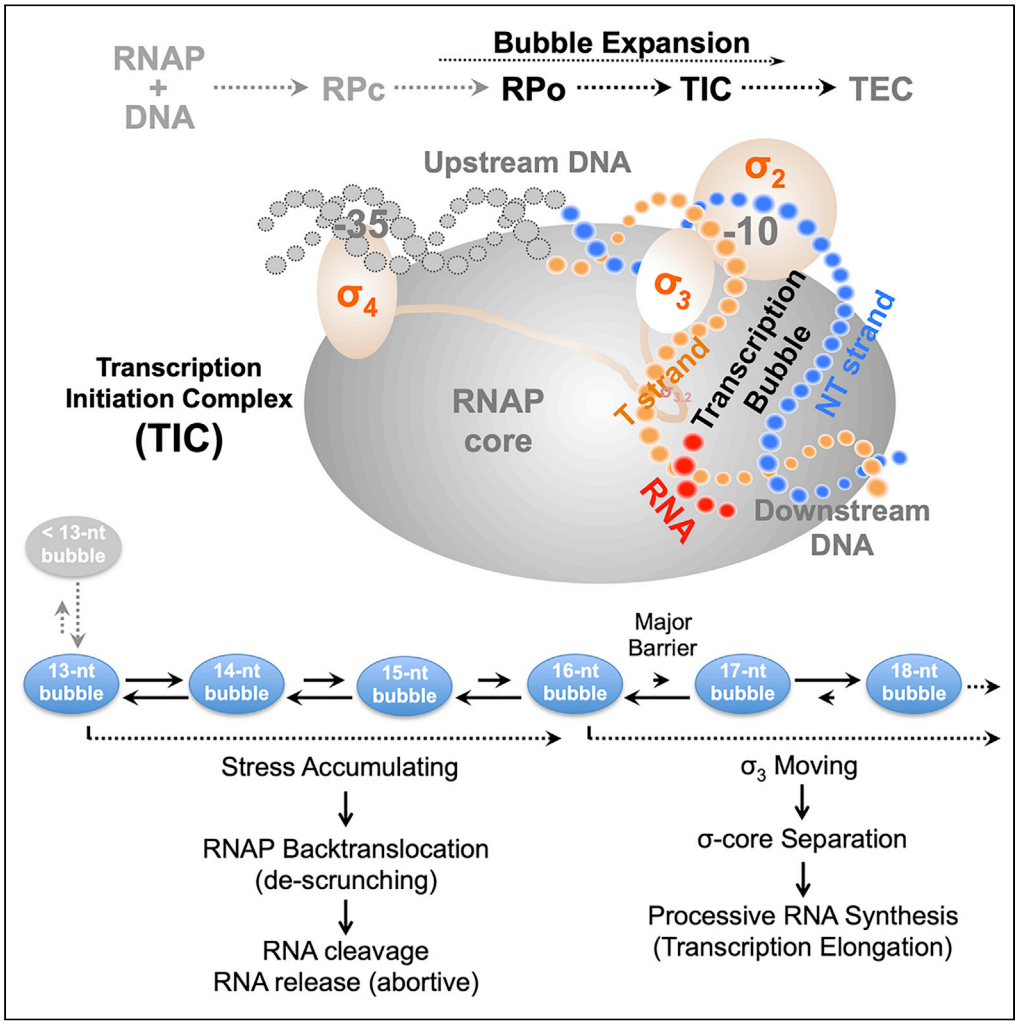
iScience

CellPress
OPEN ACCESS

Article

Structural Insights into Transcription Initiation from *De Novo* RNA Synthesis to Transitioning into Elongation



Yuhong Zuo,
Swastik De,
Yingang Feng,
Thomas A. Steitz

yuhong.zuo@yale.edu

HIGHLIGHTS

In-crystal *de novo* synthesis of RNA from 0 nt (empty bubble) up to 8 nt

Active RNA cleavage in crystals of transcription initiation complexes (TICs)

$\sigma_{3.2}$ on the RNA path is not directly involved in restricting RNA synthesis in TICs

Template DNA scrunching accumulates stress to drive σ-RNAP core separation

Zuo et al., iScience *23*, 101445
September 25, 2020
https://doi.org/10.1016/j.isci.2020.101445

iScience

CellPress
OPEN ACCESS

Article

Structural Insights into Transcription Initiation from *De Novo* RNA Synthesis to Transitioning into Elongation

Yuhong Zuo,[1,5,8,*] Swastik De,[1,2,5,6] Yingang Feng,[1,3] and Thomas A. Steitz[1,2,4,7]

SUMMARY

In bacteria, the dissociable σ subunit of the RNA polymerase (RNAP) is responsible for initiating RNA synthesis from specific DNA sites. As nascent RNA grows, downstream DNA unwinds and is pulled into the RNAP, causing stress accumulation and initiation complex destabilization. Processive transcription elongation requires at least partial separation of the σ factor from the RNAP core enzyme. Here, we present a series of transcription complexes captured between the early initiation and elongation phases via in-crystal RNA synthesis and cleavage. Crystal structures of these complexes indicate that stress accumulation during transcription initiation is not due to clashing of the growing nascent RNA with the $\sigma_{3.2}$ loop, but results from scrunching of the template strand DNA that is contained inside the RNAP by the σ_3 domain. Our results shed light on how scrunching of template-strand DNA drives both abortive initiation and σ-RNAP core separation to transition transcription from initiation to elongation.

INTRODUCTION

Transcription initiation is a dynamic process that is highly regulated during gene expression. This multistep process begins with sequence-specific interactions by RNA polymerase (RNAP) with the promoter DNA to form an open complex (RPo) (Boyaci et al., 2019; Chen et al., 2020; McClure, 1985; Saecker et al., 2011). RPo contains a transcription bubble of unwound DNA immediately downstream of the promoter region and is capable of *de novo* RNA synthesis using ribonucleoside triphosphates (NTPs). During transcription initiation, RNAP remains associated with the DNA promoter. As the downstream DNA is unwound and pulled into RNAP for DNA-templated RNA synthesis, the flexible DNA within the unwound region is accommodated in the RNAP in a process known as "DNA scrunching" (Kapanidis et al., 2006; Revyakin et al., 2006). DNA scrunching causes stress accumulation, destabilizing the initiation complex and resulting in many initial transcripts being released as short oligonucleotides (abortive initiation) (Carpousis and Gralla, 1980; Goldman et al., 2009; Hsu, 2002). Only a fraction of initial RNA synthesis events proceed to transcription elongation and produce full-length RNAs.

In bacteria, a single RNAP type is responsible for all RNA synthesis, but one of a plethora of σ factors is required to associate with the RNAP core enzyme and form an RNAP holoenzyme for promoter recognition and transcription initiation (Feklistov et al., 2014). For the primary σ^{70} factor and many related σ factors, the σ_2 and σ_4 domains of the σ factor interact with the promoter −10 and −35 elements, respectively, whereas the $\sigma_{3.2}$ loop inserts deep inside RNAP and lies on the path of the RNA transcript (Murakami et al., 2002). It was suggested that the $\sigma_{3.2}$ loop would clash with the 5′-end of the growing nascent RNA to play an important role in destabilizing the initiation complex. A recent study demonstrated a stepwise displacement of the $\sigma_{3.2}$ loop driven by RNA extension using synthetic DNA promoters corresponding to the downstream fork of an initiation bubble (Li et al., 2020). However, this stepwise displacement of the $\sigma_{3.2}$ loop is an experimental artifact that is unlikely to occur when DNA scrunching is involved as with intact DNA promoters. Multiple structural and biochemical studies have shown that the scrunching template-strand DNA (tDNA) is accommodated in an internal space of RNAP (Liu et al., 2016; Winkelman et al., 2015; Zuo and Steitz, 2015). We expect the scrunching tDNA to press on the σ_3 domain directly and contribute substantially to the stress accumulation in an initiation complex (Liu et al., 2016; Zuo and Steitz, 2015).

