS_{\text{antisense}}$ pause sequence (2333944…2333993) inserted after λPR+A26 C-less cassette	This work	#6401
pYB408 *hisG* pause sequence (2090173…2090338) inserted after λPR+A29 U-less cassette	This work	#6388
pYB460 *fabH* pause sequence (1148759…1148818) inserted after λPR+A26 C-less cassette	This work	#6440
pYB461 *fabH* pause sequence (1148759…1148818) inserted after λPR+A26 C-less cassette, with point mutations disrupting pause hairpin	This work	#6441
pYB458	This work	#6438

rpoH pause sequence (3600679…3600783) inserted after λPR+A26 C-less cassette		
pYB459 *rpoH* pause sequence (3600679…3600783) inserted after λPR+A26 C-less cassette, with point mutations disrupting pause hairpin	This work	#6439
pYB443 *fabH* pause sequence (1148744…1148818) inserted between lacUV5 promoter and full-length lacZ	This work	#6423
pYB450 *fabH* pause sequence (1148744…1148818) inserted between lacUV5 promoter and full-length lacZ, with point mutations disrupting pause hairpin	This work	#6430
pYB445 *rpoH* pause sequence (3600697…3600798) inserted between lacUV5 promoter and full-length lacZ	This work	#6425
pYB452 *rpoH* pause sequence (3600697…3600798) inserted between lacUV5 promoter and full-length lacZ, with point mutations disrupting pause hairpin	This work	#6432
Bacterial strains	Source	Stock #
Escherichia coli K12 MG1655; F- λ- ilvG- rfb-50 rph-1	(62)	#1655
MG1655 thiC39::Tn10	This work	#2643
MG1655 *rpoC*(WT), no-SCAR strain for WT	This work	#4011
MG1655 *rpoC*(A941T, ΔSI3), no-SCAR strain for ΔSI3*	This work	#4010

RL324 F- λ- recB21 recC22 sbcB15 thr-1 leuB6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his4 argE3 rpsL31(Str) tsx-33 supE44	(31)	#324
RL3000 F- λ- *ilvG*468+ *rfb-50 rph+ ybhJ*(L54->I) *yebN*(G25xxx->D) *ycfK*::97bp ΔInsB-5 ΔInsA-5 ΔInsAB-5, a lab strain as WT for sequencing	(75)	#3000
RL3000 *rpoC*(A941T, ΔSI3)::3XFLAG, ΔSI3* strain for NET-seq	This work	#4002
RL3000 *rpoC*::3XFLAG, WT strain for NET-seq	This work, constructed by K. Jansen	#3511
B. subtilis 168 trpC2 *rpoC*::3XFLAG, used as NET-seq spike-in control	(4)	#3116
VH1000, RLG3499 MG1655 pyrE lacl lacZ	(76)	#3948

3.6 Supplementary figures and tables



e S1

Figure S3.1 TL sequence comparison among bacterial lineages and the location of SI3.

(A) Top, the dynamic region in ***E. coli*** TL and the locations of SI3 and the suppressor mutation β′A941T. Bottom, the folded (pdb: 8eg8) and unfolded (modified from pdb: 6rin) conformations of ***E. coli*** TL and the positions of SI3. (B) Sequence alignments of TL regions from bacteria species (see Supplementary Table) with/without a native SI3 domain.

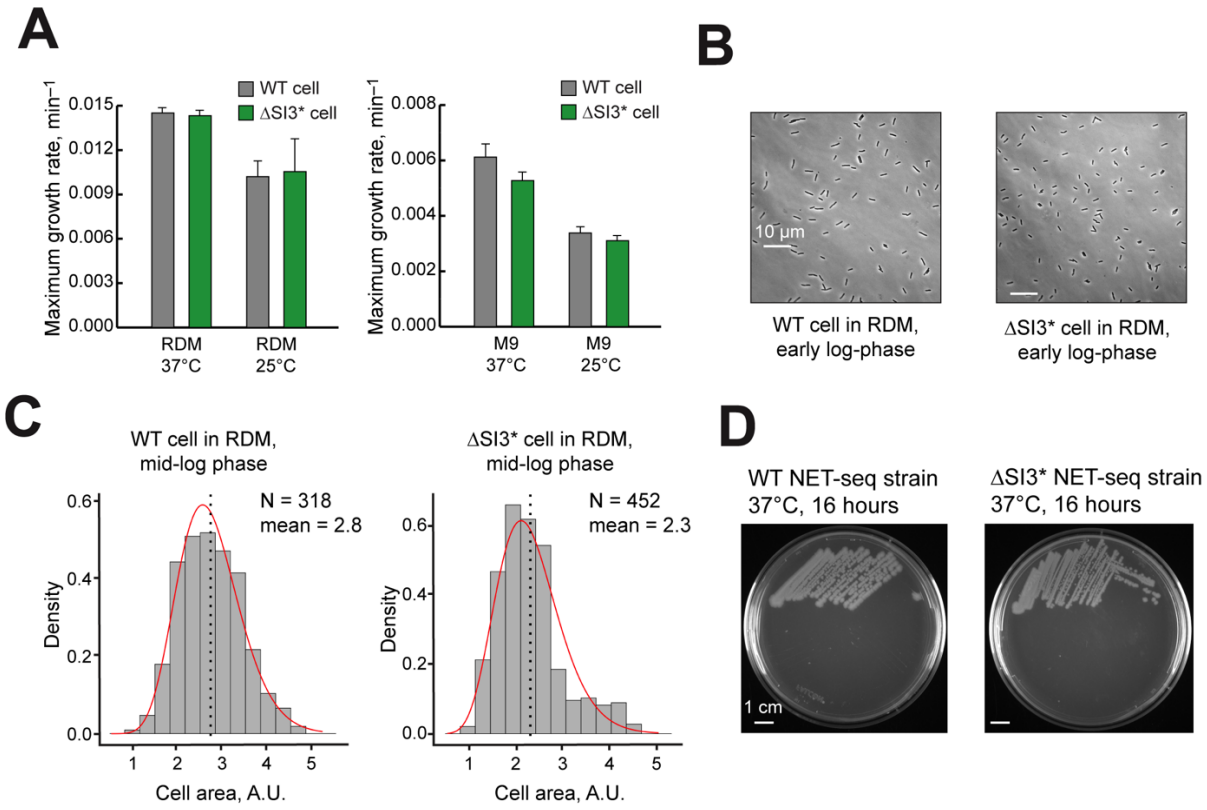


Figure S3.2. ΔSI3* strain phenotypes.

(A) Maximum growth rate comparison between ΔSI3* and WT cells (no-SCAR strains) in different culture conditions. RDM, MOPS glucose rich defined medium; M9, M9 minimal medium. (B) Microscopy images of early log-phase cells of ΔSI3* and WT, bar representing 10 μm. (C) Cell size distribution of ΔSI3* and WT cells in RDM at early log-phase. Histograms are fitted with sigma distribution in red curves. Black dotted lines represent mean of cell areas. (D) Colony size comparison between WT and ΔSI3* NET-seq strains.

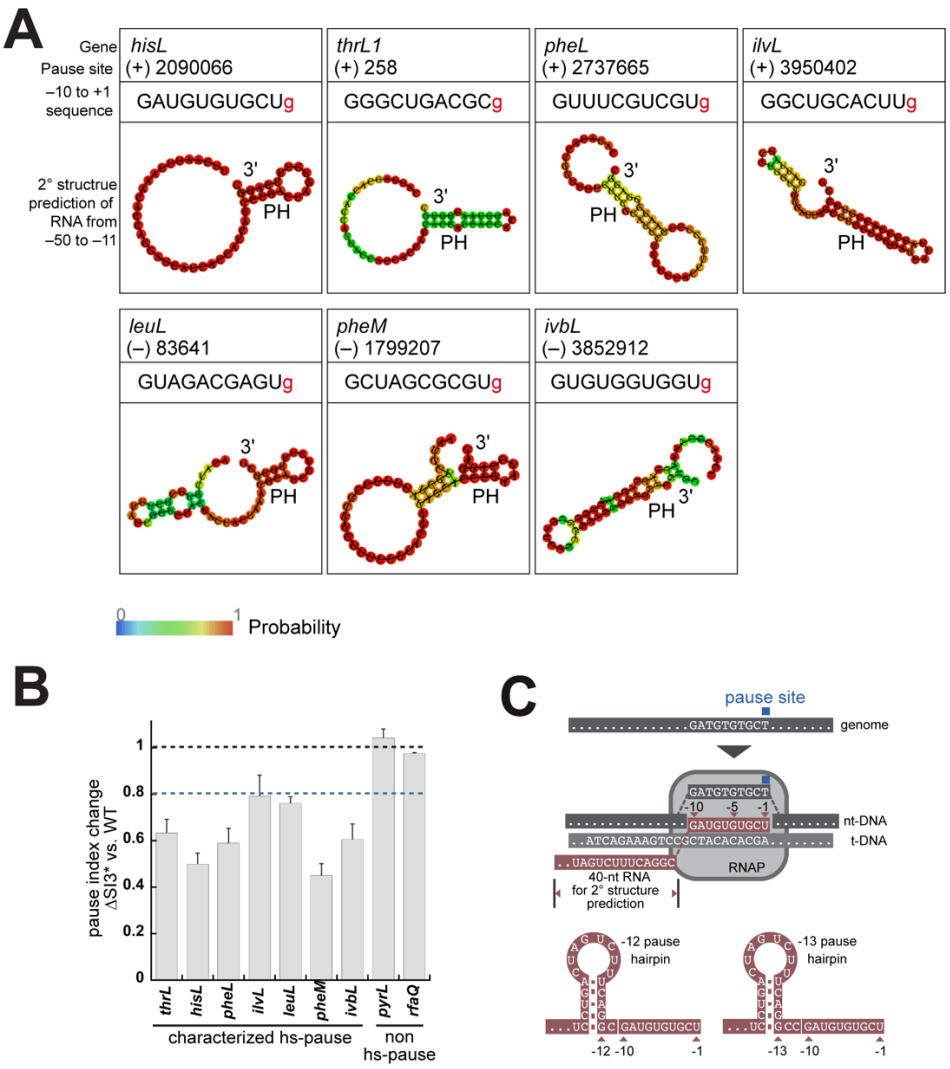


Figure S3.3. Known hs-pause sites in amino-acid biosynthetic regions.

(A) RNA secondary structure prediction of the seven characterized hs-pause sites. Pause sites are specified by genomic strand and the coordinates of the halted 3′ ends. (B) Pause index changes of the seven hs-pauses between WT and ΔSI3* datasets. In the right side of the bar graph, *pyrL* and *rfaQ* are non-hs-pause sites. Dark blue dotted line marks y = 0.8, which is used as the pause index decrease threshold in pause strength comparison (see Figure 5D–G). (C) The sequence fragment used for RNA secondary structure prediction. Below shows two examples of –12 PH and –13 PH.

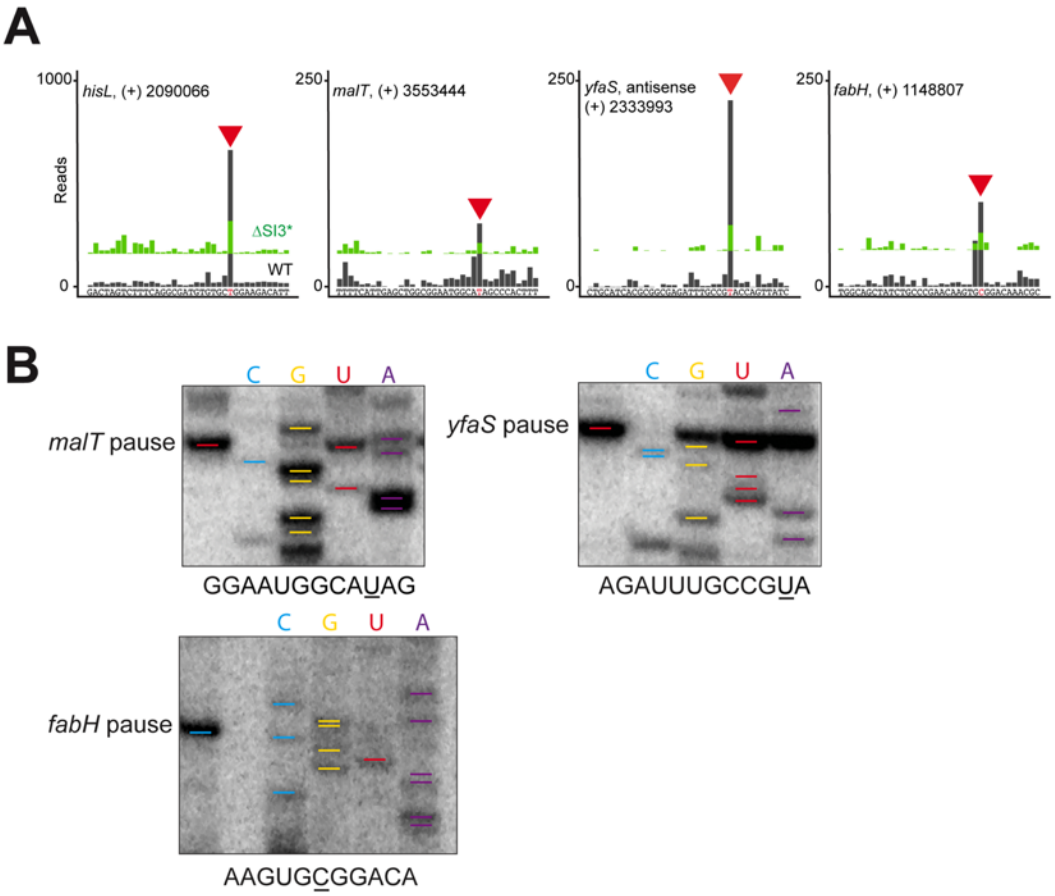


Figure S3.4. Candidates of putative hs-pause signals were picked for *in vitro* validation.

(A) NET-seq 3′-end read comparison between WT (black) and ΔSI3* (green) datasets for the three candidates. The detected pause sites are marked with red triangles. (B) *In vitro* transcription assay of pause candidates ran along with sequencing ladders.

