Structural and mechanistic basis of RNA extension in reiterative transcription initiation: RNA slipping without DNA scrunching

Yu Liu[1,2*], Jared T. Winkelman[1,2,3*], Libing Yu[1,2*,**], Chirangini Pukhrambam[2,3], Yu Zhang[4], Bryce E. Nickels[2,3,***] and Richard H. Ebright[1,2,***]

[1] Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA.

[2] Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ 08854, USA.

[3] Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.

[4] Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China, 200032

* equal contribution

** present address: Institute of Materials, China Academy of Physics, Mianyang, China 621907

*** corresponding authors

Impact statement: X-ray crystallography, protein-DNA-photocrosslinking, and single-molecule DNA nanomanipulation show that reiterative transcription initiation proceeds through RNA slipping, with sliding of RNA relative to DNA within a short, 2-3 bp, RNA-DNA hybrid.

Abstract

In standard transcription initiation, RNA polymerase (RNAP) binds to promoter DNA, unwinds promoter DNA, selects a transcription start site, and--using a "scrunching" mechanism, in which RNAP remains bound to the promoter, unwinds additional DNA, and pulls the additional unwound DNA past its active center, synthesizing an RNA product having a 5' sequence complementary to the DNA template. In an alternative pathway of transcription initiation, termed "reiterative transcription initiation," primarily observed at promoters containing homopolymeric sequences at or near the transcription start site, RNAP binds to promoter DNA, unwinds promoter DNA, selects a transcription start site, and--using a mechanism that has not previously been defined--generates an RNA product having a 5' sequence that contains a variable number of nucleotides not complementary to the DNA template. Here, using x-ray crystallography to define structures, using protein-DNA-photocrosslinking to map positions of RNAP leading and trailing edges relative to DNA, and using single-molecule DNA nanomanipulation to assess RNAP-dependent DNA unwinding, we show that RNA extension in reiterative transcription initiation (1) occurs without DNA scrunching, (2) involves a short, 2 bp (post-translocated state) to 3 bp (pre-translocated state) RNA-DNA hybrid, (3) and can involve an RNA product positioned as in standard transcription initiation and a DNA template strand positioned differently from standard transcription initiation. The results establish that, whereas RNA extension in standard transcription initiation proceeds through a scrunching mechanism, RNA extension in reiterative transcription initiation proceeds through a slippage mechanism, with sliding of RNA relative to DNA within a short, 2-3 bp, RNA-DNA hybrid.

Introduction

In standard transcription initiation, RNA polymerase (RNAP) holoenzyme binds to promoter DNA, unwinds ~13 base pairs (bp) of promoter DNA to form an RNAP-promoter open complex (RPo) containing a single-stranded "transcription bubble," selects a transcription start site, and synthesizes the first 10 nucleotides of the RNA product as an RNAP-promoter initial transcribing complex (RPitc) (Chen et al., 2021; Mazumder and Kapanidis, 2019; Ruff et al., 2015; Winkelman et al., 2021). Standard transcription initiation proceeds through a “DNA scrunching” mechanism, in which RNAP unwinds additional DNA, pulls the additional unwound DNA past its active center, and accommodates the additional unwound DNA as single-stranded bulges within the transcription bubble (Kapanidis et al., 2006; Margeat et al., 2006; Revyakin et al., 2006; Winkelman et al., 2015). During standard transcription initiation, each step of RNA extension involves (i) unwinding of 1 bp of DNA downstream of the RNAP active center, expanding the transcription bubble by 1 bp; (ii) translocation of DNA and RNA--together--by 1 bp relative to the RNAP active center by 1 bp; (iii) binding, through base pairing, of a complementary nucleoside triphosphate (NTP) to the DNA template strand in the RNAP active center, and (iv) phosphodiester bond formation, resulting in addition of a nucleotide to the RNA 3’ end (Winkelman et al., 2021). Standard transcription initiation yields an RNA product having a sequence fully complementary to the DNA template strand. Furthermore, the RNA product generated during standard transcription initiation remains fully base paired to the DNA template strand, as an RNA-DNA “hybrid.”