Previously, we reported the crystallization of the transcription initiation complexes (TICs) of *E. coli* RNAP containing the stress-responsive σ^S factor, an alternative σ factor closely related to the primary σ^{70} factor

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

[2]Howard Hughes Medical Institute, New Haven, CT 06510, USA

[3]Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, 189 Songling Road, Laoshan District, Qingdao, Shandong 266101, China

[4]Department of Chemistry, Yale University, New Haven, CT 06520, USA

[5]These authors contributed equally

[6]Present address: Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

[7]Deceased October 9, 2018

[8]Lead Contact

*Correspondence: yuhong.zuo@yale.edu

https://doi.org/10.1016/j.isci.2020.101445

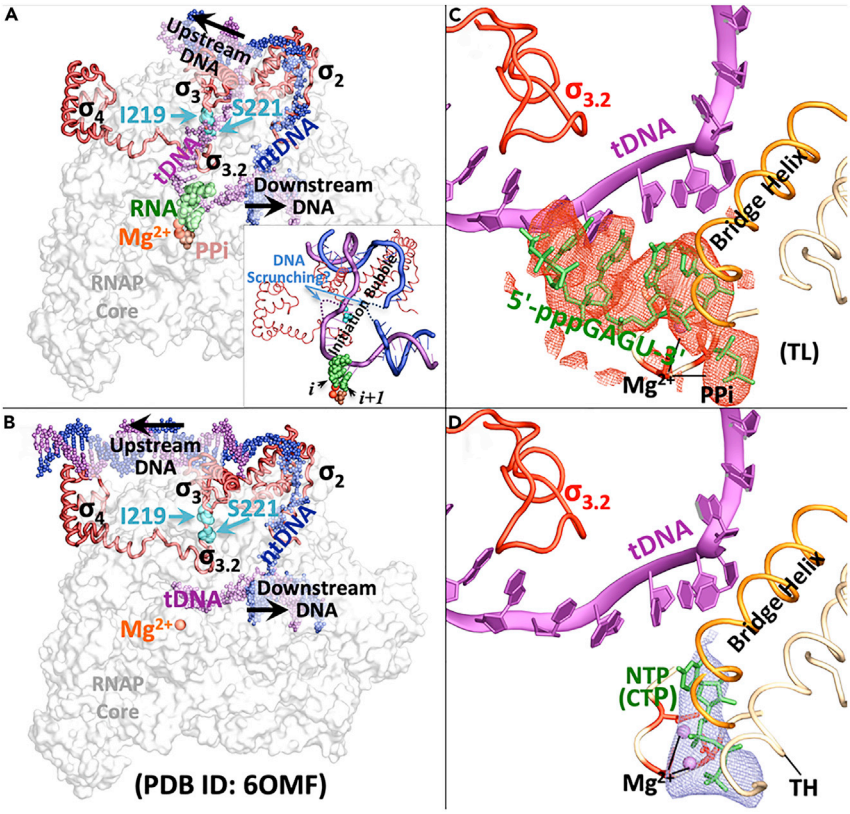


Figure 1. Crystal Structures of σS-Transcription Initiation Complexes

(A) Crystal structure of STIC4, with the core enzyme (gray surface), the σ^S factor (red), the DNA strands (blue and purple), and the nascent RNA (green) highlighted. The inset highlights the DNA bubble during transcription initiation.

(B) The STIC structure model rebuilt from a cryo-EM structure of the crl-associated STIC (PDB ID: 6OMF).

(C) An empty initiation bubble (RNA depletion) in SRPo crystals. The mesh diagram contours the $F_{o(old)}$-$F_{o(fresh)}$ density map.

(D) CTP association with the empty initiation bubble in the SRPo crystal. The F_o-F_o density maps (meshes) here and in the following figures are all computed relative to the old σ^S-TIC crystal model that carries an empty bubble (F_o-$F_{o(old)}$). All the F_o-F_o density maps are contoured at $\pm 3.0\sigma$ with positive density in blue and negative density in red. ntDNA, non-template strand DNA; tDNA, template strand DNA. Bridge helix, the RNAP structural element that branches the RNAP secondary channel from the primary channel. TH/TL, trigger helices/trigger loop, the RNAP structural element that controls the active site opening to the secondary channel.

(Liu et al., 2016). Crystallized σ^S-containing TICs (STICs) carry a complete transcription bubble and a 5′-triphosphorylated RNA tetranucleotide (4-mer) 5′-pppGAGU-3′ synthesized *de novo* using NTPs. In our current study, we captured a series of transcription complexes that are between the early initiation and elongation phases via in-crystal RNA synthesis and cleavage reactions (see Transparent Methods and Table S1). We show that RNAP backtranslocation is actively happening and RNA synthesis beyond 6 nucleotides (nt) is strongly inhibited in STIC crystals. Both RNAP backtranslocation and inhibition of RNA synthesis are likely caused by stress accumulated in the initiation complexes. We provide evidence that clashing of the growing nascent RNA with the $\sigma_{3.2}$ loop does not account for inhibiting RNA synthesis, and thus the scrunching tDNA that presses on the σ_3 domain is likely the major contributor to the stress accumulation that also drives the σ factor to separate from the RNAP core enzyme.

RESULTS

Crystallization of STICs and In-Crystal RNAP Activity Assays

E. coli STIC crystals grow to full size in about 1 week and remain enzymatically active while present in the crystallization drops (Liu et al., 2016). Crystallized STICs (Figure 1A) demonstrate essentially the same architecture as the STIC in solution (Figure 1B) (Cartagena et al., 2019), but lack the contacts between the σ_4

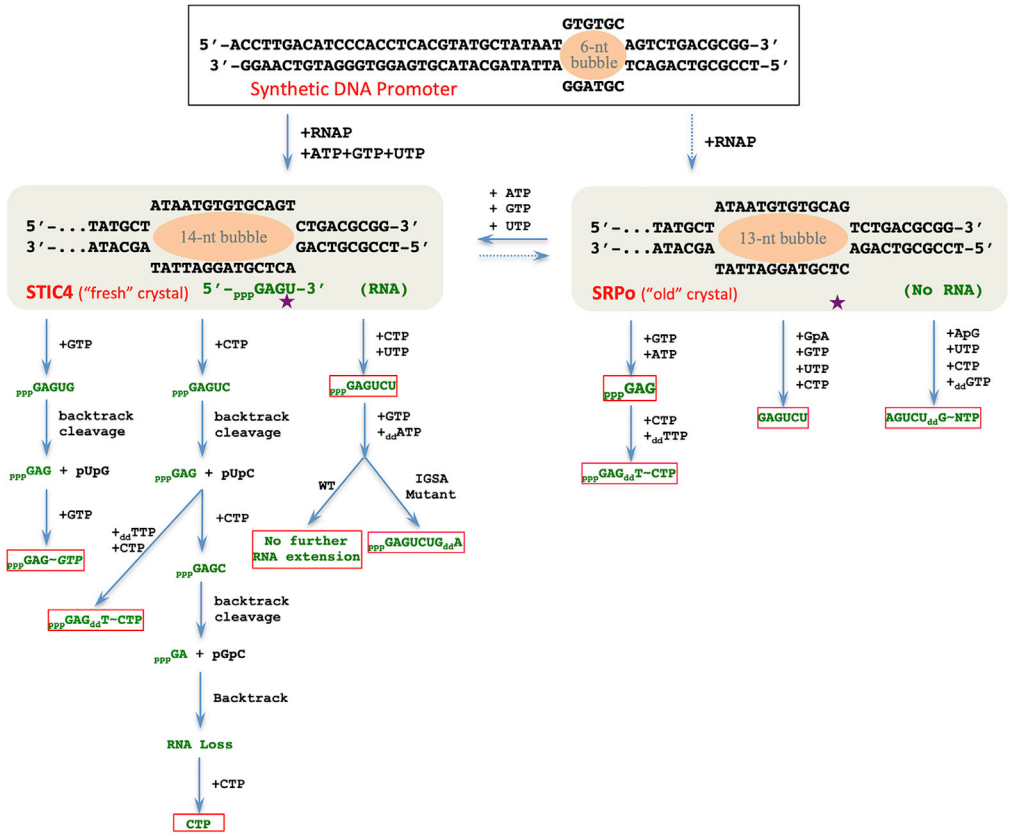


Figure 2. Schematic Diagram of RNA Syntheses and Cleavages by STICs

The RNA or nucleotides that bind at the RNAP active site chamber are shown in green, and the red frames mark some of the observed nucleotides or RNA oligonucleotides in this study. The purple star marks the RNAP catalytic site.

domain and the promoter −35 element due to crystal packing. Compared with the STIC in solution, crystallized STICs show slightly tighter clamping by the pincers and a more localized trigger loop (TL) insertion domain. If an STIC crystal is harvested within 3 weeks of crystal setup ("fresh" crystal), it always displays a well-ordered RNA 4-mer at the pre-translocated register along with a trapped pyrophosphate ion (PPi) at the active site (Liu et al., 2016), referred to as STIC4 hereafter (Figure 1A); if a crystal was harvested 6 weeks or longer after crystal setup ("old" crystal), we found that it had lost the nascent RNA. The STIC in an old crystal essentially represents the RPo form that is ready for NTP binding and *de novo* RNA synthesis, referred to as SRPo hereafter (Figures 1C and 1D).

With the specific sequence in our promoter design (Figure 2), the nascent RNA in the STIC4 crystals might be extended stepwise from 4 up to 8 nt by soaking the crystals with different NTP combinations, whereas the SRPo crystals should allow *de novo* RNA synthesis from an empty bubble (no RNA or NTP) to form a nascent RNA up to 5 nt in a stepwise manner by soaking the crystals with different NTP combinations. In addition to normal RNA addition, many competing processes occur during transcription initiation. These competing processes include nucleotide mis-incorporation, RNAP forward or backward translocation, and RNA cleavage via hydrolytic or pyrophosphorolytic mechanisms. Many of these processes could be explored using our in-crystal RNA synthesis and cleavage assays with STIC crystals (Figure 2).