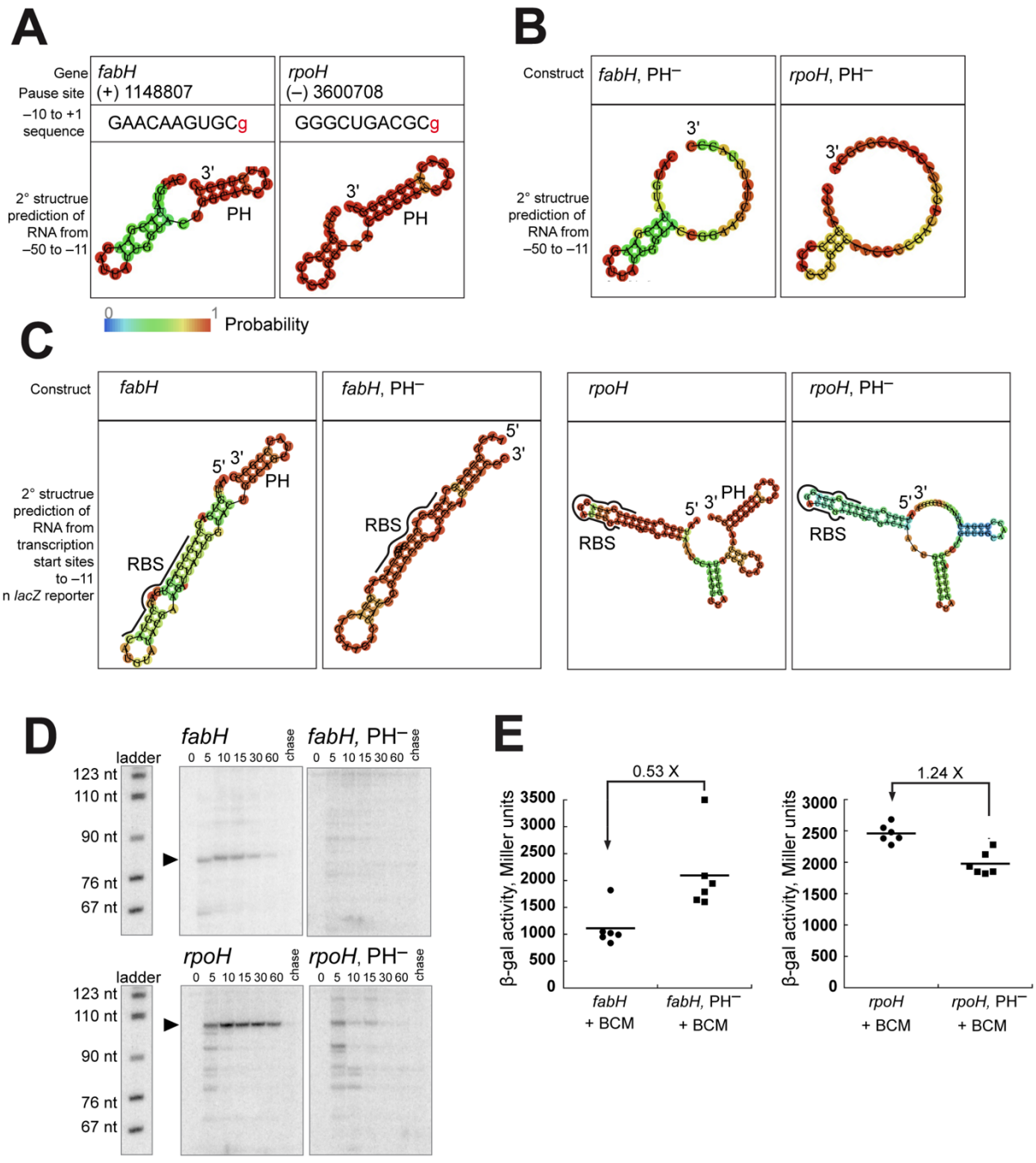


Figure S3.5. Two putative hs-pauses in fabH and rpoH were tested *in vivo*.

(A) RNA secondary structure prediction of the –50 to –11 RNA fragments of *fabH* and *rpoH* hs-pauses. Pause elements from –10 to +1 are listed in the middle. (B) RNA secondary structure prediction of the –50 to –11 RNA fragments of *fabH* and *rpoH* hs-pauses with base changes to disrupt PHs. (C) RNA secondary structure prediction of

the transcribed RNA from the lacZ reporter plasmid transcription start sites to the –11 positions of the pause signals. Native RBS sequences are labeled with black curves. (D) *In vitro* transcription assays of the two pause signals and their mutant constructs. Note that the DNA templates used in this assay are comparable to constructs in Figure 6C, which do not contain RBSs. (E) Expression comparisons of the *lacZ* reporters between hs-pause candidates and their mutants with bicyclomycin. Plot represents 6 biological replicates and their mean values.

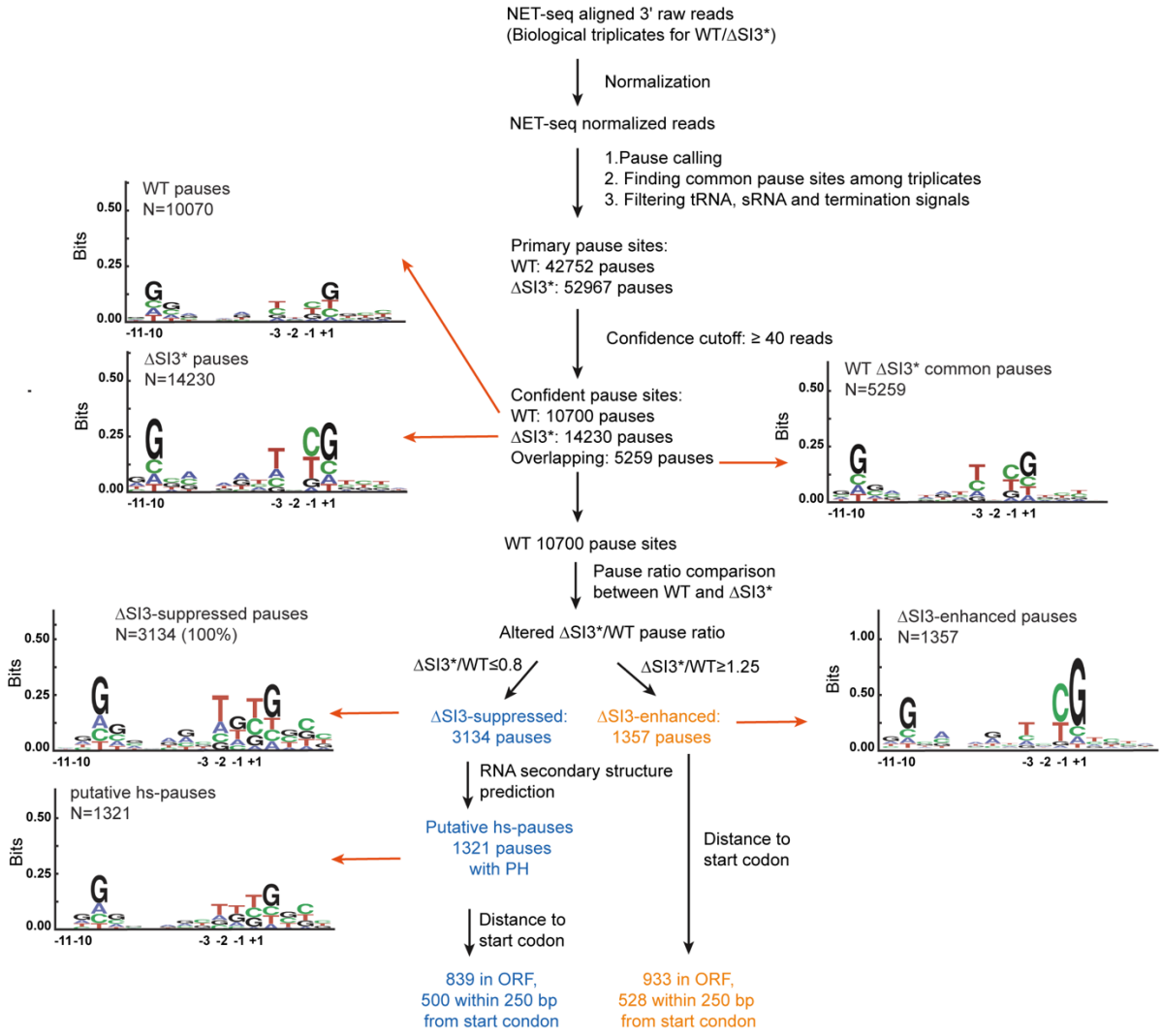


Figure S3.6. Data processing and pause signal sorting.

Sequence logos are listed along with sub-populations of the pause signals during data processing.

3.7 References

1. R. Landick, Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol* **75**, 291-314 (2021).

2. A. Mayer, H. M. Landry, L. S. Churchman, Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr Opin Cell Biol* **46**, 72-80 (2017).

3. F. Blombach, T. Fouqueau, D. Matelska, K. Smollett, F. Werner, Promoter-proximal elongation regulates transcription in archaea. *Nat Commun* **12**, 5524 (2021).

4. M. H. Larson *et al.*, A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).

5. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *Elife* **8** (2019).

6. I. O. Vvedenskaya *et al.*, Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285-1289 (2014).

7. L. S. Churchman, J. S. Weissman, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373 (2011).

8. M. Gajos *et al.*, Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res* **49**, 4402-4420 (2021).

9. A. V. Yakhnin *et al.*, NusG controls transcription pausing and RNA polymerase translocation throughout the Bacillus subtilis genome. *Proc Natl Acad Sci U S A* **117**, 21628-21636 (2020).

10. S. Gong, Y. Wang, Z. Wang, W. Zhang, Co-Transcriptional Folding and Regulation Mechanisms of Riboswitches. *Molecules* **22** (2017).

11. D. Lai, J. R. Proctor, I. M. Meyer, On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461-1473 (2013).

12. C. L. Chan, D. Wang, R. Landick, Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J Mol Biol* **268**, 54-68 (1997).

13. R. Landick, J. Carey, C. Yanofsky, Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc Natl Acad Sci U S A* **82**, 4663-4667 (1985).

14. C. L. Chan, R. Landick, The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J Biol Chem* **264**, 20796-20804 (1989).

15. J. Y. Kang *et al.*, RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* **69**, 802-815 e801 (2018).

16. J. Y. Kang *et al.*, An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. *Proc Natl Acad Sci U S A* **120**, e2215945120 (2023).

17. X. Guo *et al.*, Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e814 (2018).

18. L. M. Iyer, E. V. Koonin, L. Aravind, Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* **3**, 1 (2003).

19. M. Chlenov *et al.*, Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. *J Mol Biol* **353**, 138-154 (2005).

20. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proc Natl Acad Sci U S A* **118** (2021).

21. R. Furman, O. V. Tsodikov, Y. I. Wolf, I. Artsimovitch, An insertion in the catalytic trigger loop gates the secondary channel of RNA polymerase. *J Mol Biol* **425**, 82-93 (2013).

22. T. A. Windgassen *et al.*, Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res* **42**, 12707-12721 (2014).

23. I. Artsimovitch, V. Svetlov, K. S. Murakami, R. Landick, Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* **278**, 12344-12355 (2003).

24. E. F. Ruff *et al.*, E. coli RNA polymerase determinants of open complex lifetime and structure. *J Mol Biol* **427**, 2435-2450 (2015).

25. N. Zakharova, I. Bass, E. Arsenieva, V. Nikiforov, K. Severinov, Mutations in and monoclonal antibody binding to evolutionary hypervariable region of Escherichia coli RNA polymerase beta' subunit inhibit transcript cleavage and transcript elongation. *J Biol Chem* **273**, 24912-24920 (1998).

26. C. R. Reisch, K. L. J. Prather, Scarless Cas9 Assisted Recombineering (no-SCAR) in Escherichia coli, an Easy-to-Use System for Genome Editing. *Curr Protoc Mol Biol* **117**, 31 38 31-31 38 20 (2017).

27. T. M. Conrad *et al.*, RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proc Natl Acad Sci U S A* **107**, 20500-20505 (2010).

28. R. Weilbaecher, C. Hebron, G. Feng, R. Landick, Termination-altering amino acid substitutions in the beta' subunit of Escherichia coli RNA polymerase identify regions involved in RNA chain elongation. *Genes Dev* **8**, 2913-2927 (1994).

29. T. V. Mishanina, M. Z. Palo, D. Nayak, R. A. Mooney, R. Landick, Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci U S A* **114**, E5103-E5112 (2017).

30. L. You *et al.*, Structural basis for intrinsic transcription termination. *Nature* **613**, 783-789 (2023).

31. Z. Horii, A. J. Clark, Genetic analysis of the recF pathway to genetic recombination in Escherichia coli K12: isolation and characterization of mutants. *J Mol Biol* **80**, 327-344 (1973).

32. Y. H. Huang *et al.*, Structure-Based Mechanisms of a Molecular RNA Polymerase/Chaperone Machine Required for Ribosome Biosynthesis. *Mol Cell* **79**, 1024-1036 e1025 (2020).

33. M. Imashimizu *et al.*, Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol* **16**, 98 (2015).

34. S. Jeanneau, P. E. Jacques, D. A. Lafontaine, Investigating the role of RNA structures in transcriptional pausing using in vitro assays and in silico analyses. *RNA Biol* **19**, 916-927 (2022).

35. K. A. Salmon, C. R. Yang, G. W. Hatfield, Biosynthesis and Regulation of the Branched-Chain Amino Acidsdagger. *EcoSal Plus* **2** (2006).

36. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).

37. I. Toulokhonov, I. Artsimovitch, R. Landick, Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**, 730-733 (2001).

38. K. E. Kolb, P. P. Hein, R. Landick, Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J Biol Chem* **289**, 1151-1163 (2014).

39. N. B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512 (2001).

40. Q. Vicens, J. S. Kieft, Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci U S A* **119**, e2112677119 (2022).

41. I. Artsimovitch, R. Landick, Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**, 7090-7095 (2000).

42. S. A. Evfratov *et al.*, Application of sorting and next generation sequencing to study 5′-UTR influence on translation efficiency in Escherichia coli. *Nucleic Acids Res* **45**, 3487-3502 (2017).

43. G. A. Perdrizet, 2nd, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci U S A* **109**, 3323-3328 (2012).

44. P. Menendez-Gil, A. Toledo-Arana, Bacterial 3'UTRs: A Useful Resource in Post-transcriptional Regulation. *Front Mol Biosci* **7**, 617633 (2020).

45. J. Ederth, R. A. Mooney, L. A. Isaksson, R. Landick, Functional interplay between the jaw domain of bacterial RNA polymerase and allele-specific residues in the product RNA-binding pocket. *J Mol Biol* **356**, 1163-1179 (2006).

46. E. C. Ruteshouser, J. P. Richardson, Identification and characterization of transcription termination sites in the Escherichia coli lacZ gene. *J Mol Biol* **208**, 23-43 (1989).

47. M. E. Gottesman, A. Mustaev, Inorganic phosphate, arsenate, and vanadate enhance exonuclease transcript cleavage by RNA polymerase by 2000-fold. *Proc Natl Acad Sci U S A* **115**, 2746-2751 (2018).

48. M. H. Larson *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc Natl Acad Sci U S A* **109**, 6555-6560 (2012).

49. M. L. Kireeva *et al.*, Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell* **30**, 557-566 (2008).

50. C. Qiu *et al.*, High-resolution phenotypic landscape of the RNA polymerase II trigger loop. *PLoS genetics* **12**, e1006321 (2016).

51. J. A. Ballesteros, X. Deupi, M. Olivella, E. E. Haaksma, L. Pardo, Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. *Biophys J* **79**, 2754-2760 (2000).

52. T. M. Gray, B. W. Matthews, Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *J Mol Biol* **175**, 75-81 (1984).

53. R. Furman, A. Sevostyanova, I. Artsimovitch, Transcription initiation factor DksA has diverse effects on RNA chain elongation. *Nucleic Acids Res* **40**, 3392-3402 (2012).

54. A. Chauvier *et al.*, Structural basis for control of bacterial RNA polymerase pausing by a riboswitch and its ligand. *Nat Struct Mol Biol* 10.1038/s41594-023-01002-x (2023).

55. S. Dey *et al.*, Structural insights into RNA-mediated transcription regulation in bacteria. *Mol Cell* **82**, 3885-3900 e3810 (2022).

56. S. Hwang *et al.*, Structural basis of transcriptional regulation by a nascent RNA element, HK022 putRNA. *Nat Commun* **13**, 4668 (2022).

57. H. Bremer, P. P. Dennis, Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3** (2008).

58. D. A. Schneider, W. Ross, R. L. Gourse, Control of rRNA expression in Escherichia coli. *Curr Opin Microbiol* **6**, 151-156 (2003).

59. M. L. Kireeva, M. Kashlev, Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* **106**, 8900-8905 (2009).

60. G. Rhodes, M. J. Chamberlin, Ribonucleic acid chain elongation by Escherichia coli ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. *J Biol Chem* **249**, 6675-6683 (1974).

61. C. R. Reisch, K. L. Prather, The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Sci Rep* **5**, 15096 (2015).

62. F. R. Blattner *et al.*, The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462 (1997).

63. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

64. B. C. Hoopes, W. R. McClure, Studies on the selectivity of DNA precipitation by spermine. *Nucleic Acids Res* **9**, 5493-5504 (1981).

65. F. C. Neidhardt, P. L. Bloch, D. F. Smith, Culture medium for enterobacteria. *J Bacteriol* **119**, 736-747 (1974).

66. L. S. Churchman, J. S. Weissman, Native elongating transcript sequencing (NET-seq). *Curr Protoc Mol Biol* **Chapter 4**, Unit 4 14 11-17 (2012).