In an alternative pathway of transcription initiation, termed “reiterative transcription initiation,” “transcriptional stuttering,” or “pseudo-templated transcription,” an RNAP-promoter reiteratively transcribing complex (RPrtc) synthesizes an RNA product having a 5’-end sequence that contains a variable number of nucleotides not complementary to the DNA template (Jacques and Kolakofsky, 1991; Turnbough and Switzer, 2008; Turnbough Jr, 2011). Reiterative transcription initiation, which was first observed six decades ago (Chamberlin and Berg, 1962, 1964), competes with standard transcription initiation at many promoters, both *in vitro* and *in vivo* (Cheng et al., 2001; Guo and Roberts, 1990; Han and Turnbough, 1998; Jensen-MacAllister et al., 2007; Liu et al., 1994; Meng et al., 2004; Qi and

Turnbough, 1995; Severinov and Goldfarb, 1994; Tu and Turnbough, 1997; Vvedenskaya et al., 2015; Wagner et al., 1990). Reiterative transcription initiation is observed primarily at promoters that contain homopolymeric sequences at, or immediately downstream of, the transcription start site, resulting in low yields of standard, full-length RNA products at such promoters (Jacques and Kolakofsky, 1991; Turnbough and Switzer, 2008; Turnbough Jr, 2011). The extent of reiterative transcription initiation relative to standard initiation can change in response to changes in nucleoside triphosphate (NTP) concentrations. Accordingly, reiterative transcription initiation is often exploited as a regulatory mechanism in which changes in NTP concentrations lead to changes in gene expression (Turnbough, 2019; Turnbough and Switzer, 2008; Turnbough Jr, 2011). Classic examples of genes whose expression is regulated by reiterative transcription initiation are *Escherichia coli pyrBI* and *Bacillus subtilis pyrG* (Liu et al., 1994; Meng et al., 2004).

The mechanism of reiterative transcription initiation has not been firmly established. It has been hypothesized that reiterative transcription initiation involves an "RNA slipping" mechanism, in which RNA extension does not involve translocation of DNA relative to the RNAP active center, but, instead, involves translocation of RNA--"slippage"--relative to both DNA and the RNAP active center (Guo and Roberts, 1990; Jacques and Kolakofsky, 1991; Macdonald et al., 1993; Murakami et al., 2017; Severinov and Goldfarb, 1994; Shin et al., 2020; Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015; Wagner et al., 1990). However, direct evidence for RNA slipping, and information on the relationship between DNA scrunching and RNA slipping has not been presented.

Here, we report crystal structures of RNAP engaged in standard transcription initiation on a template containing a non-homopolymeric sequence and RNAP engaged in reiterative transcription initiation on templates containing template-strand GGG and CCC homopolymeric sequences. The structures reveal that, whereas RNA extension in standard transcription initiation involves DNA scrunching, RNA extension in reiterative transcription initiation does not. The structures further reveal that the path of the first five nucleotides of the RNA product relative to RNAP is the same in RPitc and RPrtc, but the path of the corresponding DNA-template-strand nucleotides differs in RPitc and RPrtc,

such that only two template-strand nucleotides (in a post-translocated state) or three template-strand nucleotides (in a pre-translocated state) are positioned to be able to base pair to the RNA product, resulting in a short RNA-DNA hybrid. Biochemical experiments using site-specific protein-DNA crosslinking and single-molecule DNA nanomanipulation verify that the differences in scrunching observed in the crystal structures of standard initiation and reiterative initiation also occur in solution. Taken together, our results establish that, whereas RNA extension in standard initiation involves DNA scrunching, RNA extension in reiterative initiation involves RNA slipping, with sliding of the RNA product relative to the DNA template strand within a short RNA-DNA hybrid.

Results and Discussion

Structures of RPrtc: RNA extension through RNA slipping without DNA scrunching

Crystal structures have been reported of RPrtc (Murakami et al., 2017; Shin et al., 2020). However, in those structures, (i) transcription-bubble nontemplate-strand DNA was extensively disordered (3- to 8-nt disordered), precluding assessment of the extent of DNA scrunching; (ii) the position of the RNA 3'-end relative to the DNA template strand did not permit further reiterative transcription (template-strand homopolymeric sequence not aligned with RNAP-active-center addition site), precluding assessment of the mechanism of RNA extension; and, (iii) in one case, there was an unusual and unexpected pattern of base pairing of DNA immediately downstream of the RNAP active center (Murakami et al., 2017; Shin et al., 2020). As a result of these issues, the previous crystal structures did not enable determination of the roles of DNA scrunching and RNA slipping in RNA extension in reiterative transcription initiation.