STIC4 Crystals Are Active in Both RNA Synthesis and Cleavage

In the presence of both CTP and UTP, STIC4 crystals extend the RNA 4-mer to the RNA 6-mer 5′-pppGA–GUCU-3′ (Figure 3A). However, there was no indication of the expected RNA 5-mer 5′-pppGAGUC-3′ forming in STIC4 crystals after they were soaked in the cryo buffer supplemented with CTP alone (data not shown). Interestingly, STIC4 crystals lost the RNA completely after being soaked with CTP for an extended period, and the residual electron density at the nucleotide addition (i+1) site was similar to what we observed with

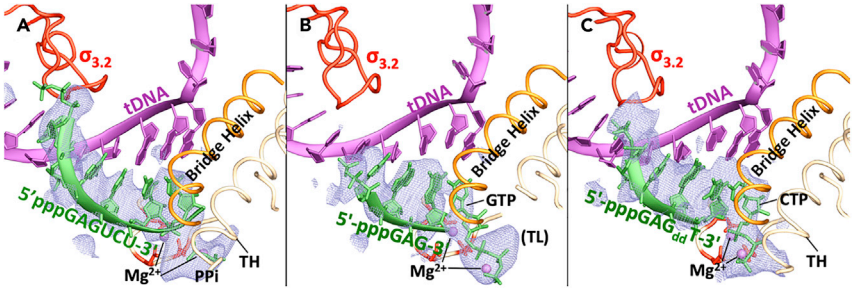


Figure 3. RNA Synthesis and Cleavage by STIC4 Crystals

(A) RNA extension by soaking STIC4 crystals with CTP and UTP.

(B) RNA cleavage by soaking STIC4 crystals with GTP.

(C) RNA synthesis and cleavage by soaking STIC4 crystals with CTP, GTP, and $_{dd}$TTP.

SRPo crystals soaked with CTP alone (Figure S1A). RNA depletion was also observed after STIC4 crystals were soaked with UTP alone (Figure S1B). However, soaking STIC4 crystals with ATP and/or GTP did not result in RNA depletion; instead, the pre-translocated RNA 4-mer was converted to a post-translocated RNA trinucleotide (3-mer) along with an NTP in a tilted conformation at the i+1 site (Figures 3B and S1C). These observations indicate that the RNA 4-mer in STIC4 crystals was extended to a 5-mer (5′-pppGAGUC-3′, 5′-pppGAGUU-3′, 5′-pppGAGUA-3′, or 5′-pppGAGUG-3′, respectively), but this RNA 5-mer was very unstable and thus translocated backward, triggering RNA cleavage to remove backtracked nucleotides (Figure 2), a proofreading activity of RNAP typically associated with nucleotide misincorporations during transcription (Orlova et al., 1995; Zenkin et al., 2006).

In a separate experiment, STIC4 crystals were soaked with CTP, GTP, and 2′,3′-dideoxy TTP ($_{dd}$TTP) to extend the RNA 4-mer to a 6-mer (5′-pppGAGUC$_{dd}$T-3′) and check whether the resulting STIC with an RNA 6-mer would allow a matching NTP (a GTP here) to bind at the i+1 site. Unexpectedly, we observed a post-translocated RNA 4-mer and an NTP at the i+1 site (Figure 3C). This observation suggests the formation of an RNA 4-mer carrying a dideoxy 3′-end (5′-pppGAG$_{dd}$T-3′), likely through a combination of nucleotide addition and backtrack-stimulated RNA cleavage steps (Figure 2). Both soaking experiments that caused either RNA depletion or formation of the dideoxy-terminated RNA 4-mer suggest relatively active RNA cleavage activity by STIC crystals. It is likely that cleavage of backtracked RNA nucleotides played an important role in the depletion of RNA in SRPo crystals.

SRPo Crystals Are Active in *De Novo* RNA Synthesis

As shown in Figures 4A and 4B, SRPo crystals can refill the RNAP active site chamber with RNA oligonucleotides when both GTP and ATP are present. Unlike CTP (Figure 1D), soaking SRPo crystals with ATP or GTP individually does not lead to well-defined electron density for an NTP at or near the active center. After being soaked with ATP, GTP, and UTP, SRPo crystals readily regenerate the same RNA 4-mer as in STIC4 crystals (data not shown). Interestingly, SRPo crystals generate an RNA 5-mer, presumably 5′-pppGGUCU-3′, resting at the pre-translocated register after being soaked with GTP, UTP, and CTP (data not shown). However, soaking SRPo crystals with ATP, UTP, and CTP does not lead to noticeable oligonucleotide formation.

It was shown that RNAP could use various nucleotide analogs or very short oligonucleotides to initiate RNA synthesis (Bird et al., 2016; Goldman et al., 2011). We used RNA dinucleotides to control the starting position of RNA synthesis in SRPo crystals. Soaking SRPo crystals with nucleotide combinations containing the RNA dinucleotide 5′-GA-3′ led to the synthesis of RNA oligonucleotides in the absence of ATP (Figures 4C and 4D). The resulting nascent transcripts are expected to have the same sequences as that initiated with GTP and ATP but lack the 5′-triphosphate. Similarly, soaking SRPo crystals with nucleotide combinations containing the RNA dinucleotide 5′-AG-3′ led to the synthesis of RNA oligonucleotides in the absence of GTP (Figures 4E and 4F).

An Energy Barrier Restricts RNA Synthesis in STIC Crystals

For both RNA extension in STIC4 crystals and *de novo* RNA synthesis in SRPo crystals, we observed full-length RNAs up to 6 nt at the shortest time point (about 20 s) of NTP soaking, and RNA synthesis did

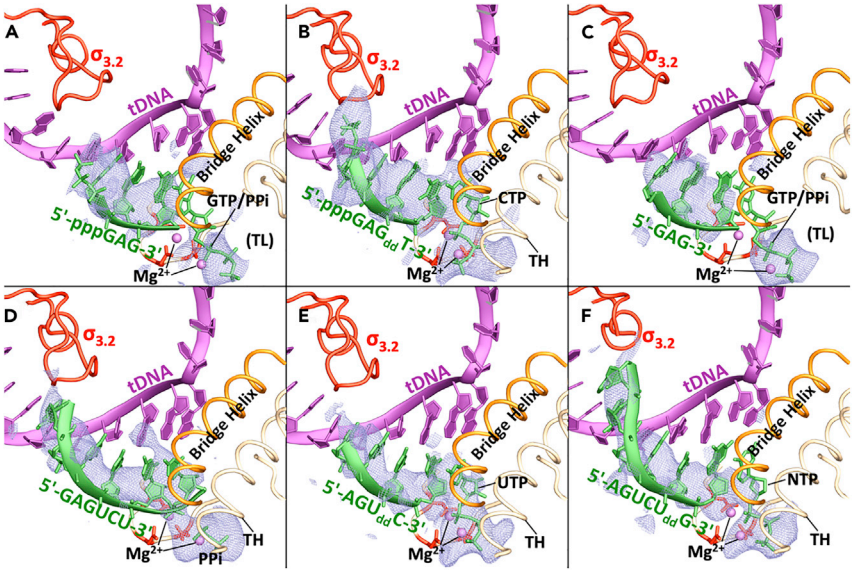


Figure 4. *De Novo* RNA Synthesis by SRPo Crystals

(A) RNA synthesis by soaking SRPo crystals with GTP and ATP.

(B) RNA synthesis with GTP, ATP, CTP, and ddTTP.

(C) RNA synthesis with GTP and RNA dinucleotide 5′-GpA-3′.

(D) RNA synthesis with GTP, UTP, CTP, and RNA dinucleotide 5′-GpA-3′.

(E) RNA synthesis with UTP, ddCTP, and RNA dinucleotide 5′-ApG-3′.

(F) RNA synthesis with UTP, CTP, ddGTP, and RNA dinucleotide 5′-ApG-3′.

not display a stepwise behavior. This non-stepwise behavior of RNA synthesis suggests that the stress accumulated in STICs can be easily surpassed by the binding event of a matching NTP at the i+1 site, possibly because the NTPs used for soaking were at relatively high concentrations (1 mM). However, there appears to be a major barrier keeping RNA synthesis in check at 6 nt. When we attempted to extend the RNA 4-mer up to 8 nt by soaking STIC4 crystals with CTP, UTP, GTP, and 2′,3′-dideoxy ATP ($_{dd}$ATP), no RNA longer than 6 nt was visible after crystal soaking for up to 2 h, but electron density for an RNA 6-mer became very well ordered at the 2-min time point (Figure S1D). This size limit on the nascent RNA was also observed with SRPo crystals (Figure 4D): soaking SRPo crystals with GTP, UTP, CTP, and the RNA dinucleotide 5′-GA-3′ gave a pre-translocated RNA 6-mer (5′-GAGUCU-3′), but not the RNA 7-mer 5′-GAGUCUG-3′.