67. G. a. D. Van Rossum, Fred L., *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

68. R. D. C. Team (2010) R: A language and environment for statistical computing. (R Foundation for Statistical Computing).

69. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).

70. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

71. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

72. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

73. X. Ju, D. Li, S. Liu, Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat Microbiol* **4**, 1907-1918 (2019).

74. I. Toulokhonov, J. Zhang, M. Palangat, R. Landick, A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell* **27**, 406-419 (2007).

75. I. N. Ghosh, R. Landick, OptSSeq: High-Throughput Sequencing Readout of Growth Enrichment Defines Optimal Gene Expression Elements for Homoethanologenesis. *ACS Synth Biol* **5**, 1519-1534 (2016).

76. T. Gaal, M. S. Bartlett, W. Ross, C. L. Turnbough, Jr., R. L. Gourse, Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**, 2092-2097 (1997).

Chapter 4: Conclusions and future directions

4.1 Conclusions and significance of this work

The investigation of the correlation between the configurations and operations of large molecular machineries has long constituted the central focus of biochemistry and molecular biology. Since the discovery of RNA polymerase (RNAP) in the 1960s, a multitude of studies encompassing biochemical, genetic, and structural approaches have unveiled intricate details regarding the functionality of various RNAPs. These investigations have shed light on the fundamental role of RNAP in the regulation of gene expression and its remarkable adaptability to regulation. Upstream of transcription, RNAP receives regulatory signals, with alternative sigma factors being expressed under diverse cellular conditions, assigning specific gene clusters to RNAP core enzymes. Furthermore, diffusive protein factors bind to RNAP, thereby modifying its operational states, while small molecules directly interact with RNAP to modulate transcription efficiency. Throughout the transcription process, the core enzyme of RNAP itself discerns signals encoded within the gene sequences. Downstream of transcription, the intricate operational states of RNAP exert influences over various aspects such as gene expression levels, transcript length, RNA folding and maturation, as well as DNA integrity and topology. To accommodate the multifaceted functional characteristics of RNAP, both the evolutionarily conserved core module and the variable modules acquired over time play pivotal roles. Deciphering the function of the variable structural modules will undoubtedly enhance our comprehension of the fundamental operational mechanisms of RNAPs, as well as the adaptive evolution of species inhabiting specific natural niches and their distinct survival strategies.

Our research objective is to investigate the functional role of a widely distributed, yet sequence-variable element known as SI3 in bacterial RNAP during transcriptional elongation and pausing. Our primary research strategy includes employing well-established biochemical assays to examine the *in vitro* transcriptional properties, leveraging the wealth of structural information obtained from the recent surge in high-resolution cryo-EM structures of RNAPs in various transcriptional states, and employing high-throughput sequencing techniques in combination with computational analysis to directly probe transcriptional pausing states *in vivo*. Additionally, complementary research methods, such as AI-assisted structure prediction and phylogenetic analysis, have also provided valuable support in our investigation.

In **Chapter 2**, we employed an *in vitro* chemical crosslinking method to investigate the conformational changes of SI3 and its involvement in transcriptional elongation and pausing. Crosslinking methods have been widely utilized to stabilize protein complexes in fixed states or examine the distribution of conformational states under specific experimental conditions. The use of a Cys triplet reporter, involving the engineering of three different cysteine residues in RNAP, has proven successful in studying the movement of the clamp domain and the folding-unfolding conformational changes of TL. Therefore, it seemed natural to apply this method to explore the conformational changes of SI3. One major challenge we encountered with this method was the biased efficiency of crosslinking using the previously employed cystamine-DTT system: it was efficient for SI3–RH crosslinking but inefficient for SI3–SI1 crosslinking. This discrepancy can be attributed to the selectivity of the thiol groups' configuration in the S_N2 reaction process. To address this issue, we conducted screenings with various oxidative reagents and identified hydrogen peroxide (H_2O_2) as a superior crosslinking reagent that exhibited high efficiency for both crosslinking species without any directional preferences. Using our SI3-CTR/CPR systems, we discovered that SI3 undergoes organized movements in the NAC, depending on the folding states of TL. Remarkably, biasing SI3 towards the closed position leads to alterations in TL's conformational changes, consequently impeding the elongation cycle. Our hypothesis is that the obligatory movements of SI3 during NAC assist in the removal of secondary binding factors, ensuring the processivity of EC. Subsequently, we examined the role of SI3 movements in transcriptional pausing. Our findings confirmed that biasing SI3 towards the swiveled position is crucial for enhancing pausing in a hairpin-stabilized pause signal by inhibiting TL folding. Interestingly, biasing SI3 towards the closed position reduced hairpin-stabilized pausing while enhancing a hairpin-less elemental pause signal, suggesting that a closed SI3 may strengthen specific pause signals (these effects are described in detail in **Appendix A**). Lastly, we investigated the interactions between the jaw domain and SI3, which dictate the movement of SI3 in accordance with the central swiveling module. The residue β′Phe1199 in the jaw domain interacts with SI3 through an alcove on its surface, forming what we refer to as the "Phe pocket." Phylogenetic analysis revealed that the Phe pocket is conserved in gammaproteobacteria, although the phenylalanine site is more commonly occupied by a histidine residue. Given that the jaw domain is universally present in bacterial RNAPs, we propose that the specific interaction between the jaw and SI3 represents an adaptive evolution of *E. coli* for fine-tuning TL's conformational changes.

In summary, this chapter presents several novel findings regarding the SI3 movement. Firstly, we expanded the application of Cys-crosslinking methods in large protein complexes by introducing a more efficient and less biased oxidant that accounts for variations in the Cysteine side chain configuration. Secondly, we revealed the functional relevance of TL insertion in relation to elongation and pausing, suggesting that the bacterial transcription process may differ depending on the presence of SI3. Lastly, we identified a previously unrecognized interaction mode known as the Phe pocket among SBHMs, which holds promise for guiding protein design and engineering.

In **Chapter 3**, we employed an *in vivo* sequencing method called NET-seq to investigate changes in pause profiles in the absence of SI3 in *E. coli*. One significant challenge we encountered was the lethality associated with SI3 deletion, as previously suggested by studies conducted by others. However, a fortuitous breakthrough occurred when we serendipitously discovered a missense mutation, β′A941T in the TL region, which rescued *E. coli* cells upon SI3 deletion. Our validation experiments demonstrated that this point mutation in the TL region might modulate TL dynamics in a biphasic manner, depending on the presence or absence of SI3. In the presence of SI3, β′A941T biased the TL region toward an unfolded state, resulting in prolonged pausing and inhibited transcription. Conversely, when β′A941T was combined with ΔSI3, the elongation rate was comparable to that of wild-type RNAP *in vitro*, and the modified RNAP supported cell growth. Further investigation confirmed that the ΔSI3* (β′ A941T-ΔSI3) RNAP retained the characteristics of the ΔSI3 RNAP, namely weakening of the hs-pause and strengthening of the ce-pause. Collectively, the ΔSI3* system represents an ideal tool to probe pause profiles using NET-seq and uncover the functionality of SI3 within the cell.

By introducing spike-in controls and conducting replicated experiments, we performed NET-seq data collection from both WT and ΔSI3* strains. This approach facilitated a quantitative comparison between the two strains, accompanied by more rigorous statistical analysis. Deletion of SI3 led to substantial global alterations in pause sites and pause strength. We leveraged these changes to effectively discriminate between different types of pause signals. Remarkably, the enhanced pausing observed in the ΔSI3 strain exhibited a sequence feature resembling the previously defined consensus elemental pause sequence, which is consistent with our *in vitro* experiments demonstrating that SI3 enhances ce-pause. In contrast, the suppressed pausing observed in the ΔSI3 strain exhibited a less uniform sequence context, with resemblance to a combination of ce-pause and backtracked 1-nt ce-pause

sequence features. We hypothesize that a portion of the suppressed pausing in the ΔSI3 strain corresponds to hs-pause, which is stabilized by RNA secondary structures and exhibits reduced dependence on DNA context. Utilizing this framework, we screened the pool of ΔSI3-suppressed pauses for pause signals that harbor defined stem-loop RNA secondary structures in the RNA exit channel. This screening yielded many putative hs-pause signals, and several candidates among them were validated as representative hs-pause signals through *in vitro* transcription assays. Moreover, we identified a significant portion of putative hs-pause signals located within the gene body, which had received less attention in previous studies. Our *in vivo* experiments clearly demonstrated that hs-pause signals contribute to changes in gene expression levels, potentially by affecting RBS accessibility and the assembly of the transcription-translation coupled complex.

During the analysis of the putative hs-pause signals, as previously mentioned, the sequence context exhibited characteristics reminiscent of backtracked ce-pause signals. This observation raised questions regarding the origin of such backtracked species. They could potentially arise from Pi/PPi-induced 3′ end cleavage during NET-seq sequencing library preparation, or they could represent genuine paused species occurring just prior to reaching the elemental pause site. We designated the preceding pause in conjunction with the subsequent elemental pause as tandem pause sites. The investigation of NET-seq results uncovered many tandem pauses, with a significant portion attributed to ΔSI3-suppressed pause signals, particularly the putative hs-pauses. *In vitro* transcription assays demonstrated the presence of certain tandem pauses during transcription, and their interconversion was found to be correlated with rNTP/PPi concentrations. These findings contribute to our understanding of the mechanisms underlying pause formation.

To summarize this section, we undertook in depth studies in the function of SI3 in pausing with quantitative NET-seq methods and our results demonstrated in many aspects that SI3 is a global pause controller in a pause type-dependent manner. In addition, our results provide resourceful information about hs-pause across the genome, giving implications about the formation of this type of pausing, as well as addressing its function in ORF.

4.2 Possible future directions of interest

4.2.1 Single molecule experiments to study SI3 dynamics during transcription elongation and pausing

Single molecule FRET experiments have been employed in *E. coli* RNAP to investigate the dynamics of the transcriptional loop (TL) during NACs (1). In the existing study, the TL's flexible regions and DNA scaffolds were modified with fluorescent labels. However, this approach primarily focuses on the closing-opening cycles of TL, making it challenging to simultaneously examine elongation and pausing processes. We propose that SI3 could serve as a superior target for single molecule experiments due to its varying positional distribution during elongation and pausing. Potential sites for fluorophore engineering include the tip/solution-exposed surface of SI3, a position in the RH region, and an immobile position in SI1. Similar to our CTR crosslinking method, the closure of SI3 induced by TL folding would bring SI3 closer to the RH region, while swiveling would direct SI3 towards the SI1 region. By investigating the movement of SI3 during several rounds of NACs, it would be possible to accurately map specific positions of the EC on the nucleic acids scaffold, providing insights into conformational changes occurring at multiple consecutive sites. The obtained results could illuminate conformational changes of RNAP during the transition from the EC state to the PEC state and enable intriguing comparisons among multiple pausing states.

4.2.2 Screening of molecules to target jaw-SI3 interactions

Our biochemical findings suggest that the interaction between the jaw and SI3 is facilitated through a specific Phe-Phe pocket docking. Disrupting this interaction has a significant impact on transcriptional pausing as it prevents the swiveling of SI3 (2). Furthermore, we observed an increase in the elongation rate and a decrease in transcription fidelity, likely attributed to a more biased movement pattern towards the closed position in NACs. We hypothesize that chemical compounds mimicking the aromatic ring of phenylalanine side chain may disrupt the Phe-Phe pocket interactions. Taking a bold assumption, it is plausible that the Phe pocket acts as a metabolite sensor, influencing the transcription elongation rate by modulating the jaw-SI3 interactions. It would be valuable to investigate whether aromatic amino acids, related peptides, or RNA oligos can target the SI3 Phe pocket and evaluate their potential roles in transcriptional regulation.

4.2.3 ΔSI3 NET-seq in different culture conditions

Our quantitative NET-seq analysis compared pause profiles for WT and ΔSI3* strains, indicating a large scale of pause alterations (see **Chapter 3**). We also observed different phenotypes of ΔSI3* from WT strain in NET-seq cultural conditions, such as reduced cell sizes and reduced genome content per cell. All above information indicates the roles of SI3 function in pause regulation under rich medium condition. The role of SI3 transcription regulation in other conditions can be addressed, if the cell culture conditions are altered accordingly, such nutrient stress, stationary phase, viral infection and DNA damage. Under these conditions, RNA expression change and pause profile change could be compared side by side to unveil how SI3 and pausing regulate gene clusters in response to the environmental stress.

4.2.4 ΔSI3-suppressed pausing and complex RNA secondary structures

In our analysis of NET-seq data (see **Chapter 3**), we conducted a search for potential hairpin-stabilized pauses within the pool of ΔSI3-suppressed pause signals. This was based on the observation that ΔSI3 weakens the hs-pause *in vitro*. Additionally, we confirmed several hypothesized hs-pauses that exhibited RNA hairpin structures within the RNA exit channel. Recent investigations in the field of structural and biochemical studies (3-5) have revealed that RNA secondary structures, beyond the traditional stem-loop, can influence RNAP conformation and pausing. We propose that complex RNA secondary structures, such as pseudoknots, may enhance pausing and therefore could be captured by our NET-seq results.

To identify these pause signals stabilized by RNA structures, we intend to reanalyze the NET-seq data by extending the RNA window for secondary structure prediction. We will then select candidates that exhibit strong pause-enhancing effects. Subsequently, we will clone the identified signals into *in vitro* transcription systems to validate their pausing effects. However, a significant challenge in this endeavor is the underdeveloped nature of structure prediction for long RNA sequences. The folding of lengthy RNA molecules in solution is influenced not only by base-pairing and base-stacking interactions but also by solution conditions and the varying rates of RNA synthesis in different regions (6, 7). Despite this obstacle, we hope that our work will uncover novel instances of RNA structure-regulated pausing events in *E. coli*.

4.2.5 Deconvoluting pause subtypes with deep learning

The current approach to identifying and defining elemental pause signals involves generating a consensus sequence logo from a large set of pause signals detected by NET-seq (8-10), without considering the varying strengths of individual signals. However, based on *in vitro* verifications of pause signals, it is possible that elemental pausing is guided by more than one pausing sequence. The presence of different elements, such as upFJ, Hyb, dsFJ, and dsDNA, may contribute to the final pause strength in a non-additive manner (11). Deep learning techniques, such as deep neural networks and autoencoders, can be employed to learn hierarchical representations of the combined data (12). By training the network on a substantial dataset consisting of mixed examples with known source components, the network can learn to disentangle and separate the individual sources. Deep neural networks have demonstrated success in sequence motif recognition and classification of transcription factor binding sites (13). Thus, we propose the application of deep learning to our NET-seq data in order to identify pause signals with distinct sequence features.

4.3 References

1. A. Mazumder, M. Lin, A. N. Kapanidis, R. H. Ebright, Closing and opening of the RNA polymerase trigger loop. *Proc Natl Acad Sci U S A* **117**, 15642-15649 (2020).

2. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proc Natl Acad Sci U S A* **118** (2021).

3. S. Hwang *et al.*, Structural basis of transcriptional regulation by a nascent RNA element, HK022 putRNA. *Nat Commun* **13**, 4668 (2022).

4. A. Chauvier, J. F. Nadon, J. P. Grondin, A. M. Lamontagne, D. A. Lafontaine, Role of a hairpin-stabilized pause in the Escherichia coli thiC riboswitch function. *RNA Biol* **16**, 1066-1073 (2019).

5. A. Chauvier *et al.*, Structural basis for control of bacterial RNA polymerase pausing by a riboswitch and its ligand. *Nat Struct Mol Biol* 10.1038/s41594-023-01002-x (2023).

6. D. E. Draper, A guide to ions and RNA structure. *RNA* **10**, 335-343 (2004).

7. T. Pan, T. Sosnick, RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35**, 161-175 (2006).

8. M. H. Larson *et al.*, A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).

9. I. O. Vvedenskaya *et al.*, Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285-1289 (2014).

10. M. Gajos *et al.*, Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res* **49**, 4402-4420 (2021).

11. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *Elife* **8** (2019).

12. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436-444 (2015).

13. J. Lanchantin, R. Singh, B. Wang, Y. Qi, Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks. *Pac Symp Biocomput* **22**, 254-265 (2017).

Appendix A: An ensemble of different states in the elemental paused transcription complexes

Adapted from:

Kang, J. Y., Mishanina, T. V., Bao, Y., Chen, J., Llewellyn, E., Liu, J., Darst, S. A., & Landick, R. (2023). An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. Proceedings of the National Academy of Sciences, 120(8).

Yu Bao's contribution:

In vitro experiments of *his* elemental pause and consensus elemental pause, Figure 5 and Figure 6

A.1 Abstract

Transcriptional pausing plays a critical role in regulating cellular RNA synthesis; however, its underlying mechanism remains incompletely elucidated. The dynamic and multi-domain RNA polymerase (RNAP) engages in sequence-specific interactions with DNA and RNA, leading to reversible conformational changes at pause sites that transiently interrupt the nucleotide addition cycle (NAC). These interactions initiate the rearrangement of the elongation complex (EC) into an elemental paused EC (ePEC). Additional rearrangements or interactions with diffusible regulators can transform ePECs into longer-lived paused ECs (PECs). For both bacterial and mammalian RNAPs, a half-translocated state in which the next DNA template base fails to load into the active site appears central to the ePEC. Some RNAPs also exhibit swiveling motion between interconnected modules, potentially contributing to the stabilization of the ePEC. However, it remains uncertain whether swiveling and half-translocation are essential characteristics of a singular ePEC state or if multiple ePEC states exist. To address this question, we employed biochemical probes that target ePECs with distinct DNA-RNA sequences, enabling the identification of discrete ePEC states. Notably, these ePECs occupy different translocation states with varying degrees of swiveling, suggesting that the formation of the posttranslocated state at specific DNA-RNA sequences may be a central feature of the ePEC. The existence of multiple ePEC conformations has significant implications for our understanding of transcriptional regulation.

A.2 Introduction

Transcriptional pausing represents an inherent characteristic observed in all DNA-dependent RNA polymerases (RNAPs) within cells, playing a crucial role in the regulated expression of genes. When encountering pause sites, RNAP interactions with specific RNA-DNA sequences lead to temporary halting of transcription, which persists for durations ranging from tens to thousands of times longer than the average nucleotide addition cycle (NAC) of 10 to 20 ms, as observed in *Escherichia coli* RNAP (1). These transient pauses provide the necessary time for essential processes such as nascent RNA folding into biologically active conformations, interactions with diffusible regulators, coordination with translation or splicing, termination of transcription, and other regulatory events involved in gene expression (2-6).

While previous findings suggest that the paused state is characterized by half-translocation and swiveling (7, 8), it remains uncertain whether all paused elongation complex (PEC) states exhibit these features, whether multiple PEC states undergo interconversion, and whether different pause signals generate distinct pause states. To gain deeper insights into the elemental pause state(s), we utilized biochemical techniques to investigate RNAP conformation and translocation using strong, consensus elemental pause signals (*con*-ePEC) and a weaker signal that forms ePECs prior to stabilization by a hairpin (PH) structure in the leader region of the *his* operon (*his*-ePEC). Our findings reveal that the two types of ePECs have distinct translocation states, with the common attribute of preventing transition to the posttranslocated state. Furthermore, the two ePECs exhibit distinct global RNAP conformations.

A.3 Results

A.3.1 Cys-triplet reporter (CTR) verifies large differences in *con*-ePEC and *his*-ePEC states.

The strong pre-translocation bias of the *con*-ePEC differs from the half-translocated, swiveled states that dominated the directly reconstituted *his*-ePEC. We next asked if these differences were also evident under conditions of active transcription in solution, using longer, fully complementary scaffolds that lack perturbing influences of an artificial transcription bubble and short downstream DNA present in cryo-EM complexes (**Figure A.1** and **Figure A.2**). We first used a Cys-triplet reporter (CTR) assay to distinguish closed and swiveled positions of SI3 in *con*-ePEC and *his*-ePEC (**Figure A.1B** and **Figure A.2B**) (9). In the CTR assay, the closed SI3 forms a disulfide between engineered β′1051C in the SI3 hairpin loop and β′671C at the tip of the RH. Alternatively, the swiveled SI3 instead forms a β′1051C disulfide with β267C in SI1. These two disulfides are distinguishable by non-reducing SDS-PAGE (**Figure A.1A** and **Figure A.2C**). The ratio of the closed to swiveled disulfide, denoted SI3 positional bias (SPB), is a relative measure of influence of transcription complex conformation on SI3 position (9).

To ask if SPB differed in *con*-ePEC and *his*-ePEC, we reconstituted the two ePECs by active elongation from an EC positioned one nucleotide upstream of the pause site (G16 for *con*-ePEC or C19 for *his*-ePEC). *his*-ePEC$_{-1}$ is known to also pause weakly(10). When compared to the –1 ECs, SPB increased for *con*-ePEC and decreased for *his*-ePEC (**Figure A.1C**). Consistent with previous results, formation of a pause-hairpin mimic using an antisense RNA further decreased SPB for *his*-ePEC. These results indicate that SI3 became more biased toward the closed position in *con*-ePEC, consistent with the pre-translocated *con*-ePECs observed by cryo-EM. Conversely, SI3 became more swiveled in the *his*-ePEC, also consistent with predominant occupancy of swiveled states observed by cryo-EM. Using this assay orthogonal to cryo-EM and scaffolds devoid of potentially perturbing influences, these results confirm that active *con*-ePEC and *his*-ePEC formed on scaffolds occupy predominantly pre-translocated and predominantly swiveled conformations, respectively.

The disulfides used in the CTR assay also can be used individually to bias ECs toward the SI3-closed or swiveled conformations (9). To ask if biasing SI3 toward the closed conformation would affect the pause strength of *con*-ePEC and *his*-ePEC similarly or differently, we measured the effect of the closed disulfide on pause kinetics (**Figure A.1D** and **Figure A.2D, E**). Trapping the closed SI3 with a disulfide dramatically enhanced the pause

strength and lifetime of the *con*-ePEC (**Figure A.1D** and **Figure A.2D**). In contrast, the same disulfide modestly increased the strength of the his-ePEC and decreased the strength of the hairpin-stabilized hisPEC (**Figure A.1D** and **Figure A.2E**). These contrasting effects of the closed SI3 disulfide are consistent with bias toward the pre-translocated state in the con-ePEC that is greatly strengthened by the disulfide and a bias toward the swiveled conformation in the his-ePEC that is countered by the disulfide, resulting in a modest effect on pause lifetime.

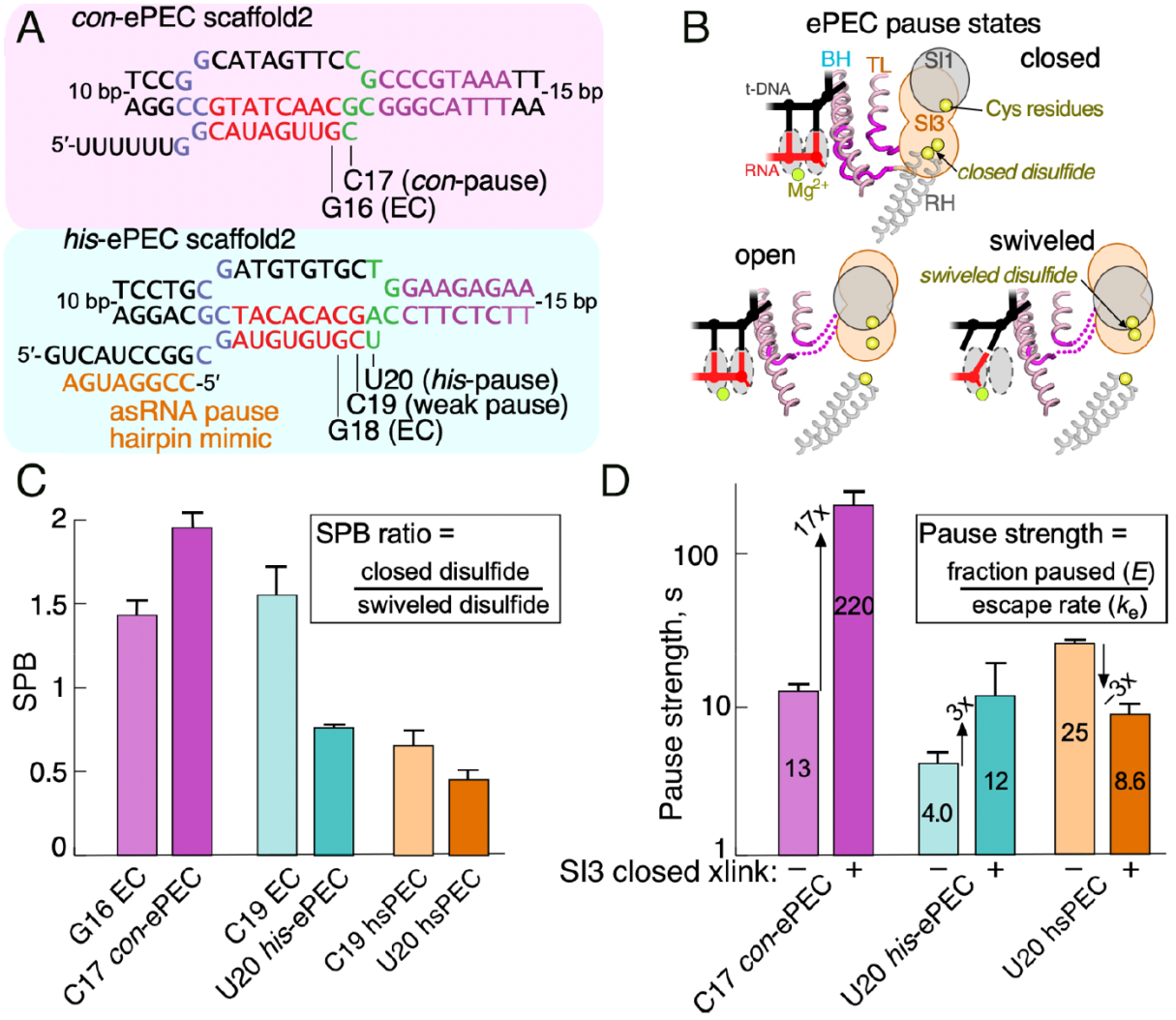


Figure A.1. Measurement of SI3 location by Cys-triplet reporter (CTR) assay and pause kinetics for *con*-ePEC and *his*-ePEC.

(A) The scaffold sequences used for CTR and pausing assays of *con*-ePEC and *his*-ePEC. These scaffolds are fully complementary and contain sufficient duplex downstream DNA to avoid perturbing effects on translocation register (see **Figure A.2** for complete sequences). (B) Diagrammatic representation of Cys residue locations near the active site of RNAP for CTR assay in SI3 closed, open, and swiveled states (9). (C) SI3 positional bias (SPB) as calculated from the CTR assay for ECs and PECs. A higher SPB indicates greater occupancy of the SI3 closed conformation. Note that absolute SPB values do not correspond directly to SI3 closed-to-swiveled ratios but that changes in SPB

roughly approximate the changes in these ratios (9). The increase in SPB for *con*-ePEC vs *con*-ePEC$_{-1}$ and decrease in SPB for *his*-ePEC vs. *his*-ePEC$_{-1}$ indicates that *con*-ePEC favors the closed SI3, presumably pre-translocated, conformation related to an EC whereas *his*-ePEC favors the swiveled SI3, presumably half-translocated conformation, related to an EC. (D) Pause strengths of paused ECs with and without a disulfide crosslink that restrains SI3 in the closed position (closed SI3 xlink). The *his*-ePEC results depicted here are from experiments reported previously (9).

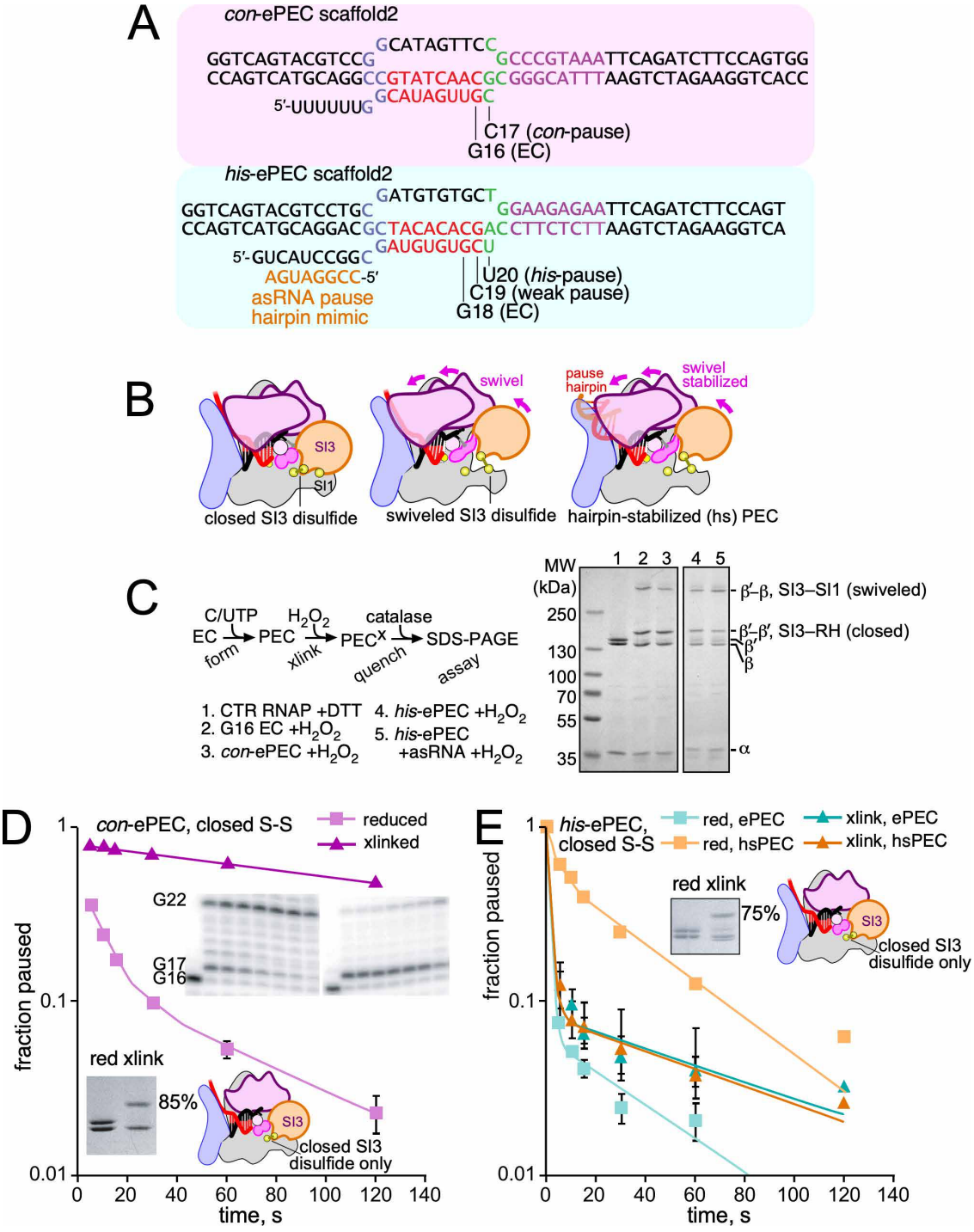


Figure A.2. Cys-triplet reporter analysis of ePEC conformations.