Here, we report crystal structures of *Thermus thermophilus* RPrtc in which nontemplate-strand DNA is fully ordered, enabling assessment of DNA scrunching, and in which the template-strand homopolymeric sequence is aligned with the RNAP-active-center addition site, enabling assessment of RNA slipping (Figures 1-2). We obtained a crystal structure of RPrtc containing a 5 nt RNA product by incubation of a nucleic-acid scaffold containing a $G_{+1}G_{+2}G_{+3}$ template-strand homopolymeric sequence, where +1 is the transcription start site, with RNAP and CTP (RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]; Figures 1C, S1), and we obtained a crystal structure of RPrtc containing a 4 nt RNA product by incubation of a nucleic-acid scaffold containing a $C_{+1}C_{+2}C_{+3}$ template-strand homopolymeric sequence with RNAP and GTP (RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]; Figures 1D, S1). For reference, we compared these structures to a previously reported crystal structure of *T. thermophilus* RPo (RPo; Figure 1A, Zhang et al., 2012)) and a crystal structure of *T. thermophilus* RPitc containing a 5 nt RNA product by incubation of a nucleic-acid scaffold lacking a template-strand homopolymeric sequence with RNAP, ATP, UTP, and CTP (RPitc,5; Figure 1B).

The crystal structure of RPo shows ordered density for all nucleotides of the transcription-bubble nontemplate strand: 5 nt in the -10 element (promoter element recognized by conserved region 2 of transcription initiation factor σ; Feklistov et al., 2014), 4 nt in the discriminator element (another promoter element recognized by conserved region 2 of transcription initiation factor σ; Feklistov et al., 2014), and 4 nt between the discriminator element and downstream dsDNA; Figure 1A, blue, light blue, and pink; (Zhang et al., 2012).

The crystal structure of RPitc,5 shows an initially transcribing complex with a 5 nt RNA product and an RNAP-active-center post-translocated state (Figure 1B). The RNA product is fully base paired to the DNA template strand as an RNA-DNA hybrid, with the RNA 3' nucleotide and the corresponding DNA template-strand nucleotide located in the RNAP-active-center product site ("P site"), and the next DNA template-strand nucleotide in the RNAP-active-center addition site ("A site") available for base pairing with an incoming NTP. The positions of the RNA and DNA relative to the RNAP active center indicate that, as compared to in RPo, 4 bp of downstream dsDNA have been unwound, 4 nt of each strand of has been translocated relative to the RNAP active center, and the RNA product has translocated 4 nt, in lock-step register with template-strand DNA. The crystal structure of RPitc,5 indicates that the 4 nt of nontemplate-strand DNA translocated relative to the RNAP active center are accommodated through DNA scrunching, with bulging of the nontemplate-strand DNA segment between the discriminator element and downstream dsDNA. Thus, the crystal structure of RPitc,5 shows ordered density for all transcription-bubble nontemplate-strand nucleotides of the -10 element and the upstream half of the discriminator element, with the same positions and the same σ-DNA interactions as in RPo, and shows disorder for 8 nt of nontemplate-strand DNA, corresponding to the downstream half of the discriminator element and DNA immediately downstream of the discriminator element (Figure 1B, cyan boxes). The 8 nt segment of disordered nontemplate-strand DNA in RPitc,5 has exactly the same endpoints, and spans exactly the same distance, as a 4 nt segment of ordered nontemplate-strand DNA in RPo (Figures 1A-B), suggesting that 4 nt of the 8 nt segment of disordered nontemplate-strand DNA are flipped out and/or

bulged out relative to the path of the nontemplate strand in RPo (Figure 1B, right, flipped and/or bulged-nucleotides in cyan box). The disorder of the 8 nt segment of disordered nontemplate-strand DNA indicates that the segment adopts an ensemble of distinct conformations (Figure 1B, cyan boxes). We conclude that formation of RPitc,5 involves 4 bp of DNA scrunching.