$\sigma_{3.2}$ Loop Unlikely to Block RNA Synthesis

Interestingly, when RNA synthesis was started with the RNA dinucleotide 5′-AG-3′ by soaking SRPo crystals with UTP, CTP, 2′,3′-dideoxy GTP ($_{dd}$GTP), and 5′-AG-3′, we observed a post-translocated RNA 6-mer (5′-AGUCU$_{dd}$G-3′) and an NTP at the i+1 site (Figure 4F), which suggests that this complex is capable of forming an RNA 7-mer if the RNA 6-mer is not terminated with a dideoxy 3′-end. Because 5′-AG-3′ primes RNA synthesis one nucleotide further downstream than 5′-GA-3′, the difference between 5′-AG-3′ and 5′-GA-3′ in initiating RNA synthesis in SRPo crystals strongly suggests that accommodation of longer nascent RNA does not involve the growing RNA pushing away the $\sigma_{3.2}$ loop. In other words, although the $\sigma_{3.2}$ loop lies on the RNA path inside RNAP, it is not directly responsible for blocking the nascent RNA from growing longer than 6 nt during normal transcription initiation. A recent study by Li et al. (2020) demonstrated that, in the absence of DNA scrunching, RNAs rest at the post-translocated register while displacing the $\sigma_{3.2}$ loop, which also suggests that the $\sigma_{3.2}$ loop exerts little resistance to the RNA extension.

Mutations in $\sigma_{3.1}$ Linker Alleviate the Barrier for RNA Synthesis

Structural and biochemical studies suggested that the scrunching tDNA is enclosed in an RNAP chamber capped by the σ_3 domain during transcription initiation (Liu et al., 2016; Winkelman et al., 2015; Zuo and Steitz, 2015). Based on a structural study with σ^{70}-containing TICs, we proposed previously that the scrunching tDNA presses directly on the σ_3 domain and the RNAP chamber might accommodate the

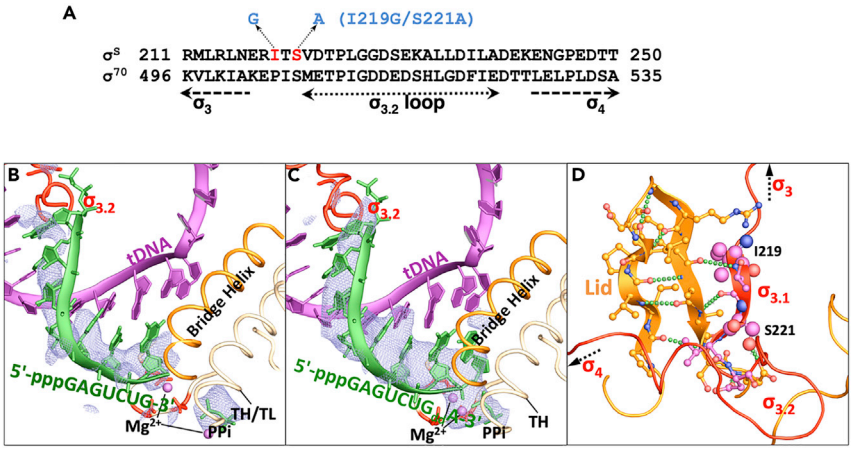


Figure 5. STIC Crystals Carrying a Mutant σ^S Factor

(A) Aligned $\sigma_{3.2}$ and flanking sequences of the *E. coli* σ^S and σ^{70} factors with the σ^S IGSA mutations (I219G/S221A) highlighted. Positions of the I219 and S221 residues in the RNAP holoenzyme are shown in Figure 1B.

(B) RNA extension by soaking IGSA mutant STIC4 crystals with CTP, UTP, and GTP.

(C) RNA extension by soaking IGSA mutant STIC4 crystals with CTP, UTP, GTP, and ddATP.

(D) Interactions between $\sigma_{3.1}$ and the RNAP lid element. σ^S I219 and S221 residues are shown as large ball and sticks. Inter-strand hydrogen bonds are shown as green dashed lines.

tDNA strand of an initiation bubble no more than about 16 nt (Zuo and Steitz, 2015). Accommodating the scrunching tDNA in this limited internal space likely contributes substantially to the stress accumulation and thus plays a major role in restricting the growth of the nascent RNA. To test this hypothesis, we prepared STICs with mutations on the σ^S factor that would increase the mobility of the σ_3 domain. One mutant STIC that carries mutations I219G and S221A on the σ^S factor (IGSA mutant STIC) forms essentially the same crystals as the wild-type (WT) STIC. Both I219 and S221 are residues in the $\sigma_{3.1}$ region that connect the σ_3 globular domain and the $\sigma_{3.2}$ loop of the σ^S factor (Figures 1A and 5). Like WT STIC crystals, fresh IGSA mutant STIC crystals also contain an RNA 4-mer, trap a PPi at the active site, and are able to extend the RNA 4-mer to a 6-mer after being soaked with CTP and UTP. In contrast to what we observed with WT STIC crystals, the RNA in IGSA mutant STIC crystals was extended noticeably longer at the 5′-end, forming an RNA 7-mer (5′-pppGAGUCUG-3′) positioned at the post-translocation register after the crystals were soaked with CTP, UTP, and GTP (Figure 5A). When 2′,3′-dideoxy ATP ($_{dd}$ATP) was included along with CTP, UTP, and GTP during the soaking, an RNA 8-mer (5′-pppGAGUCUG$_{dd}$A-3′) was observed at the pre-translocation register (Figure 5B). These observations suggest that the energy barrier that limits the nascent RNA to 6 nt in WT STIC crystals does not block RNA synthesis in IGSA mutant STIC crystals.

DISCUSSION

Active Site Conformations of Transcription Complexes

Previous structural studies showed that transcription complexes display various conformational states with regard to NTP/PPi association, RNAP translocation, and active site opening (Liu et al., 2016; Zuo and Steitz, 2017). In the absence of an incoming NTP, the post-translocation state with an open active site was frequently observed in crystal structures of transcription elongation complexes, whereas TICs containing a complete DNA bubble frequently displayed a closed active site with the RNA at the pre-translocation register (Liu et al., 2016; Zuo and Steitz, 2015). In this study, we observed that the STICs containing an RNA 3-mer (5′-pppGAG-3′ or 5′-GAG-3′) form a 14-nt bubble with the RNA at the post-translocation register (e.g., Figures 3B, 4A and, 4C). These post-translocated complexes display an unfolded TL and an active site open to the secondary channel. In some of these complexes, we observed non-matching NTPs at the i+1 site (e.g., Figures 3B and S1C), suggesting a non-selective NTP association by these complexes. For the STICs that contain a longer nascent RNA and form a larger DNA bubble (e.g., Figures 3A and 3C), the RNAs remain at the pre-translocation register unless a matching NTP binds at the i+1 site. In both pre-translocated STICs and STICs with a matching NTP at the i+1 site, the RNAP active site is closed by TL folding into trigger helices. The conformation of an STIC appears to be affected by the size of the DNA bubble, the length of the nascent RNA, or both. As the nascent RNA grows and the DNA bubble expands,

the post-translocation state becomes more and more disfavored and eventually even a matching NTP becomes barely capable of promoting RNAP forward translocation. The different STIC conformations are thus indicative of stress accumulation in these initiation complexes.

Bubble Size Matters during Transcription Initiation

Transcription initiation involves formation of the RPo with a DNA bubble of appropriate size to start DNA-templated RNA synthesis. Although details of the dynamic RPo formation process remain obscure, a recent study showed that an 8-nt bubble is not large enough to position the single-stranded DNA template into the RNAP active site for *de novo* RNA synthesis (Boyaci et al., 2019). Structures reported here and in many other studies (Bae et al., 2015; Hubin et al., 2017; Liu et al., 2016; Zhang et al., 2012; Zuo and Steitz, 2015) suggest that RNAP interacts most comfortably with a 13-nt single-stranded segment of the non-template strand DNA (ntDNA), whereas up to 16 nt of the tDNA segment of an initiation bubble might be enclosed inside the RNAP (Liu et al., 2016; Zuo and Steitz, 2015). Due to the dynamic aspect of DNA scrunching that starts before RNA synthesis begins, the DNA bubble is flexibly positioned relative to the RNAP active site in the RPo, and thus selection of the transcription start site frequently displays heterogeneity (Robb et al., 2013; Winkelman et al., 2016a, 2016b). The structural observations explain why most *E. coli* transcriptions start at 7–8 nt downstream of the conserved −10 element (Vvedenskaya et al., 2015). This transcription start site preference is likely specific to the σ factor; other σ factors, such as the extrocytoplasmic σ factors, appear to favor DNA bubbles of different sizes and thus initiate RNA synthesis from other preferred positions (Li et al., 2019; Lin et al., 2019).