(A) The scaffold sequences used for CTR and pausing assays of con-ePEC and *his*-ePEC. These scaffolds are fully

complementary and contain sufficient duplex downstream DNA to avoid perturbing effects on translocation register (B) Conformations of the EC and ePEC sampled during the CTR assay. (C) The CTR assay (11). The representative gel panel shows relative migration in SDS-PAGE of the SI3–RH (β′–β′) and SI3–SI1 (β–β′) crosslinks for samples listed by lane number. The RNAP diagrams depict relative Cys locations in SI3 closed, open, and swiveled states showing the swivel module for RNAP. (D) Pausing kinetics of con-ePEC on the complete complementary scaffold (panel A) when the SI3–RH (closed) disulfide is formed using H_2O_2 or is reduced using DTT (11). The RNAP used for this experiment contained only two added Cys residues so the closed (SI3–RH) disulfide was formed exclusively. The insets show representative gel panels of the denaturing RNA gel for the pause assay and the extent of crosslinking assayed by SDS-PAGE. The strong increase in pausing when the crosslink is formed is consistent with a predominant role of the pretranslocated state for con-ePEC. (E) Pausing kinetics of *his*-ePEC on the complete complementary scaffold (panel A) when the SI3–RH (closed) disulfide is formed or reduced as described for panel D. The minimal effect of the crosslink on pausing by *his*-ePEC compared to the strong inhibition pausing for the hairpin-stabilized his-PEC is consistent with a lesser role of the pre-translocated state for *his*-ePEC compared to *con*-ePEC (panel D). The *his*-ePEC results depicted here are replotted from data reported previously (11).

A.3.2 A translocation register assay confirms different *con*-ePEC and *his*-ePEC states.

As a further test of differences in translocation bias of the *con*-ePEC and *his*-ePEC, we next examined reconstituted ePECs using an assay that traps ECs or PECs in their current translocation register with the phage HK022 Nun protein. This assay reports the post-translocated state based on its capacity for nucleotide addition and the pre-translocated state based on its sensitivity to pyrophosphorolysis (12). For this assay, we reconstituted *con*-ePEC and *his*-PEC on scaffolds similar to those used for cryo-EM but fully complementary to eliminate possible perturbation of translocation register caused by non-complementarity (**Figure A.3** and **Figure A.4**). In the Nun-trapped translocation assay, Nun binds the upstream fork junction and downstream DNA of ECs or PECs to block translocation (12, 13), and then NTPs or PPi are added to detect fractions of post- and pre-translocated complexes. It was unclear how the half-translocated PEC would be affected by Nun. We formed $ePEC_{-1}$ by reconstitution, bound to Ni^{2+}-NTA beads via an RNAP His_{10} tag, extended to ePECs by incubation with CTP (for *con*-ePEC) or UTP (for *his*-ePEC), treated with Nun, washed away NTPs, and then incubated with 1 mM GTP or 2.5 mM PPi to assay translocation register (**Figure A.4**). Nun-treated *con*-ePEC was more sensitive to pyrophosphorolysis and incorporated less GTP, whereas the opposite was true for *his*-ePEC and to an even greater extent for the hairpin-stabilized *his*PEC (**Figure A.3B** and **Figure A.4D–G**). Since the *his*PEC is known to be half-translocated, we infer the Nun-bound, half-translocated EC can react with NTP substrate. Taken together, these data are thus fully consistent with the cryo-EM structures (14) suggesting that *con*-ePEC is principally pre-translocated and *his*-ePEC is principally half-translocated.

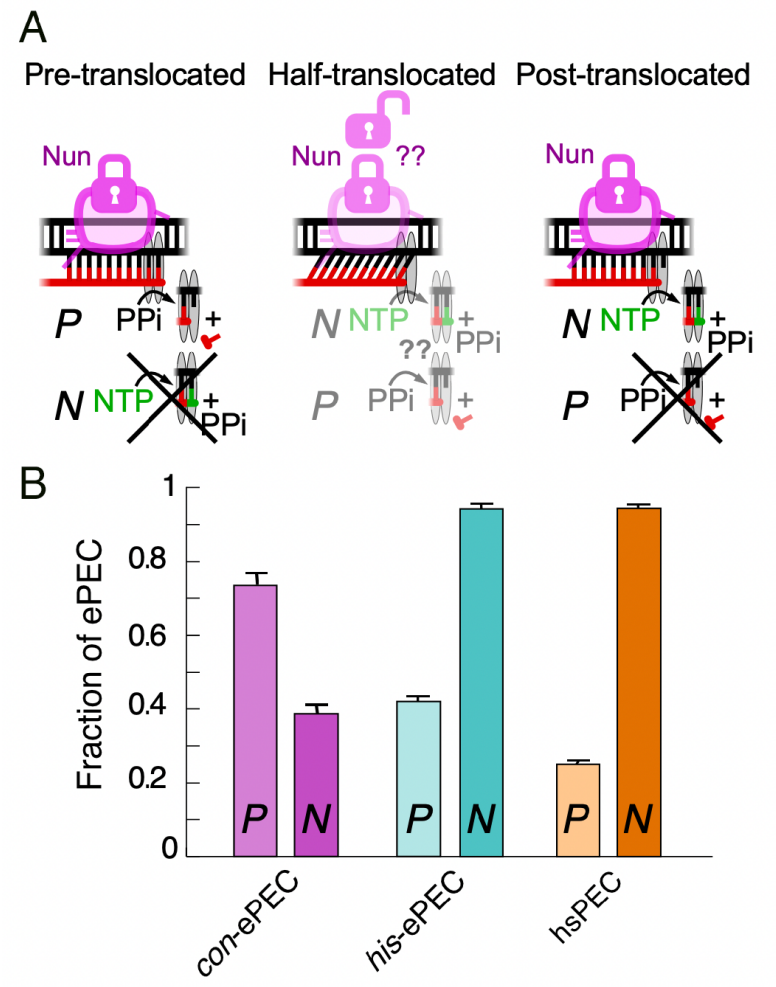


Figure A.3. Translocation registers of *con*-ePEC and *his*-ePEC determined by Nun-locked assay (12).

(A) Diagrammatic representation of Nun-locked translocation assay for pre-, half-, and post-translocated states. Nun locks translocation register via interactions with the upstream fork junction and downstream duplex DNA. Pre-translocated ECs are susceptible to pyrophosphorolysis but not nucleotide addition. Post-translocated ECs are susceptible to nucleotide addition but not pyrophosphorolysis. Whether Nun locks the half-translocated state and its subsequent reactivity are uncertain, but results presented here are consistent with either reactivity with NTP not PPi, or Nun-induced conversion to a post-translocated state reactive with NTP but not PPi. (B) Fractions of PECs reactive with PPi (P) or NTP (N) after incubation with Nun. Greater reactivity with PPi and lesser reactivity with NTP of *con*-ePEC that *his*-ePEC or hairpin stabilized (hs) *his*-hsPEC is consistent with greater occupancy of pre-translocated register for the *con*-ePEC and greater occupancy of half-translocated registers by *his*-ePEC and *his*-

hsPEC as also observed by cryo-EM analyses. Data are average and s.d. of 3 independent replicates from gels shown in **Figure A.4C–E**.

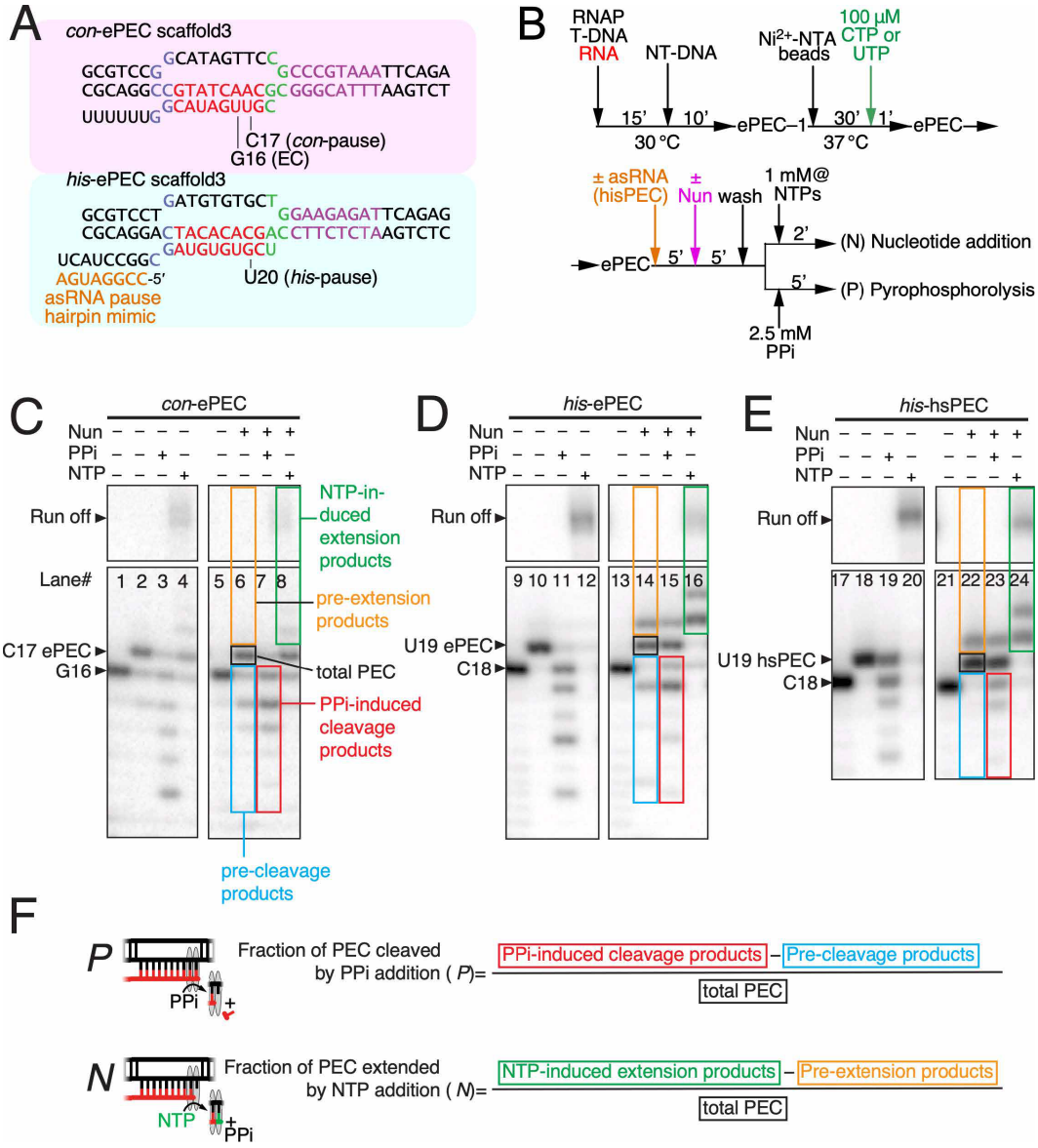


Figure A.4. Nun-locked catalysis assay of ePEC translocation register.

(A) Scaffolds used for Nun-locked translocation assay. These scaffolds differ from those used for cryo-EM (Fig. 1B) only in being fully complementary to avoid any perturbing effects of fork-junction base-pairing equilibria on translocation register. (B) Steps in the Nun-locked translocation assay. (C, D and E) Gel images for Nun-locked translocation register assays with different scaffolds. The right panel of each image set (lanes 1−4, 9−12, 17−20) shows the NTP extension and pyrophosphorolysis species without Nun addition. Left panel of each image set: red boxes (lanes 7, 15, 23), shortened bands were induced by pyrophosphorolysis; green boxes (lanes 8, 16, 24),

extended bands were induced by NTP addition. Note that even without PPi/NTP addition (lanes 6, 14, 22), there are distinct species from PEC, possibly resulting from RNAP backtracking/autocleavage (blue boxes) and residual NTPs in Nun protein stock (orange boxes). These pre-reaction species were subtracted during quantitation to accurately calculate reaction products generated in Nun-locked complexes. (F) Quantification formula for fraction of PEC cleaved by PPi addition (P) and fraction of PEC extended by NTP addition (N). Colors in the formula correspond to the boxes on the gel images.

A.4 Discussion

An elemental pause arises when EC encounters a special DNA sequencing encoding a pause signal and enters an offline state, in which the PEC is neither backtracked nor stabilized by an RNA hairpin. It remains unclear why elemental pause signals show variable duration. In a simple model, the elemental pause complex adopts a single conformation, and the pause duration is determined by the translocation difficultness dictated by the sequence context. However, multiple conformations could explain differences in pause strength.

By comparing two ePECs reconstituted on different nucleic acid scaffolds, we probed the conformational properties of them with two biochemical reporting systems. First, the CTR system focused on the SI3 positions, an indicator of RNAP swiveling state and TL folding state. The formation of a con-ePEC significantly induced TL folding, whereas the formation of a *his*-ePEC induced swiveling, typical for a hairpin-stabilized paused complex. Second, we tested the translocation state of the two ePECs, by utilizing PPi-induced cleavage assay and NTP-induced nucleotide incorporation assay. The *con*-ePEC's resistance to NTP addition and ability to undergo NTP addition is consistent with the pre-translocated register, while the opposite characteristics of the *his*-ePEC support a half-translocated register. These biochemical results matched well with the published cryo-EM results obtained using the exact same scaffolds of the two ePECs. Collectively, these results are supportive of a model in which the elemental pausing state involves multiple pause species with distinct conformational properties.

A.5 Materials and methods

*Eco*RNAP expression and purification

*Eco*RNAP was prepared as described previously (11, 15). Briefly, plasmid pRM756 that contains *rpoA* (α), *rpoB* (β), *rpoC* (β′) with C-terminal deca-histidine tag and *rpoZ* (ω) was expressed in BL21(DE3) (Novagen). The cells were grown at 37 ºC in LB broth in the presence of 60 ug kanamycin/mL. Protein expression was induced at apparent OD_{600} of 0.6-0.8 with 1 mM isopropyl β-D-thigalactopyranoside (IPTG) for 4 h. Cells were harvested, resuspended in lysis buffer (50 mM Tris pH 8.0, 5% glycerol, 1 mM EDTA (pH 8.0), 1 mM $ZnCl_2$, 10 mM DTT, protease inhibitor cocktail), and lysed by sonication at 4 ºC. The lysate was precipitated by adding polyethyleneimine (PEI, 10% (w/v), Sigma Aldrich) to a final concentration of 0.6% (w/v) dropwise. The pellets were washed three times with wash buffer containing TGED (10 mM Tris-HCl, pH 8.0, 5% glycerol, 0.1% EDTA pH 8.0, 10 mM DTT) + 0.5M NaCl, and the RNAP was eluted from the pellet with elution buffer (TGED + 1 M NaCl). The eluted RNAP was precipitated by adding ammonium sulfate (35 g per 100 ml eluted protein solution), redissolved in chelating buffer (20 mM Tris pH 8.0, 1 M NaCl, 5% glycerol), and loaded onto a HisTrap HP column (Cytiva) for purification by Ni^{2+}-affinity chromatography using an imidazole gradient. The recovered RNAP was dialyzed in TGED + 100 mM NaCl buffer and loaded onto a Heparin FF column (Cytiva) for affinity chromatography using 0.5 M NaCl. The eluted RNAP by was concentrated by Amicon Ultra centrifugal filter (Merck Millipore), dialyzed into RNAP Storage buffer (10 mM Tris-HCl, 25% v/v glycerol, 100 mM NaCl, 100 µM EDTA, 1 mM $MgCl_2$, 20 µM $ZnCl_2$, 10 mM DTT) and stored at –80 ºC until use.