DNA Scrunching and Stress Accumulation

As the nascent RNA grows and the initiation bubble expands, stress accumulates in the initiation complex to suppress RNAP forward translocation and to promote RNAP backtranslocation, as evidenced by the predominance of pre-translocated STICs and the occurrence of RNA cleavage in STIC crystals when longer RNA transcripts and larger DNA bubbles are formed. The $\sigma_{3.2}$ loop that sits on the path of the RNA starts to make contact with the nascent RNA when the RNA grows to about 5 nt (Zuo and Steitz, 2015). Although there appears to be a major energy barrier that blocks RNA synthesis beyond 6 nt in STIC crystals, this study suggests that the RNA-$\sigma_{3.2}$ loop clash is not responsible for restricting nascent RNA synthesis. Therefore, DNA scrunching must be the major contributor to the stress accumulation during transcription initiation. As the ntDNA strand is partially exposed and should start to bulge to the outside of the RNAP at a size of about 13–14 nt (Zuo and Steitz, 2015), ntDNA scrunching might affect only the selection of the transcription start site, and contribute very little to stress buildup as the bubble expands beyond 14 nt. On the other hand, the tDNA strand is held completely inside the RNAP, so tDNA scrunching could keep building up stress as the bubble expands during transcription initiation.

Relaxation of Stressed Initiation Complexes

The stress in the initiation complex keeps accumulating as the DNA bubble expands. To relieve the stress, one of two processes has to occur: (1) RNAP must backtranslocate to de-scrunch the DNA and decrease the size of the DNA bubble or (2) the multi-domain σ factor must separate at least partially from the RNAP core enzyme to create space for the scrunching tDNA.

We demonstrate here that the nascent RNA is actively cleaved in the STIC crystals. This suggests that RNAP backtranslocation occurs frequently, because the intrinsic RNA cleavage activity of RNAP is very weak except when the RNA is at a backtracked state (Orlova et al., 1995; Zenkin et al., 2006). Processive backtranslocation of RNAP could lead to release of the nascent RNAs (abortive initiation). In our STIC crystals, it appears that many RNA cleavage events may occur before an RNA oligonucleotide gets released. However, RNA cleavage events likely occur less frequently in solution, where RNAP translocates much more dynamically.

The single-stranded tDNA of the initiation bubble is accommodated in an internal space capped by the σ_3 globular domain of the σ factor on the outer surface of the RNAP. We argued previously that the internal chamber of the σ^{70}-RNAP is not spacious enough to accommodate a DNA bubble larger than about 16 nt (Zuo and Steitz, 2015). This appears to be true for the σ^{S}-RNAP as well. As the RNA extends to 7 or 8 nt after NTP soaking, the DNA bubble expands to 18 nt in IGSA mutant STIC crystals. Similarly, formation of the post-translocated RNA 6-mer 5′-AGUCU$_{aa}$G-3′ in WT STIC crystals also expands the initiation bubble to 18 nt, which is in contrast to the 16-nt DNA bubbles associated with the pre-translocated RNA 6-mers

(5′-pppGAGUCU-3′ and 5′-GAGUCU-3′). In both situations, we noticed a likely outward shift by the σ_3 domain to make space for the scrunching tDNA (Figure S2). Interestingly, we observed 18-nt bubbles with both pre- and post-translocated RNAs (Figures 4F and 5A), but there is no clear evidence of a stable 17-nt DNA bubble in our STIC crystals. This may imply that an 18-nt bubble is energetically more favorable than a 17-nt bubble in these complexes. It is likely that the σ_3 domain moves at the time the bubble expands from 16 nt to 17 nt, which breaks a major energy barrier and releases the stress accumulated as the bubble expands up to 16 nt.

Beginning of σ-RNAP Core Separation and Transitioning from Initiation to Elongation

Shifting the σ_3 domain outward would move the $\sigma_{3.2}$ loop along and create space for the growing nascent RNA. This is reminiscent of the transition of transcription from initiation to elongation by T7 RNAP, in which the scrunching tDNA pushes around the promoter-binding domain to clear the path for the nascent RNA (Durniak et al., 2008). Therefore, this σ_3 movement likely marks the beginning of the σ separation from the RNAP core enzyme for transitioning transcription from initiation into the elongation phase. Outward motion of the σ_3 domain would break multiple inter-strand hydrogen bonds and disrupt the β-β interactions between the $\sigma_{3.1}$ linker region and the β-hairpin of the lid element deep inside the well-folded RNAP holoenzyme (Figure 5C). It is conceivable that considerable energy may be required to disrupt the interactions between $\sigma_{3.1}$ and the RNAP lid, consistent with our observed energy barrier restricting RNA synthesis to 6 nt in the STIC crystals. Mutations I219G and S221A on the σ^S factor could lower the energy requirement for disrupting the RNAP lid-$\sigma_{3.1}$ interactions. Mutating P504 and S506, the structural counterparts of I219 and S221 residues, on the primary σ^{70} factor was shown to reduce abortive RNA synthesis and increase the rate of promoter escape (Cashel et al., 2003), likely functioning similarly by facilitating the disruption of β-β interactions between the σ^{70} factor and the RNAP core enzyme.

Concluding Remarks

Based on this study, stress accumulation during transcription initiation is largely the result of tDNA scrunching, and larger initiation bubbles accumulate more stress. RNA synthesis from some promoters, such as those of the ribosomal RNA (rRNA) genes, starts at 9 or 10 nt downstream of the conserved −10 element. Compared with RNA synthesis starting at the more common positions (7 or 8 nt downstream of the −10 element), these promoters require larger bubbles to be pre-formed in the RPos. Therefore, transcription from these promoters should require higher concentrations of the initiation NTPs to overcome the greater stress accumulation. Unsurprisingly, it was shown previously that transcription initiation from rRNA promoters displays a strong dependency on concentrations of the initiation NTPs (Schneider et al., 2002). In addition, a larger pre-formed initiation bubble may also mean that σ-RNAP core separation starts at an earlier phase of the initial RNA synthesis, likely making the transition of transcription from initiation to elongation look more efficient.

Limitations of the Study

Here we captured a series of transcription complexes between the early initiation and elongation phases by employing in-crystal RNA synthesis and cleavage reactions. As this study is based on in-crystal activity assays, the DNA design for crystallizable complete TICs and for achieving stepwise RNA extension prevented us from observing transcription bubbles larger than 18 nt for further σ-RNAP core separation. Nevertheless, our crystals provide valuable clues why initial RNA synthesis is ineffective and how stress accumulates and drives both RNA backtranslocation and σ-RNAP core separation during transcription initiation.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yuhong Zuo (yuhong.zuo@yale.edu).

Materials Availability

All unique reagents generated in this study are available from the Lead Contact without restriction.

Data and Code Availability

Coordinates and structure factor files have been deposited in the RCSB Protein DataBank under accession codes: 6UTW (STIC4, 4-nt RNA), 6UTX (SRPo, empty DNA bubble), 6UTY (SRPo soaked with CTP), 6UTZ

(STIC4 soaked with CTP and UTP), 6UU0 (STIC4 soaked with GTP), 6UU1 (STIC4 soaked with CTP, GTP and ddTTP), 6UU2 (SRPo soaked with GTP and ATP), 6UU3 (SRPo soaked with GTP, ATP, CTP and ddTTP), 6UU4 (SRPo soaked with GTP and dinucleotide 5′-GA-3′), 6UU5 (SRPo soaked with GTP, UTP, CTP and dinucleotide 5′-GA-3′), 6UU6 (SRPo soaked with UTP, ddCTP and dinucleotide 5′-AG-3′), 6UU7 (SRPo soaked with UTP, CTP, ddGTP and dinucleotide 5′-AG-3′), 6UU8 (IGSA mutant STIC4 soaked with CTP, UTP and GTP), 6UU9 (IGSA mutant STIC4 soaked with CTP, UTP, GTP and ddATP), 6UUA (STIC4 soaked with CTP), 6UUB (STIC4 soaked with UTP), 6UUC (STIC4 soaked with ATP), 6UTV (STIC4 soaked with CTP, UTP, GTP and ddATP).

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101445.

ACKNOWLEDGMENTS

We thank the staff of Argonne National Laboratory beamlines 24-ID-C and 24-ID-E for their help during data collection, and the Center for Structural Biology Facility at Yale University for computational support. We also thank Ronald R. Breaker and Michael M. Zuo for critical reading of the manuscript. This work was supported by NIH grant GM22778 (to T.A.S.). T.A.S. was an investigator of the Howard Hughes Medical Institute. This work was also partly supported by the National Natural Science Foundation of China (Grant 31670735 to Y.F.) and the Chinese Government Scholarship (Grant 201604910035 to Y.F.).

AUTHOR CONTRIBUTIONS

Y.Z. and T.A.S. designed the project; Y.Z., S.D., and Y.F. performed the experiments; Y.Z. and S.D. processed the data and interpreted the results; Y.Z. and S.D. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 1, 2020
Revised: July 27, 2020
Accepted: August 6, 2020
Published: September 25, 2020

REFERENCES

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S.A. (2015). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. Elife *4*, e08504.

Bird, J.G., Zhang, Y., Tian, Y., Panova, N., Barvik, I., Greene, L., Liu, M., Buckley, B., Krasny, L., Lee, J.K., et al. (2016). The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. Nature *535*, 444–447.

Boyaci, H., Chen, J., Jansen, R., Darst, S.A., and Campbell, E.A. (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. Nature *565*, 382–385.

Carpousis, A.J., and Gralla, J.D. (1980). Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. Biochemistry *19*, 3245–3253.

Cartagena, A.J., Banta, A.B., Sathyan, N., Ross, W., Gourse, R.L., Campbell, E.A., and Darst, S.A. (2019). Structural basis for transcription activation by Crl through tethering of sigma(S) and RNA polymerase. Proc. Natl. Acad. Sci. U S A *116*, 18923–18927.

Cashel, M., Hsu, L.M., and Hernandez, V.J. (2003). Changes in conserved region 3 of Escherichia coli sigma 70 reduce abortive transcription and enhance promoter escape. J. Biol. Chem. *278*, 5539–5547.

Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A.Y., Olinares, P.D.B., Saecker, R.M., Winkelman, J.T., Maloney, M.F., Chait, B.T., Ross, W., et al. (2020). Stepwise promoter melting by bacterial RNA polymerase. Mol. Cell *78*, 275–288 e276.

Durniak, K.J., Bailey, S., and Steitz, T.A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. Science *322*, 553–557.

Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. Annu. Rev. Microbiol. *68*, 357–376.

Goldman, S.R., Ebright, R.H., and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts in vivo. Science *324*, 927–928.

Goldman, S.R., Sharp, J.S., Vvedenskaya, I.O., Livny, J., Dove, S.L., and Nickels, B.E. (2011). NanoRNAs prime transcription initiation in vivo. Mol. Cell *42*, 817–825.

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

Hubin, E.A., Lilic, M., Darst, S.A., and Campbell, E.A. (2017). Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. Nat. Commun. *8*, 16072.

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

Li, L., Fang, C., Zhuang, N., Wang, T., and Zhang, Y. (2019). Structural basis for transcription initiation by bacterial ECF sigma factors. Nat. Commun. *10*, 1153.

Li, L., Molodtsov, V., Lin, W., Ebright, R.H., and Zhang, Y. (2020). RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. Proc. Natl. Acad. Sci. U S A *117*, 5801–5809.

Lin, W., Mandal, S., Degen, D., Cho, M.S., Feng, Y., Das, K., and Ebright, R.H. (2019). Structural basis of ECF-sigma-factor-dependent transcription initiation. Nat. Commun. *10*, 710.

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

McClure, W.R. (1985). Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. *54*, 171–204.

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

Orlova, M., Newlands, J., Das, A., Goldfarb, A., and Borukhov, S. (1995). Intrinsic transcript cleavage activity of RNA polymerase. Proc. Natl. Acad. Sci. U S A *92*, 4596–4600.

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

Robb, N.C., Cordes, T., Hwang, L.C., Gryte, K., Duchi, D., Craggs, T.D., Santoso, Y., Weiss, S., Ebright, R.H., and Kapanidis, A.N. (2013). The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. J. Mol. Biol. *425*, 875–885.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol. *412*, 754–771.

Schneider, D.A., Gaal, T., and Gourse, R.L. (2002). NTP-sensing by rRNA promoters in Escherichia coli is direct. Proc. Natl. Acad. Sci. U S A *99*, 8602–8607.

Vvedenskaya, I.O., Zhang, Y., Goldman, S.R., Valenti, A., Visone, V., Taylor, D.M., Ebright, R.H., and Nickels, B.E. (2015). Massively systematic transcript end readout, "MASTER": transcription start site selection, transcriptional slippage, and transcript yields. Mol. Cell *60*, 953–965.

Winkelman, J.T., Chandrangsu, P., Ross, W., and Gourse, R.L. (2016a). Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of E. coli ribosomal RNA promoters. Proc. Natl. Acad. Sci. U S A *113*, E1787–E1795.

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

Winkelman, J.T., Winkelman, B.T., Boyce, J., Maloney, M.F., Chen, A.Y., Ross, W., and Gourse, R.L. (2015). Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. Mol. Cell *59*, 768–780.

Zenkin, N., Yuzenkova, Y., and Severinov, K. (2006). Transcript-assisted transcriptional proofreading. Science *313*, 518–520.

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

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

Zuo, Y., and Steitz, T.A. (2017). A structure-based kinetic model of transcription. Transcription *8*, 1–8.

iScience, Volume *23*

Supplemental Information

Structural Insights into Transcription Initiation from *De Novo* RNA Synthesis to Transitioning into Elongation

Yuhong Zuo, Swastik De, Yingang Feng, and Thomas A. Steitz

Table S1. Data Collection and Refinement Statistics. Related to All Figures.

PDB code	6UTW	6UTX	6UTY	6UTZ	6UU0	6UU1	6UU2	6UU3	6UU4	6UU5	6UU6	6UU7	6UU8	6UU9	6UUA	6UUB	6UUC	6UTV
Crystals[1]	STIC4	SRPo	SRPo	STIC4	STIC4	STIC4	SRPo	SRPo	SRPo	SRPo	SRPo	SRPo	Mutant STIC4	Mutant STIC4	STIC4	STIC4	STIC4	STIC4
Soaking	-	-	CTP	CTP, UTP	GTP	CTP, GTP, ddTTP	GTP, ATP	GTP, ATP, CTP, ddTTP	GpA, GTP	GpA, GTP, UTP, CTP	ApG, UTP, ddCTP	ApG, UTP, CTP, ddGTP	CTP, UTP, GTP	CTP, UTP, GTP, ddATP	CTP	UTP	ATP	CTP, UTP, GTP, ddATP
Data Collection																		
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)																		
a	132.99	132.07	132.58	132.93	132.80	132.85	131.47	132.56	132.14	132.36	132.40	132.94	133.46	132.18	132.96	132.94	133.10	132.53
b	155.45	153.30	152.73	155.99	156.16	154.25	153.23	153.58	153.43	153.77	154.11	153.77	154.50	153.47	153.83	154.99	154.92	155.43
c	234.74	230.90	229.85	233.58	233.37	230.73	229.17	231.21	230.96	231.69	232.15	232.62	235.70	230.83	231.23	233.25	231.78	232.60
Resolution (Å)	3.85 (4.09-3.85)	4.05 (4.29-4.05)	4.15 (4.40-4.15)	3.80 (4.03-3.80)	3.90 (4.13-3.90)	4.10 (4.34-4.10)	4.40 (4.67-4.40)	4.00 (4.24-4.00)	4.30 (4.56-4.30)	5.40 (5.73-5.40)	4.20 (4.45-4.20)	4.40 (4.66-4.40)	4.40 (4.66-4.40)	5.40 (5.72-5.40)	4.00 (4.24-4.00)	3.95 (4.19-3.95)	4.10 (4.34-4.10)	3.45 (3.66-3.45)
Rmeas (%)	12.0 (413.9)	18.9 (327.9)	19.4 (493.1)	14.0 (425.2)	13.8 (537.0)	11.6 (415.4)	17.2 (387.8)	12.4 (393.3)	15.1 (473.0)	20.1 (305.8)	20.0 (436.8)	20.8 (475.7)	22.1 (442.6)	25.7 (396.7)	10.2 (389.6)	9.7 (326.3)	10.0 (330.3)	15.4 (487.2)
I/σI	8.85 (0.49)	8.19 (0.56)	10.24 (0.50)	9.01 (0.46)	11.63 (0.48)	9.94 (0.50)	8.01 (0.51)	10.59 (0.48)	11.74 (0.54)	6.68 (0.55)	7.53 (0.46)	7.04 (0.43)	6.92 (0.44)	6.93 (0.47)	9.22 (0.43)	9.80 (0.56)	10.14 (0.55)	12.95 (0.52)
$CC_{1/2}$	0.999 (0.193)	0.998 (0.181)	0.999 (0.139)	0.999 (0.169)	0.999 (0.145)	0.999 (0.097)	0.999 (0.110)	0.999 (0.171)	0.999 (0.220)	0.996 (0.133)	0.999 (0.114)	0.998 (0.165)	0.998 (0.178)	0.995 (0.157)	0.999 (0.109)	0.999 (0.146)	0.999 (0.108)	0.999 (0.216)
Redundancy	3.5 (3.3)	5.6 (5.6)	13.4 (13.1)	6.7 (6.2)	13.4 (12.2)	6.1 (6.2)	6.6 (6.7)	6.8 (6.5)	13.4 (13.1)	6.2 (5.0)	6.8 (6.8)	6.7 (6.7)	6.0 (5.9)	5.9 (5.8)	5.3 (4.9)	5.0 (4.5)	5.0 (5.1)	12.3 (12.2)
Completeness (%)	93.7 (90.2)	99.3 (97.1)	99.3 (95.9)	99.5 (98.5)	99.4 (96.9)	98.6 (94.2)	98.6 (92.2)	99.3 (96.4)	99.6 (98.6)	99.4 (97.8)	99.1 (95.2)	99.4 (97.3)	99.4 (97.2)	99.0 (95.4)	93.1 (91.3)	99.1 (95.9)	99.3 (97.0)	99.7 (98.4)
Refinement																		
Resolution (Å)	49.4-3.85	48.9-4.05	49.0-4.15	49.4-3.80	49.4-3.90	49.2-4.10	49.7-4.40	49.1-4.00	49.0-4.30	49.0-5.40	49.1-4.20	49.2-4.40	49.4-4.40	49.0-5.40	47.9-4.00	38.9-3.95	49.3-4.10	49.3-3.45
No. reflections	43,641	38,603	35,751	48,197	44,697	37,445	29,564	40,237	32,436	16,704	35,121	30,874	31,470	16,580	37,926	42,501	38,147	63,761
R_{work} / R_{free} (%)	26.8 / 32.9	27.9 / 35.0	30.2 / 37.0	27.4 / 34.3	31.4 / 36.8	33.2 / 39.6	29.7 / 36.3	30.3 / 38.0	30.4 / 36.4	29.5 / 36.2	34.2 / 38.9	34.4 / 39.0	35.7 / 37.9	31.5 / 36.4	30.2 / 36.5	30.3 / 37.5	32.2 / 38.0	28.2 / 34.8
No. atoms	29,078	28,841	28,961	29,063	28,977	29,064	28,949	29,084	28,959	29,050	29,026	28,934	28,931	28,959	28,961	28,871	28,970	29,063
RMS Deviations																		
Bond lengths (Å)	0.006	0.005	0.005	0.006	0.006	0.006	0.004	0.006	0.006	0.005	0.006	0.006	0.009	0.006	0.005	0.005	0.005	0.006
Bond angles (°)	1.39	1.23	1.15	1.35	1.41	1.45	1.15	1.40	1.35	1.35	1.45	1.40	1.61	1.38	1.18	1.18	1.16	1.34
Structural Features																		
Bubble size (nt)	14	13	13	16	14	15	14	15	14	16	16	18	18	18	13	13	14	16
RNA/NTP/PPi	pppGAGU ~PPi	–	CTP	pppGAGUC U~PPi	pppGAG ~*GTP*	pppGAG$_{dd}$T ~CTP	pppGAG +PPi	pppGAG$_{dd}$T ~CTP	GAG+PPi	GAGUCU ~PPi	AGU$_{dd}$C ~UTP	AGUCU$_{dd}$G ~NTP	pppGAGUC UG+PPi/NTP?	pppGAGUC UG$_{dd}$A~PPi	CTP	*UTP*	pppGAG ~*ATP*	pppGAGUC U~PPi
Translocation[2]	Pre-	-	-	Pre-	Post-	Post-	Post-	Post-	Post-	Pre-	Pre-	Post-	Pre-/Post-?	Pre-	-	-	Post-	Pre-
TL/TH[3]	TH	TL	TH	TH	TL	TH	TL	TH	TL	TH	TH	TH	TH/TL ?	TH	TH	TL	TL	TH