In vitro transcription pause assays with wild-type or Cys-pair reporter RNAPs

CPR RNAPs were prepared as described previously (9). The transcription assay scaffold was made by mixing 5 µM RNA with 10 µM template strand DNA (t-DNA) in Reconstitution Buffer (10 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$, 40 mM KCl) or cryo-EM buffer (20 mM Tris-Cl, pH 8.0, 150 mM potassium glutamate, 5 mM $MgCl_2$, 5 mM DTT) followed by heat-denaturing and gradual cooling in thermal cycler. Elongation complex (EC) was formed by mixing scaffold with CPR or ΔαCTD RNAP at 1: 3 ratio in Elongation Buffer (25 mM HEPES-KOH, pH 8.0, 130 mM KCl, 5 mM $MgCl_2$, 0.15 mM EDTA, 5% v/v glycerol, 25 µg acetylated BSA/mL) or cryo-EM buffer above for 15

min at 37 °C or 23 ºC followed by adding non-template strand DNA (nt-DNA) for 15 min (1 μM RNA, 2 μM T-DNA, 3 μM RNAP, 5 μM NT-DNA). CHAPSO was added to a final concentration of 8 mM to ECs reconstituted in cryo-EM buffer. Free RNAP was blocked by adding 0.1 mg heparin/mL. To induce disulfide bond formation, 5 mM H_2O_2 was added for 15 min at 37 °C followed by adding 0.1 U/μL catalase for 5 min at room temperature to remove access H_2O_2.

For *con*-ePEC pause assay with varying delay times of GTP addition (Fig. S2), G16 RNA was 5′-^{32}P labeled prior to scaffold annealing. Following *con*-ePEC$_{-1}$ assembly at 23 °C (mimicking cryo-EM sample preparation conditions), complexes were diluted to a final concentration of 50 nM based on RNA with cryo-EM buffer and reacted with CTP (100 μM final) for the indicated amount of time at 23 °C, to position RNAP at the pause (Fig. S2). Escape from the pause was then initiated by addition of limiting GTP (10 μM) at 23 °C and followed by withdrawing aliquots at various time points.

For *con*-ePEC pause assay using CPR RNAP, reconstituted EC was adjusted to 200 nM with EB and isotope-labeled with [α-^{32}P] GTP (33 Ci/mmol) to reach G16 pause –1 register. Transcription was initiated by adding equal volume of GTP and CTP mixture (100 μM final) and incubated at 37 °C. 5 μL of reaction samples were taken at different time points over 2 min and mixed with 5 μL of Stop Buffer (8 M urea, 50 mM EDTA, 90 mM Tris-borate buffer, pH 8.3, 0.02% bromophenol blue, 0.02% xylene cyanol) to quench the reaction. RNAs were analyzed by denaturing PAGE (15% 19:1 acrylamide: bis-acrylamide, 45 mM Tris-borate, pH=8.3, 1.25 mM Na_2EDTA, 8 M urea). For *his*-ePEC/*his*PEC pause assay, reconstituted EC was adjusted to 200 nM with EB, isotope-labeled with [α-^{32}P]CTP (33 Ci/mmol), and extended to U20 pause register with 2 μM UTP. To mimic pause RNA hairpin, 2 μM antisense RNA oligo was added for 10 min at 37 °C. Transcription was initiated by adding equal volume of GTP mixture (10 μM final) and incubating at 37 °C. For non-CPR transcription assays, WT RNAP was used and crosslinking steps were omitted.

Cys-triplet reporter (CTR) crosslinking assay

The CTR RNAP was prepared as described previously (9). The CTR EC was reconstituted similarly to the transcription assay with a few adjustments in component ratio (5 μM RNA, 10 μM t-DNA, 2.5 μM CTR, 10 μM nt-

DNA). To induce disulfide bond formation, 5 mM H_2O_2 was added for 15 min at 37 °C followed by addition of 0.1 U catalase/μL for 5 min at room temperature to remove access H_2O_2. NTPs (50 μM each) were used for extension to relevant EC/PEC registers and 25 μM antisense RNA oligo was used to mimic the pause RNA hairpin. Samples were mixed with 4× LDS Sample Buffer (ThermoFisher) and analyzed on a thin layer 4-15% gradient polyacrylamide gel (GE Healthcare) using a PhastSystem Electrophoresis Unit (Pharmacia). The intensities of β–β′ and β′–β′ species were gel-quantified and used for SPB calculation: SPB = (β′–β′%) / (β–β′%).

Translocation register assay

EC scaffolds were prepared as described above with 5′ end ^{32}P isotope labeled RNAs. The ECs were reconstituted by mixing scaffold with β′-His10 tagged WT RNAP at 1:1 ratio in cryo-EM PEC Buffer (20 mM Tris-HCl, pH=8.0, 150 mM potassium glutamate, 5 mM $MgCl_2$) for 15 min at 25 °C followed by adding non-template strand DNA (NT-DNA) for 10 min (2 μM RNA, 4 μM T-DNA, 2 μM RNAP, 10 μM NT-DNA). Both *con*-ePEC and *his*-ePEC were directly reconstituted at –1 pause register. Reconstituted ECs were adjusted to 200 nM with buffer and incubated with 8 mM CHAPSO (to mimic cryo-EM sample preparation condition) and 0.1 mg/mL heparin. ECs were immobilized on Ni^{2+} NTA magnetic beads (Dynabeads, ThermoFisher) and incubated with 200 μM CTP or UTP for 1 min to make con-ePEC and *his*-ePEC respectively. To form *his*PEC, 2 μM antisense RNA oligo was added to *his*-ePEC for 5 min. To trap translocation states of PECs, 15 μM coliphage HK022 Nun protein (16)was added for 5 min incubation. The supernatant was then removed and bead-tethered PECs were washed twice with buffer. To elute PECs, beads were incubated with cryo-EM PEC buffer containing 150 μM imidazole for 10 min at 25 °C. PECs were then incubated with either 2.5 μM pyrophosphate for 5 min to test pyrophosphorolysis efficiency or 1 mM NTPs for 2 min to test nucleotide addition efficiency. Samples were mixed with equal volume of Stop buffer for PAGE analysis.

A.6 References

1. R. Landick, Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol* **75**, 291-314 (2021).

2. S. Gong, Y. Wang, Z. Wang, W. Zhang, Co-Transcriptional Folding and Regulation Mechanisms of Riboswitches. *Molecules* **22** (2017).

3. O. Laptenko, J. Lee, I. Lomakin, S. Borukhov, Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J* **22**, 6322-6334 (2003).

4. R. Kohler, R. A. Mooney, D. J. Mills, R. Landick, P. Cramer, Architecture of a transcribing-translating expressome. *Science* **356**, 194-197 (2017).

5. L. You *et al.*, Structural basis for intrinsic transcription termination. *Nature* **613**, 783-789 (2023).

6. C. Mora Gallardo, A. Sanchez de Diego, A. C. Martinez, K. H. M. van Wely, Interplay between splicing and transcriptional pausing exerts genome-wide control over alternative polyadenylation. *Transcription* **12**, 55-71 (2021).

7. J. Y. Kang *et al.*, RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* **69**, 802-815 e801 (2018).

8. J. Y. Kang *et al.*, Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e1614 (2018).

9. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2101805118 (2021).

10. C. Chan, D. Wang, R. Landick, Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J. Mol. Biol.* **268**, 54-68 (1997).

11. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proc Natl Acad Sci U S A* **118** (2021).

12. C. L. Vitiello, M. L. Kireeva, L. Lubkowska, M. Kashlev, M. Gottesman, Coliphage HK022 Nun protein inhibits RNA polymerase translocation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2368-2375 (2014).

13. J. Y. Kang *et al.*, Structural basis of transcription arrest by coliphage HK022 nun in an Escherichia coli RNA polymerase elongation complex. *eLife* **6**, e25478 (2017).

14. J. Y. Kang *et al.*, An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. *Proc Natl Acad Sci U S A* **120**, e2215945120 (2023).

15. I. Artsimovitch, V. Svetlov, K. S. Murakami, R. Landick, Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* **278**, 12344-12355 (2003).

16. J. Y. Kang *et al.*, Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife* **6** (2017).

Appendix B: Alarmone ppGpp strengthens elemental pause via binding site 1

Yu Bao's contribution:

In vitro experiments of pause assays with ppGpp

B.1 Abstract

In many bacterial species, the alarmone (p)ppGpp serves as a key signaling molecule in the stringent response, facilitating the regulation of gene expression during conditions of nutrient stress. (p)ppGpp exerts its influence on transcriptional processes by directly binding to two distinct sites on RNA polymerase (RNAP), thereby inducing allosteric modifications in transcription initiation efficiency. Although previous *in vitro* studies have suggested that (p)ppGpp may also impact transcription elongation rate, the precise mechanism behind remains inadequately understood. It is conceivable that ppGpp could interfere with transcription elongation by means such as substrate competition, stimulation of pausing, or modulation of RNAP through unidentified factors. To investigate the mechanisms underlying (p)ppGpp's regulation of transcriptional pausing, we conducted *in vitro* experiments to assess the effects of ppGpp on various pause complexes. Our findings demonstrate that (p)ppGpp directly stabilizes a consensus elemental paused complex by biasing the complex towards a swiveling conformation. Furthermore, we have established that (p)ppGpp's pause-stabilizing effects are mediated specifically through its binding to the site 1 of RNAP, independent of substrate competition. The insights provided by our results establish a novel framework for investigating the allosteric effects of (p)ppGpp on transcriptional elongation and pausing.

B.2 Introduction

Bacterial transcription regulation by the signal molecule (p)ppGpp was first discovered fifty years ago (1). The so-called alarmone, (p)ppGpp, is synthesized by RelA/SpoT in E. coli in response to stressed conditions (2), in which the elevated level of uncharged t-RNAs bound to the A-site of the ribosome, thereby stabilizing RelA association with the ribosome and triggering its (p)ppGpp synthesizing activities. pppGpp is the primary product of RelA catalysis, which is then converted into ppGpp by GppA hydrolase. pppGpp and ppGpp are similar in terms of their regulatory functions, with ppGpp being more potent (3). (p)ppGpp binds RNAP via two characterized sites, the binding site 1 is at the interface between β′ and ω subunits, while the binding site 2 is formed by RH domain and a dissociable factor DskA (3-7). Site 1 and site 2 are both involved in the full inhibitory effect of gene expression, while site 2 alone is predominately responsible for positive regulation of certain genes. It has been proposed that the different kinetics of the transcription initiation stages determined the effect of (p)ppGpp (8).

ppGpp's inhibitory effects on transcription elongation were observed over forty years ago. ppGpp increases the dwell time of most transcriptional pauses by a factor of 2–3 via direct RNAP contacts (9). Because ppGpp enhances pausing in vitro without DksA and does not synergize with DksA to affect elongation in vitro (9-12), site 1 is the likely but as yet untested target for ppGpp allosteric stimulation of pausing. ppGpp could stimulate different steps in pausing. Pausing occurs by an initial, sequence-dependent isomerization of RNAP into an elemental paused elongation complex (ePEC) (13-16). The ePEC enters more long-lived pause states by backtracking (reverse translocation) of RNA and DNA through RNAP, by binding of regulators, or by formation of a pause RNA hairpin in the RNA exit channel of RNAP. The ePEC itself can assume multiple conformations. In some, translocation of the RNA forms a half-translocated (tilted) RNA–DNA hybrid. DNA translocation is inhibited in all ePEC states. Rotation of a portion of the flexible, multidomain RNAP termed the swivel module (e.g., the clamp, shelf, and jaw) may further inhibit both DNA translocation and an active-site trigger loop (TL)–trigger helices (TH) transition required for catalysis. ppGpp is proposed to promote backtracking (17), which could explain its effect on pausing. Alternatively, ppGpp could affect swiveling or translocation in the ePEC.

To elucidate how ppGpp controls pausing by RNAP in bacteria such as *E. coli* where direct ppGpp–RNAP interactions occur, we utilized *in vitro* transcription assays to explore pause enhancing effects on various DNA

context and found an optimum scaffold that recapitulates ppGpp's allosteric effects. We also revealed the swiveling conformational change of a ppGpp-bound ePEC with the Cys-triplet reporter system.

B.3 Results and discussion

B.3.1 ppGpp strengthens the consensus elemental pause *in vitro.*

To investigate how ppGpp affects transcriptional pausing, we first studied the well-studies consensus elemental pause signal. The consensus elemental pause sequence was generated from global pause signals in NET-seq screening, featured by a sequence motif of 5′-$G_{-11}G_{-10}t_{-3}g_{-2}Y_{-1}G_{+1}$ (–1 corresponds to 3′ end position of nascent RNA) (18). We cloned the consensus elemental pause signal into a ~300 bp linear DNA transcription template controlled by a λP_R promoter and tested the pause kinetics with/without 100 μM in the presence of 100 μM NTPs. The *in vitro* transcription kinetics analysis showed a significant of pause enhancement by a factor of ~2.6 with ppGpp addition (**Figure B.1**). However, when replacing the DNA template into a nucleic acid scaffold to reperform the pause assay, the effect of ppGpp is weakened to 1.4× fold change (**Figure B.2**). We hypothesized that the weaker effect of ppGpp arose from the unnatural configuration of the nucleic acids scaffold distinct from that of the template DNA, since (a) the poly-U sequence at the 5′ end of the RNA cannot base pair with the upstream t-DNA, which does not allow backward translocation of the PEC; (b) the poly-U sequence does not allow RNA secondary structure to form at the RNA exit channel and (c) the EC is halted at pause –1 position (G36 EC) prior to the pause formation (C37 ePEC) and pause escape, probably causing some conformational changes of the EC that rarely occur in continuous transcription. To address this problem, we redesigned the scaffold, to let EC reconstitute at an earlier position G17 EC) and let the EC to elongation over a C-less stretch until reaching pause –1 position with an incoming CTP (**Figure B.2**). At this point, the halted G36 EC is incubated with an antisense DNA oligo to form RNA-DNA duplex that prevents the backtracking and intrinsic cleavage of the G36 EC. This RNA-DNA duplex forms at the –14 position to avoid having pause-stabilizing effects. In this framework, we successfully recapitulated the pause enhancing effects of 100 μM ppGpp with 1.9× fold change. In the same experimental condition, pppGpp shows a weaker effect than ppGpp (~1.5× fold enhancement). We conclude that ppGpp can specifically enhance the elemental pause *in vitro*.

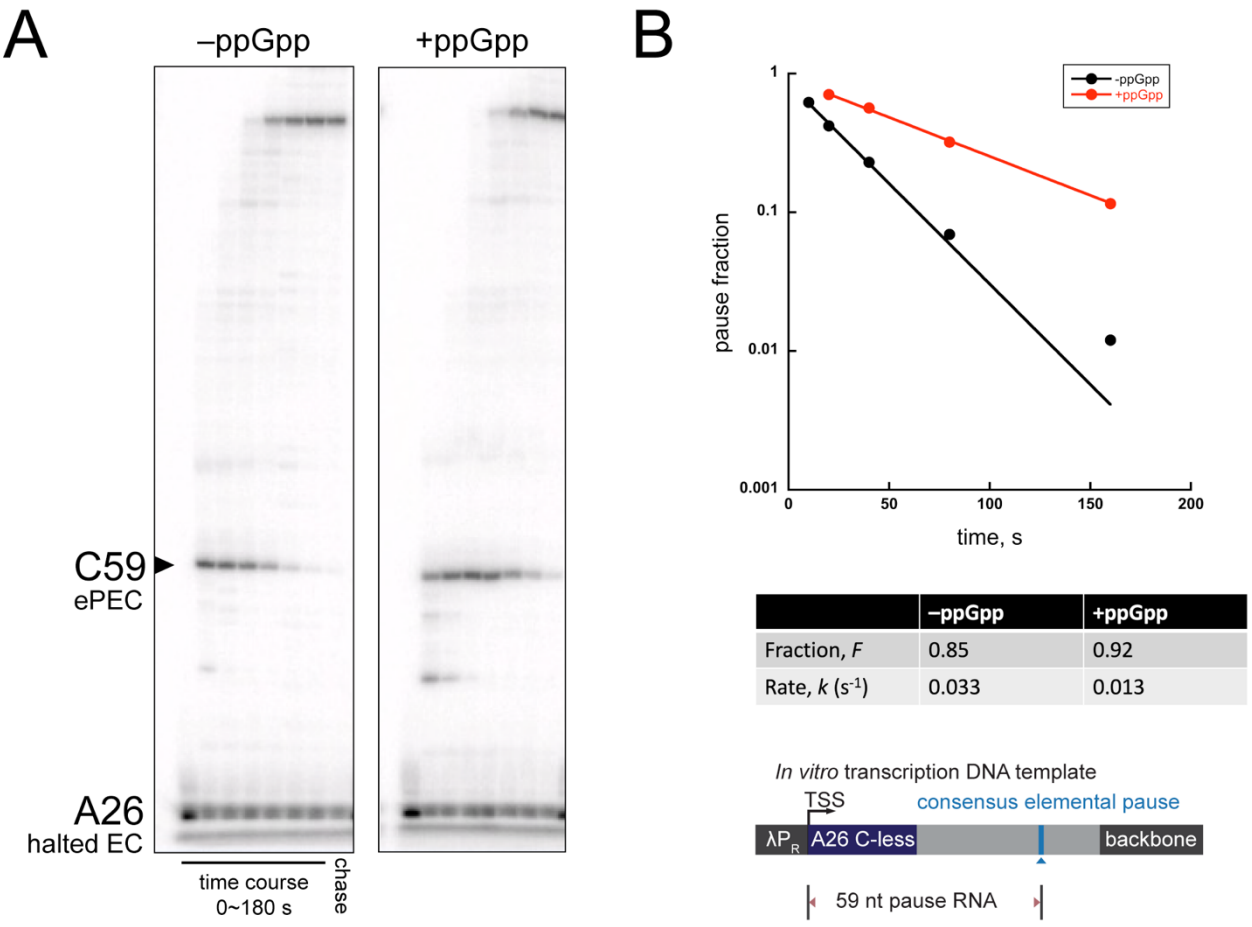
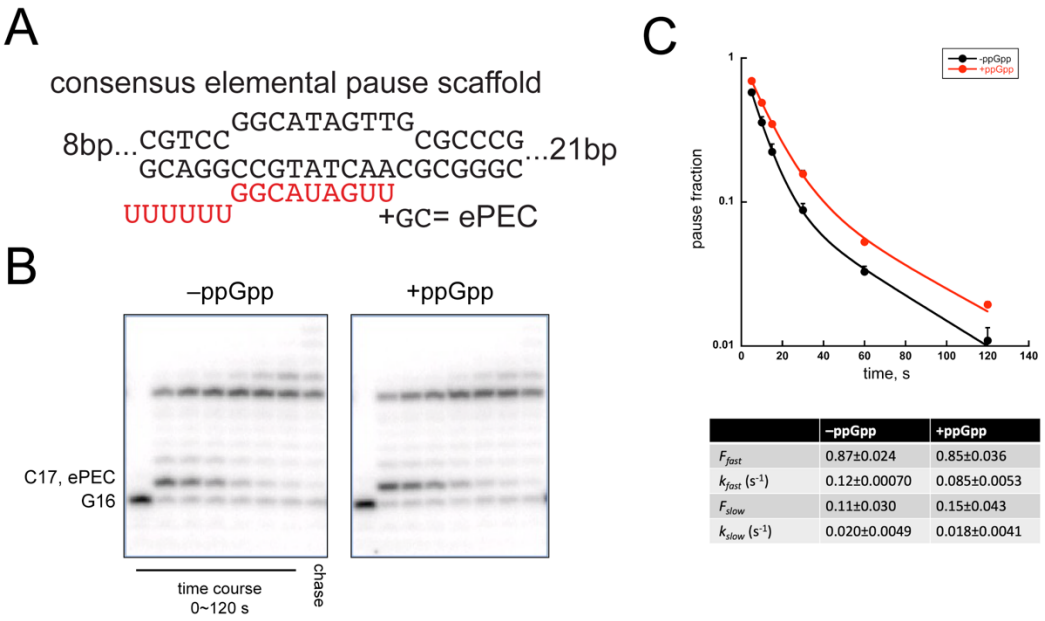


Figure B.1. ppGpp enhances consensus elemental pause *in vitro.*

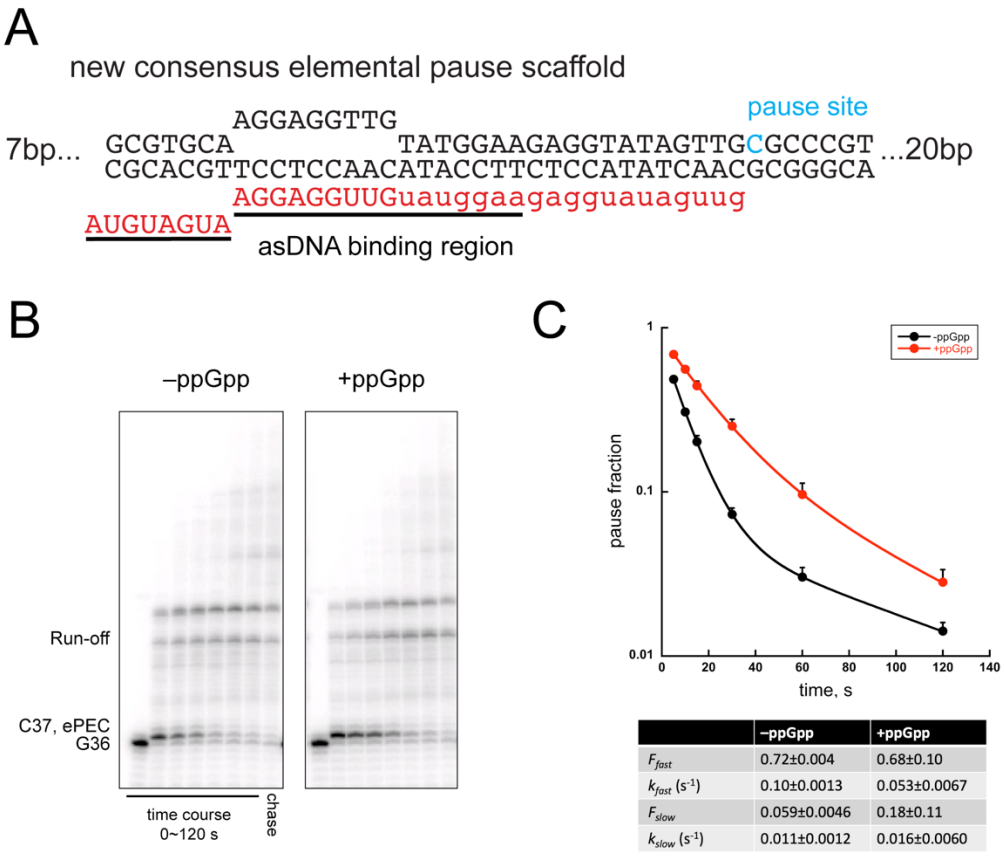
(A) The gel image of the *in vitro* transcription assay of a consensus elemental pause signal in a DNA template. ppGpp is used at 100 μM as the final concentration. (B) The pause kinetics of the consensus elemental pause with/without ppGpp. The design of the transcription template is shown at the bottom.



	–ppGpp	+ppGpp
F_{fast}	0.87±0.024	0.85±0.036
k_{fast} (s^{-1})	0.12±0.00070	0.085±0.0053
F_{slow}	0.11±0.030	0.15±0.043
k_{slow} (s^{-1})	0.020±0.0049	0.018±0.0041

Figure B.2. ppGpp slightly enhances consensus elemental pause reconstituted with a scaffold.

(A) The nucleic acids scaffold to reconstitute the consensus elemental paused complex. (B) The gel image of the pause assay with the scaffold in (A). (C) The pause kinetics comparison of the pause assay with/without ppGpp addition.



	–ppGpp	+ppGpp
F_{fast}	0.72±0.004	0.68±0.10
k_{fast} (s^{-1})	0.10±0.0013	0.053±0.0067
F_{slow}	0.059±0.0046	0.18±0.11
k_{slow} (s^{-1})	0.011±0.0012	0.016±0.0060

Figure B.3. ppGpp significantly enhances consensus elemental pause reconstituted with a new scaffold.

(A) The new nucleic acids scaffold to reconstitute the consensus elemental paused complex. The capital letters of the RNA strand indicate the RNA used when assembling the scaffold. (B) The gel image of the pause assay with the scaffold in (A). (C) The pause kinetics comparison of the pause assay with/without ppGpp addition.

B.3.2 ppGpp strengthens the consensus elemental pause with modifications in dsFJ sequence element.

The upFJ sequence element, including the base (normally a pyrimidine, C or U) at –1 position and the incoming base (normally a guanine, while an adenine can also contribute to pause *in vitro*) at +1 position, plays critical roles of elemental pause formation. Previously studies show that the shuffling of the bases at this position greatly reduced the strength of elemental pause. We replaced the –1 C in the scaffold into a U, and found that ppGpp effects are barely changed (**Figure B.4**). We also replaced the the +1 G into A/T/C and as expected, in cases of the +1 T/C, pause strength was greatly reduced while the +1 A maintained pause strength (Figure). Interestingly, the ppGpp promoters pausing in all these 4 templates, regardless of the actual pause strength. We speculate that the ppGpp effects on pausing is not specific to a single pause sequence. The shortage of this set of experiments is lack of replicates to perform rigorous statistically analysis, therefore determine if there is any synergy between ppGpp effects and the strength of the pause sequence.

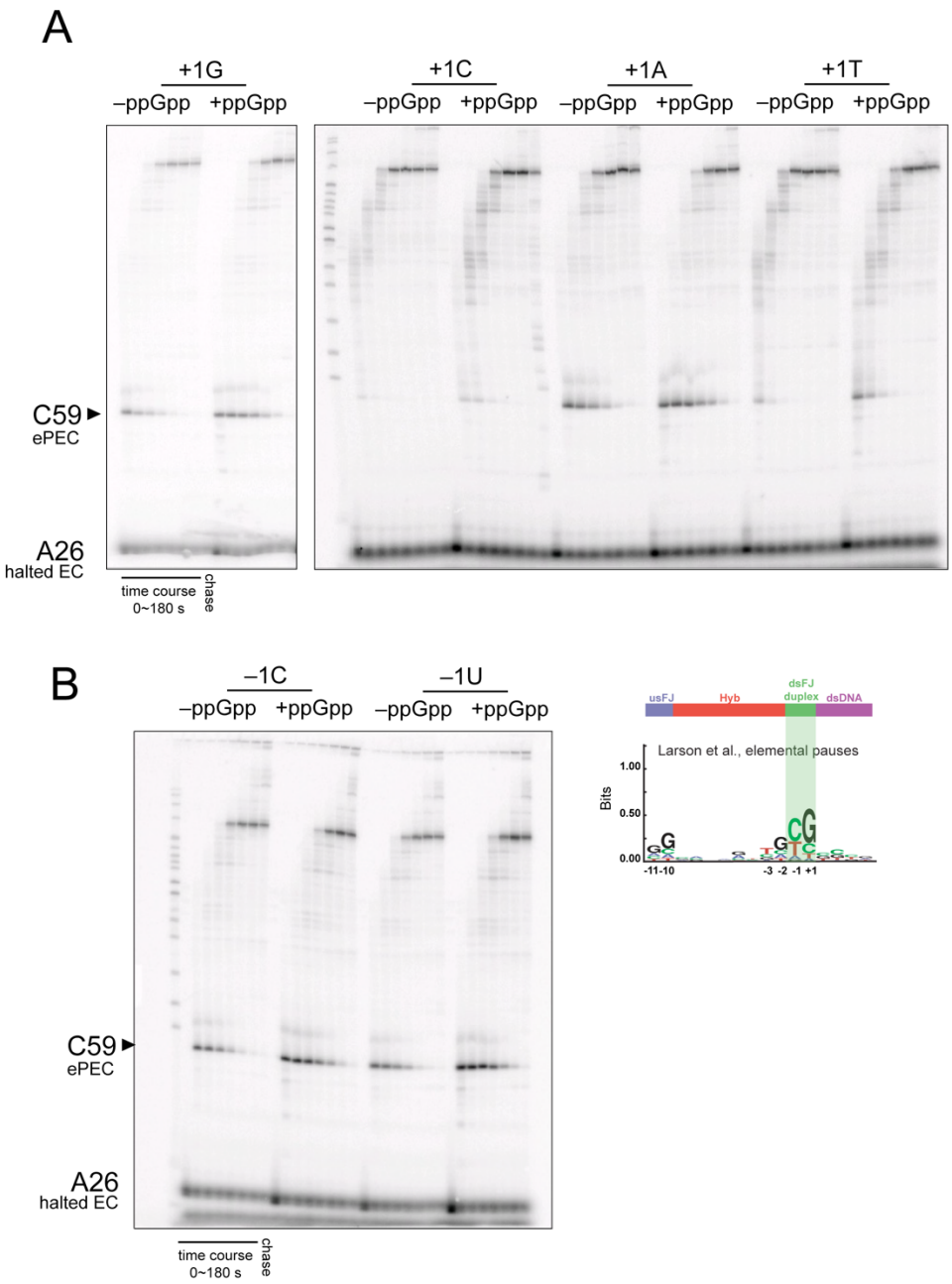


Figure B.4. ppGpp strengthens the consensus elemental pause with modifications in dsFJ sequence element.

(A) DNA template-based transcription assay with +1 base alteration. (B) Left, DNA template-based transcription assay with –1 base alteration. Right, the position of dsFJ in the sequence logo of the consensus elemental pause.

B.3.3 ppGpp's pause-enhancing effects is based on inducing RNAP swiveling.

Based on studies on elemental pause, two major components exist in the elemental pause complex: a half-translocated register with swiveling and a pre-translocated register with TL folded. Stabilizing either of these components will promoter pause strength. To tell which conformational change ppGpp induced in consensus elemental pause, we probed the SI3 positions with the established Cys-triplet reporter. The SPB index, calculated by the amount of SI3-closed crosslinking species, divided by the amount of SI3-swiveled crosslinking species, denotes the extent of swiveling in EC/PEC. We observed significantly SPB decrease upon 100 μM ppGpp addition to the ePEC, indicating SI3 domain is more biased to the swiveling position when ppGpp binds ePEC. To tell if the ppGpp-induced swiveling is directly related to the pause enhancement, we compared the pause kinetics between ePECs formed by WT RNAP and β′F1199A RNAP, a mutant RNAP that has reduced SI3 swiveling due to the disruption of SI3-jaw interactions. For WT RNAP, the ePEC estimated half-life is increased from 40 s to 142 s (3.6×), in contrast, for β′F1199A RNAP, the ePEC estimated half-life is increased from 30 s to 50 s (1.7×) (**Figure B.5**). The significant decrease of pause enhancement of ppGpp for the mutant RNAP is a supportive piece of evidence that the swiveling is contributing to ppGpp's pause enhancing effects.