Highest resolution shells are shown in parentheses. [1]STIC4 – fresh σ^S-TIC crystals; SRPo – old σ^S-TIC crystals; Mutant – σ^S I219G/S221A (IGSA) mutant. [2]Pre – pre-translocation; Post – post-translocation. [3]TL – trigger loop (flexible/delocalized, active site opened); TH – trigger helix (ordered, active site closed).

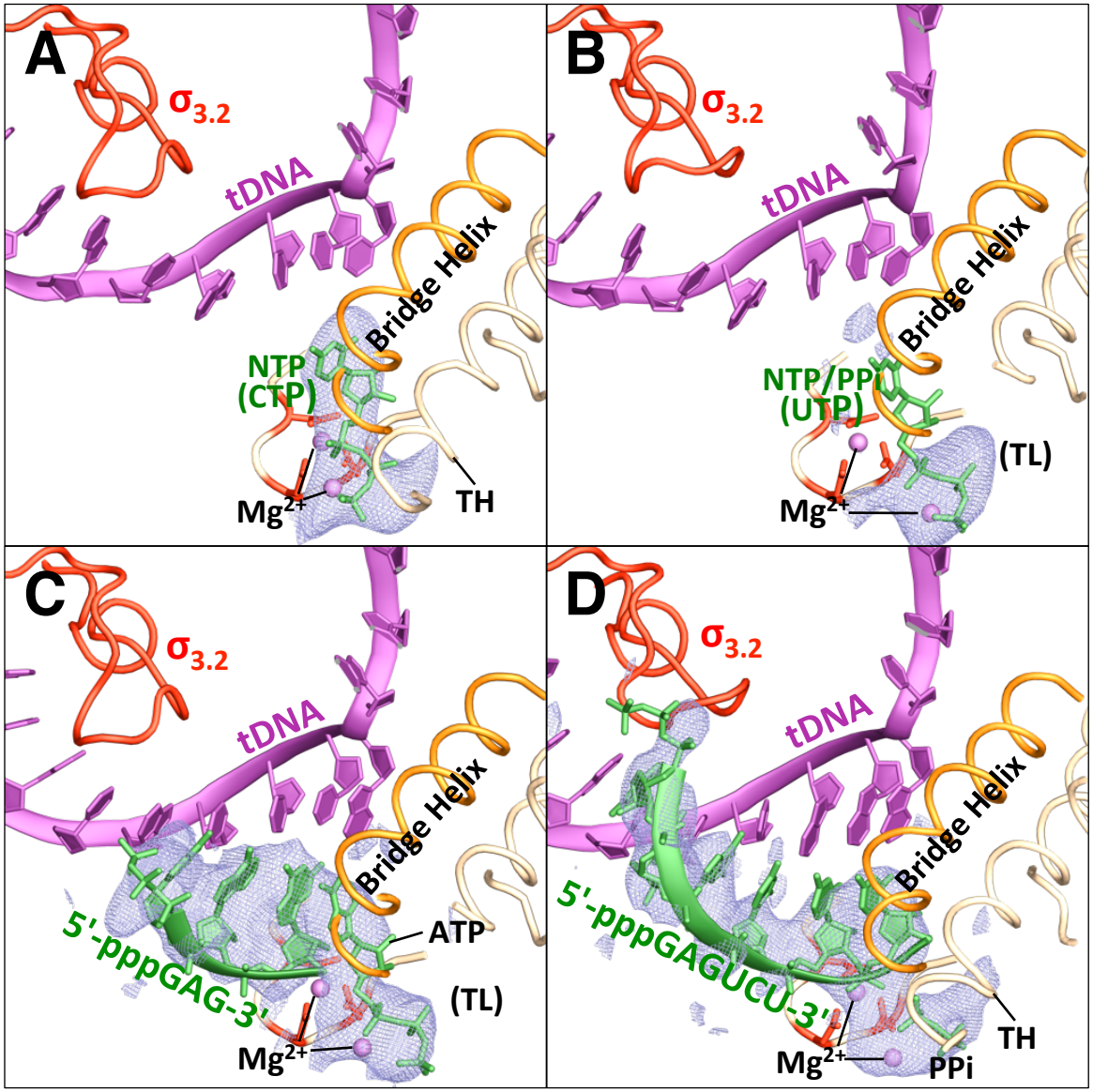


Fig. S1. RNA Synthesis and Cleavage by Soaking STIC4 Crystals with NTPs. Related to Figure 3. (A) CTP soaking. (B) UTP soaking. (C) ATP soaking. (D) Soaking with CTP, UTP, GTP, and ddATP for about 150 seconds.