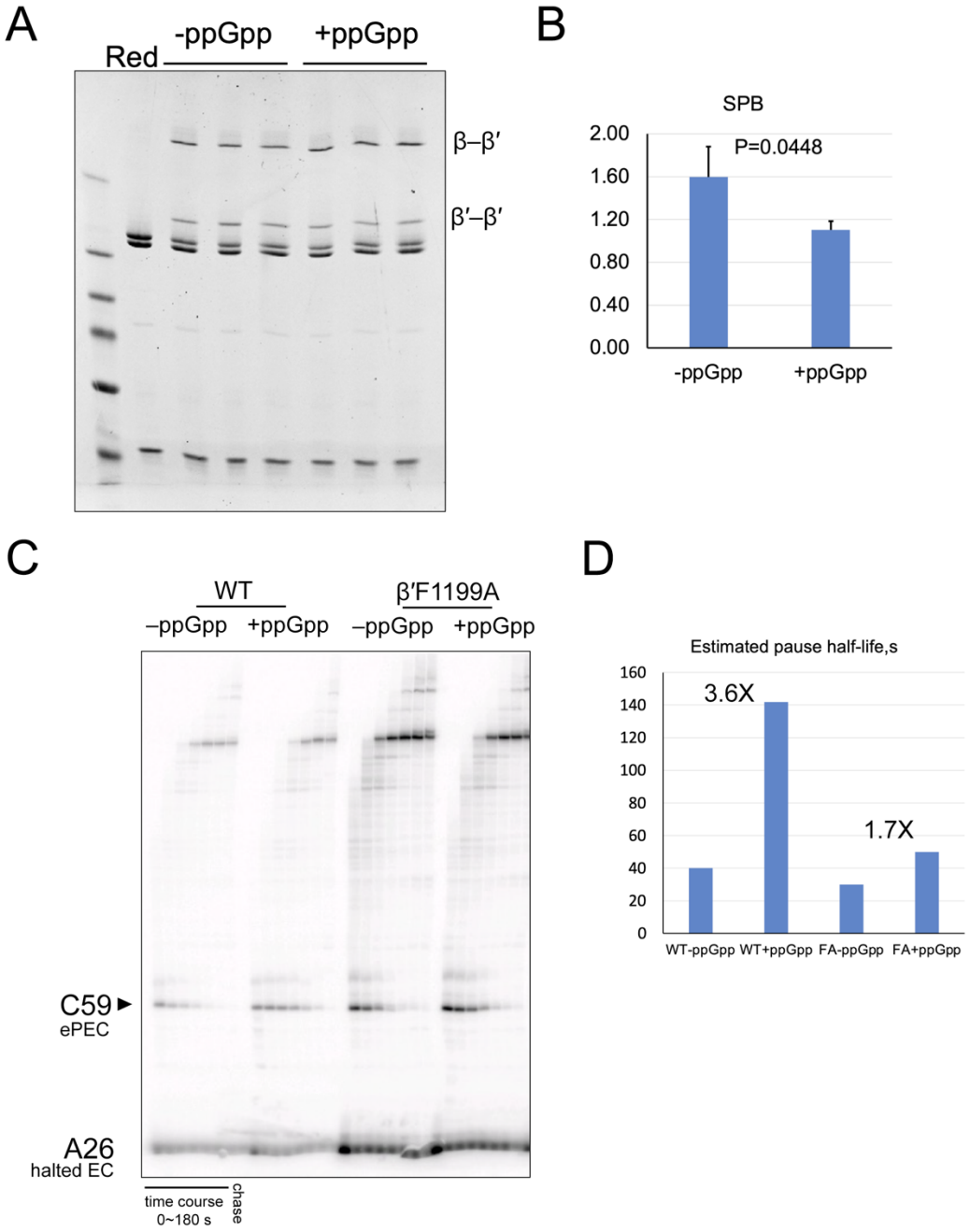

A
-ppGpp
+ppGpp
Red
β–β′
β′–β′
B
SPB
P=0.0448
2.00
1.60
1.20
0.80
0.40
0.00
-ppGpp
+ppGpp
C
WT
β′F1199A
–ppGpp
+ppGpp
–ppGpp
+ppGpp
C59
ePEC
A26
halted EC
time course
0~180 s
chase
D
Estimated pause half-life,s
160
140
120
100
80
60
40
20
0
3.6X
1.7X
WT-ppGpp
WT+ppGpp
FA-ppGpp
FA+ppGpp

Figure B.4. ppGpp-induced RNAP swiveling caused pause stabilization.

(A) CTR crosslink assay with ppGpp, the scaffold is the same as that in Figure B.3.A.(B) SPB calculation of the PAGE. ppGpp significantly reduced SPB value, indicative of more swiveling of the PEC. P value is from two-tailed t-test based on the technical triplicates. (C) *In vitro* transcription with two RNAPs and ppGpp treatment. (D) Estimated pause half-life in the transcription assay. FA is the short form of β′F1199A RNAP.

B.4 Materials and methods

Cys-triplet reporter (CTR) crosslinking assay

The CTR RNAP was prepared as described previously (19). The CTR EC was reconstituted similarly to the transcription assay with a few adjustments in component ratio (5 μM RNA, 10 μM t-DNA, 2.5 μM CTR, 10 μM nt-DNA). To induce disulfide bond formation, 5 mM H_2O_2 was added for 15 min at 37 °C followed by adding 0.1 U catalase/μL for 5 min at room temperature to remove access H_2O_2. NTPs (50 μM each) were used for extension to relevant EC/PEC registers and 25 μM antisense RNA oligo was used to mimic pause RNA hairpin. Samples were mixed with 4× LDS Sample Buffer (ThermoFisher) and analyzed on a thin layer 4-15% gradient polyacrylamide gel (GE Healthcare) using a PhastSystem Electrophoresis Unit (Pharmacia). The intensities of β–β′ and β′–β′ species were gel-quantified and used for SPB calculation: SPB = (β′–β′%) / (β–β′%).

DNA template-based *in vitro* transcription assay

The DNA template was prepared by PCR amplification from a plasmid harboring the λP_R promoter, the C-less cassette and the DNA fragment containing the modified consensus elemental pause signal and purified by spermine precipitation. The holoenzyme was reconstituted by mixing 5 μM σ70 with 1 μM RNAP core enzyme at 37 °C for 20 min. To form the A26 halted complex, mix the following components and incubate at 37 °C for 15 min: 50 nM holoenzyme, 40 nM DNA template, 0.2 Unit/μL RNasin ribonuclease inhibitor (Promega), 150 μM ApU dinucleotide, 0.1 μCi/μL [α-^{32}P] GTP, 2 μM GTP, 5 μL ATP and 5 μM UTP in 1× Elongation Buffer (10 mM HEPES, pH = 7.9, 0.1 mM EDTA, 50 mM potassium glutamate, 10 mM magnesium acetate, 5 μg/mL acetylated BSA and 1 mM DTT). Afterwards, 0.1 mg heparin/mL was added to bind free RNAPs at 25 °C for 20 min. To resume transcription, equal volume of halted complex is mixed with 200 μM rNTPs at 37 °C. Samples of 5 μL at various time points were withdrawn from the mixture and mixed with 5μL 2× Stop Buffer (8 M urea, 50 mM EDTA, 90 mM Tris-borate buffer, pH 8.3, 0.02% bromophenol blue, 0.02% xylene cyanol). RNAs were analyzed by denaturing PAGE (8% 19:1 acrylamide: bis-acrylamide, 45 mM Tris-borate, pH=8.3, 1.25 mM Na_2EDTA, 8 M urea).

DNA scaffold-based *in vitro* transcription assay

The transcription assay scaffold was made by mixing 5 μM RNA with 10 μM template strand DNA (t-DNA) in Reconstitution Buffer (10 mM Tris-HCl, pH 7.9, 5 mM $MgCl_2$, 40 mM KCl) or cryo-EM buffer (20 mM Tris-Cl, pH 8.0, 150 mM potassium glutamate, 5 mM $MgCl_2$, 5 mM DTT) followed by heat-denaturing and gradual cooling in thermal cycler. Elongation complex (EC) was formed by mixing scaffold with RNAP at 1: 3 ratio in Elongation Buffer (25 mM HEPES-KOH, pH 8.0, 130 mM KCl, 5 mM $MgCl_2$, 0.15 mM EDTA, 5% v/v glycerol, 25 μg acetylated BSA/mL) for 15 min at 37 °C or 23 ºC followed by adding non-template strand DNA (nt-DNA) for 15 min (1 μM RNA, 2 μM T-DNA, 3 μM RNAP, 5 μM NT-DNA). Free RNAP was blocked by adding 0.1 mg heparin/mL.

For *con*-ePEC pause assay, reconstituted EC was adjusted to 200 nM with EB and isotope-labeled with [α-^{32}P] GTP (33 Ci/mmol) to reach G16 pause –1 register. Transcription was initiated by adding equal volume of GTP and CTP mixture (100 μM final) and incubated at 37 °C. 5 μL of reaction samples were taken at different time points over 2 min and mixed with 5 μL of Stop Buffer (8 M urea, 50 mM EDTA, 90 mM Tris-borate buffer, pH 8.3, 0.02% bromophenol blue, 0.02% xylene cyanol) to quench the reaction. RNAs were analyzed by denaturing PAGE (15% 19:1 acrylamide: bis-acrylamide, 45 mM Tris-borate, pH=8.3, 1.25 mM Na_2EDTA, 8 M urea).

Table of oligonucleotides, plasmids and strains

Oligonucleotides (5′ to 3′)	Assay	Stock #
CGTTAAATCTATCACCGCAAGGG	PCR	#3071
CAGTTCCCTACTCTCGCATG	PCR	#645
pRM1002, plasmid harboring *in vitro* transcription template λP_R-A26-consensus elemental pause (+1 G)	PCR template	#5302

pYB412, plasmid harboring *in vitro* transcription template λP_R-A26-consensus elemental pause (+1 C)	PCR template	#6392
pYB413, plasmid harboring *in vitro* transcription template λP_R-A26-consensus elemental pause (+1 A)	PCR template	#6393
pYB414, plasmid harboring *in vitro* transcription template λP_R-A26-consensus elemental pause (+1 T)	PCR template	#6394
pYB415, plasmid harboring *in vitro* transcription template λP_R-A26-consensus elemental pause (–1 U)	PCR template	#6395
hp-cePEC template DNA +1G CGGTCTGAACCGGTGAATTTACGGGCGCAACTATACCTCTTCCATACAAC CTCCTTGCACGC	*In vitro* transcription	# 14953
hp-cePEC non-template DNA +1G GCGTGCAAGGAGGTTGTATGGAAGAGGTATAGTTGCGCCCGTAAATTCAC CGGTTCAGACCG	*In vitro* transcription	# 14952
hp-cePEC RNA GCGGUCUAGUCUUACCGCGAGGUAUAGUUG	*In vitro* transcription	# 15289
hp-cePEC template DNA +1A CGGTCTGAACCGGTGAATTTACGGGTGCAACTATACCTCTTCCATACAAC CTCCTTGCACGC	*In vitro* transcription	# 15191
hp-cePEC non-template DNA +1A GCGTGCAAGGAGGTTGTATGGAAGAGGTATAGTTGCACCCGTAAATTCAC CGGTTCAGACCG	*In vitro* transcription	# 15190

B.5 References

1. A. Travers, Modulation of RNA polymerase specificity by ppGpp. *Molecular & general genetics : MGG* **147**, 225-232 (1976).

2. A. B. Loveland *et al.*, Ribosome*RelA structures reveal the mechanism of stringent response activation. *eLife* **5**, e17029 (2016).

3. U. Mechold, K. Potrykus, H. Murphy, K. S. Murakami, M. Cashel, Differential regulation by ppGpp versus pppGpp in Escherichia coli. *Nucleic acids research* **41**, 6175-6189 (2013).

4. W. Ross, C. E. Vrentas, P. Sanchez-Vazquez, T. Gaal, R. L. Gourse, The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Molecular cell* **50**, 420-429 (2013).

5. W. Ross *et al.*, ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. *Mol Cell* **62**, 811-823 (2016).

6. V. Molodtsov *et al.*, Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol Cell* **69**, 828-839 e825 (2018).

7. Y. Zuo, Y. Wang, T. A. Steitz, The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Molecular cell* **50**, 430-436 (2013).

8. R. L. Gourse *et al.*, Transcriptional Responses to ppGpp and DksA. *Annu Rev Microbiol* **72**, 163-184 (2018).

9. M. Krohn, R. Wagner, Transcriptional pausing of RNA polymerase in the presence of guanosine tetraphosphate depends on the promoter and gene sequence. *The Journal of biological chemistry* **271**, 23884-23894 (1996).

10. R. E. Kingston, W. C. Nierman, M. J. Chamberlin, A direct effect of guanosine tetraphosphate on pausing of Escherichia coli RNA polymerase during RNA chain elongation. *The Journal of biological chemistry* **256**, 2787-2797 (1981).

11. R. E. Kingston, M. J. Chamberlin, Pausing and attenuation of in vitro transcription in the rrnB operon of E. coli. *Cell* **27**, 523-531 (1981).

12. M. Roghanian, N. Zenkin, Y. Yuzenkova, Bacterial global regulators DksA/ppGpp increase fidelity of transcription. *Nucleic Acids Res* **43**, 1529-1536 (2015).

13. X. Guo *et al.*, Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e814 (2018).

14. J. Y. Kang *et al.*, RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Molecular cell* **69**, 802-815.e805 (2018).

15. R. Landick, Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol* **75**, 291-314 (2021).

16. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *eLife* **8**, e40981 (2019).

17. V. Kamarthapu *et al.*, ppGpp couples transcription to DNA repair in E. coli. *Science* **352**, 993-996 (2016).

18. M. H. Larson *et al.*, A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).

19. Y. Bao, R. Landick, Obligate movements of an active site-linked surface domain control RNA polymerase elongation and pausing via a Phe pocket anchor. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2101805118 (2021).

Acknowledgement

I would like to express my deepest gratitude to all those who have supported me throughout my journey in pursuing my Ph.D. in Biochemistry. This thesis would not have been possible without the unwavering guidance, encouragement, and assistance of numerous individuals and institutions.

First and foremost, I am immensely grateful to my supervisor, Professor Robert Landick, for his invaluable guidance, expertise, and continuous support throughout the duration of my research. His profound knowledge and commitment to excellence have shaped me into a better man in research field. I am grateful for his patience, insightful feedback, and dedication to my growth as a researcher. He perfectly played the role of an elder full of wisdom, capability and patience, like Albus Dumbledore in the magic world, Gandalf in Middle-earth and Professor Ding Yi in Remembrance of Earth's Past.

I would like to extend my sincere appreciation to the members of my thesis committee, Professor Richard Gourse, Professor James Keck, Professor Aaron Hoskins and Professor Charlie Y. Mo, for their valuable insights, constructive criticisms, and thoughtful suggestions. Their expertise and willingness to share their knowledge have been instrumental in shaping the direction of my research. The atmosphere during committee meeting has always been cozy and peaceful, allowing me to share my ideas and questions freely.

I am grateful to the faculty and staff of the Department of Biochemistry at University of Wisconsin–Madison for providing an enriching academic environment. Their commitment to excellence in teaching and research has inspired and motivated me throughout my doctoral studies. I am indebted to the technicians, research assistants, and lab mates who have contributed to my research in various ways. Their collaborative spirit, assistance, and camaraderie have made the research journey more enjoyable and productive. I would like to extend my special thanks to Dr. Rachel Mooney, Dr. Sherry Xinyun Cao, Dr. Michael Engstrom, Dr. Michael Wolfe, Jason Saba and Expery Omollo, for their great passion in instructing me with experimental techniques, discussion on scientific problems and sharing sparkling moments in life.

I would like to acknowledge the financial support provided by National Institute of Health. This support has enabled me to carry out my research and pursue my academic goals. I am deeply appreciative of their investment in scientific research and their belief in my potential.

Finally, I would like to express my heartfelt gratitude to my family their unconditional love, encouragement, and belief in me through the long past time and long current distance. Their unwavering support, understanding, and patience have been a constant source of strength throughout this challenging journey. Their presence in my life has made the pursuit of this Ph.D. degree more rewarding and meaningful. I would like to express my deepest gratitude and profound love to my incredible fiancée, Dr. Huping Wang. Your enduring support, encouragement, and understanding throughout this challenging journey have been my guiding star and a constant source of strength. I am truly fortunate to have you as my partner in life and science. I am excited for our future together and the adventures that await us in the celestial voyage traversing the seas of stars in the universe of life.

To all those who have played a part, no matter how big or small, in shaping my academic and personal growth, I offer my sincere appreciation. Your contributions have been invaluable, and I am forever grateful for your support. I would like to use Chinese poem sentences from the great poet Li Bai to end my acknowledgement:

"Only wishing to sing and toast with wine,

as the moonlight casts a long glow within the golden goblet."

"唯愿当歌对酒时，月光长照金樽里"

May we cherish the moonlight.