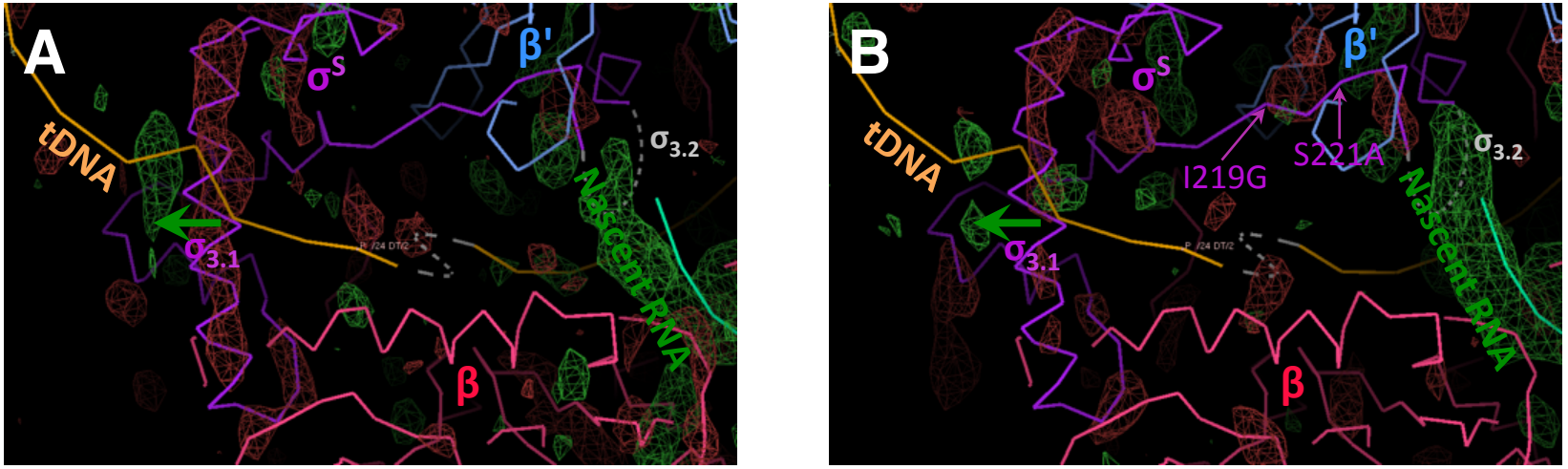


Fig. S2. Movement of the σ_3 Domain Associated with Nascent RNA Synthesis. Related to Figures 4 and 5. (A) RNA extension by soaking IGSA mutant STIC4 crystals with CTP, UTP, GTP, and $_{dd}$ATP. (B) *de novo* RNA synthesis by soaking WT SRPo crystals with UTP, CTP, $_{dd}$GTP, and RNA dinucleotide 5'-ApG-3'. The two image panels in this figure were screen-captured from coot with the F_o-F_o density maps contoured at ±3.0σ. The positive density is marked in green, and the negative density is marked in red.

TRANSPARENT METHODS

Preparation of the *E. coli* σ^S Factor

A DNA fragment encoding *E. coli* σ^S factor was amplified from the genomic DNA of *E. coli* K12 using DNA primers 5'-AAACCATGGGTCAGAATACGCTGAAAGTTCATG-3' and 5'-AAACTCGAGGAACAGCGCTTCGATATTCAG-3' and cloned into a pET21d plasmid between the NcoI and XhoI sites to form the pRS21h plasmid for σ^S expression. Two point-mutations (S2G and R329L) were introduced for cloning, and a 6xHis tag was engineered into the C-terminus of the σ^S gene to facilitate protein purification. This σ^S clone that carries a C-terminal 6xHis tag and the S2G and R329L mutations is considered the wild-type (WT) σ^S clone. To express the IGSA mutant σ^S factor, the I219G and S221A mutations were introduced to the pRS21h plasmid using DNA primers 5'–GGTGTCTACCGCGGTACCGCGCTCGTTAAGACGAAGCATACG-3' and 5'-AACGAGCGCGGTACCGCGGTAGACACCCCGCTGGGTGGTGATTCC-3'.

After transforming the expression vectors that carry the WT or IGSA mutant σ^S gene into *E. coli* Rosetta2(DE3)pLysS cells, overexpression of the σ^S protein was induced by addition of 1 mM isopropyl-β,D-thiogalactopyranoside (IPTG) at the late log phase. Cells were harvested after overnight induction at 30 °C and lysed using a continuous flow French press. WT and IGSA mutant σ^S proteins were purified with the same protocol that involves a series of Ni^{2+}-affinity, monoQ anion exchange and size-exclusion chromatography. Purified σ^S proteins (about 3 mg/ml) in the storage buffer (20 mM Tris-Cl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, and 0.1 mM EDTA) were flash frozen in liquid nitrogen or used directly for assembling the *E. coli* σ^S-RNAP holoenzyme.

Preparation of *E. coli* σ^S-RNAP Holoenzyme

The *E. coli* RNA polymerase core enzyme that lacks the C-terminal 94 amino acids of the α subunit (Δα236-329, or ΔαCTD) was prepared as described previously (Zuo and Steitz, 2015). The ΔαCTD-RNAP holoenzyme containing either WT or mutant σ^S factor (WT or IGSA mutant σ^S-RNAP) was prepared by mixing the purified ΔαCTD-RNAP core enzyme with the purified WT or IGSA mutant σ^S protein (about 1:3 molar ratio) at room temperature for 15 minutes followed by size exclusion chromatography to remove the extra σ^S protein. The holoenzymes were concentrated to about 30 mg/ml in the storage buffer (20 mM Tris-Cl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, and 0.1 mM EDTA) and stored in small aliquots at -80°C after flash freezing in liquid nitrogen.

Assembly and Crystallization of *E. coli* σ^S-Transcription Initiation Complex (STIC)

The STIC was assembled by incubating the σ^S-RNAP holoenzyme (5 mg/ml) with a preformed 50-bp DNA promoter (25 μM, non-template strand DNA sequence: 5'-ACCTTGACATCCCACCTCACGTATGCTATAATgtgtgcAGTCTGACGCGG-3', template strand DNA sequence: 5'-TCCGCGTCAGACTcgtaggATTATAGCATACGTGAGGTGGGATGTCAAGG-3') and 2 mM each of GTP, ATP and UTP at 37°C for 10 minutes as described previously (Liu et al., 2016). The storage buffer (20 mM Tris-Cl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$ and 0.1 mM EDTA) was also used as the reaction buffer for assembling the initiation complex. This reaction mixture was then mixed with the reservoir solution (1:1 in volume) and used for crystallization at room temperature by vapor diffusion. The reservoir solution contains 18% (wt/vol) PEG 3350, 120 mM NaCl, and 100 mM HEPES-Na (pH 8.0). σ^S-TIC crystals grew to full size in about 1 week and were kept in the crystallization drop before being harvested. At the time of crystal

harvest, STIC crystals were quickly washed with the mother liquor to get rid of unused NTPs and any free reaction products, and then cryo-protected in the cryo buffer containing 20% (wt/vol) PEG 3350, 15% (wt/vol) ethylene glycol, 120 mM NaCl, 5 mM $MgCl_2$, and 100 mM HEPES-Na (pH 8.0) by buffer exchange in multiple steps. Cryo-protected STIC crystals were then flash-frozen in liquid nitrogen directly or used in the soaking experiments before flash freezing. The soakings took about 30 minutes unless specified otherwise. "Fresh" crystals were typically harvested about 14 to 20 days after crystal setup; "old" crystals were typically harvested about 50 days after crystal setup.

Data Processing and Structural Determination of STICs

X-ray diffraction data were collected at 100 K at the beamlines 24-ID-C and 24-ID-E at Argonne National Laboratory, Chicago, IL. All data were integrated and scaled with XDS (Kabsch, 2010). The structures were solved by molecular replacement with PHASER (McCoy et al., 2007) using the previously determined crystal structure of the E. coli STIC (Liu et al., 2016) as the starting model. The molecular replacement solution was subjected to rigid body refinement with Refmac5 (Murshudov et al., 2011) using multiple rigid groups. After model building in Coot (Emsley and Cowtan, 2004), five cycles of TLS (translation libration screw-motion) and restrained refinement were performed using Refmac5 in the CCP4 suite (Winn et al., 2011). Data collection and structural refinement statistics are summarized in Table S1. Unless specified, the figures were created using PyMOL (Delano, 2002).

For all Fast Fourier Transform (fft) of the *Fo-Fo* difference maps, the structural factors from different datasets were first scaled using the SCALEIT program in the CCP4 suite (Winn et al., 2011) to assess the isomorphism between datasets. The scaling factor refinements were

performed using the default parameters of SCALEIT. The same set of phases extracted from a refined SRPo dataset (the second dataset in Table S1) was used for all the *Fo-Fo* map computations shown in the figures.

In-crystal RNA Synthesis and Cleavage

To investigate the enzymatic activity of the crystallized STICs, cryo-protected STIC crystals were immersed in soaking buffers at room temperature for 30 minutes or the specified time periods before being flash frozen without additional cryo buffer exchange. The cryo buffer [20% (wt/vol) PEG 3350, 15% (wt/vol) ethylene glycol, 120 mM NaCl, 5 mM $MgCl_2$, and 100 mM HEPES-Na (pH 8.0)] supplemented with various combinations of nucleotides was used as the buffers for various soaking experiments. The synthetic DNA promoter was designed to make it convenient to synthesize nascent RNAs (runoff transcript sequence 5'-GAGUCUGACGCGG-3') of various lengths up to about 10 nt by soaking STIC crystals with different nucleotide combinations. Where included, individual NTPs and dideoxy NTPs (ddNTPs) were used at a final concentration of 1 mM, and the dinucleotide (5'-GA-3' or 5'-AG-3') was used at a final concentration of 200 μM for all soaking experiments. Soaking experiments were typically carried out in 100 μl soaking buffers.

SUPPLEMENTAL REFERENCES

Delano, W.L. (2002). The PYMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography *60*, 2126-2132.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

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

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of applied crystallography *40*, 658-674.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta crystallographica Section D, Biological crystallography *67*, 355-367.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography *67*, 235-242.

Zuo, Y., and Steitz, T.A. (2015). Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. Mol Cell *58*, 534-540.