The crystal structure of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] shows a reiteratively transcribing complex with a 5 nt RNA product and an RNAP-active-center post-translocated state (Figure 1C). Only part of the RNA product--the part comprising the 3' nucleotide and the adjacent nucleotide--is base paired to the DNA template strand as a 2 bp RNA-DNA hybrid, with the RNA 3' nucleotide and the corresponding DNA template-strand nucleotide located in the RNAP-active-center P site, and the next DNA template-strand nucleotide in the RNAP-active-center A site available for base pairing with an incoming NTP. The positions of the RNA and DNA relative to the RNAP active center indicate that, as compared to in RPo, 1 bp of downstream dsDNA has been unwound, 1 nt of each strand of DNA has been translocated relative to the RNAP active center, and the RNA product has been translocated by 4 nt relative to the RNAP active center, translocating 1 nt in register with template-strand DNA and 3 nt out of register with template-strand DNA. The crystal structure of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] shows that the 1 nt of nontemplate-strand DNA translocated relative to the RNAP active center is accommodated through DNA scrunching, with unstacking and flipping of 1 nt of nontemplate-strand DNA between the discriminator element and downstream dsDNA (Figure 1C, cyan boxes). The crystal structure of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] shows ordered density for all transcription-bubble nontemplate-strand nucleotides, including the scrunched, unstacked, flipped nucleotide: 5 nt in the -10 element, 4 nt in the discriminator element, and 5 nt between the discriminator element and downstream dsDNA (Figure 1C). The crystal structure shows graphically that formation of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] involves 1 bp of DNA scrunching and 3 nt of RNA slipping.

The crystal structure of RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] shows a reiteratively transcribing complex with a 4 nt RNA product and an RNAP-active-center post-translocated state, a 2 bp RNA-DNA hybrid, 1 bp of unwinding of downstream dsDNA, 1 nt of translocation of each DNA strand, 1 nt of translocation of the

RNA product in register with template-strand DNA, and 2 nt of translocation of the RNA product out of register with template-strand DNA (Figure 1D). The crystal structure shows graphically that formation of RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] involves 1 bp of DNA scrunching and 2 nt of RNA slipping.

We infer, based on the structures in Figure 1, that standard transcription initiation to generate a post-translocated state of RPitc,x involves x-1 bp of DNA scrunching, whereas reiterative transcription initiation to form a post-translocated state of RPrtc,x involves 1 bp of DNA scrunching and x-2 nt of RNA slipping. Expressing these inferences in terms of mechanism, we infer that, in standard transcription initiation, following synthesis of a 2 nt initial RNA product, 1 bp of DNA scrunching occurs for each 1 nt of RNA extension (Revyakin et al., 2006), and we infer that, in contrast, in reiterative transcription initiation, following synthesis of a 2 nt initial RNA product, 1 bp of DNA scrunching occurs to position the 3' end of the initial RNA product in the RNAP-active-center P site, and no further DNA scrunching--just RNA slipping--occurs in RNA extension (see "Mechanistic Implications," below).

Structures of RPrtc: short RNA-DNA hybrid

In the crystal structure of RPitc,5, all nucleotides of the RNA product are complementary to, and base paired to, the DNA template strand, yielding a 5 bp RNA-DNA hybrid (Figures 1B and 2A). In contrast, in the crystal structures of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] and RPrtc,4 [$C_{+1}C_{+2}C_{+3}$], only the 2 nt at the 3' end of the RNA product are complementary to, and base paired to, the DNA template strand, yielding a 2 bp RNA-DNA hybrid (Figures 1C-D and 2A). We conclude that, unlike standard transcription initiation, reiterative transcription initiation involves a short RNA-DNA hybrid: a hybrid that is only 2 bp in length in the RNAP-active-center post-translocated state of the crystal structures, and that would be only 3 bp in length upon NTP binding and phosphodiester bond formation to yield the RNAP-active-center pre-translocated state.

In RPrtc, the conformation and interactions with RNAP of the RNA product--even the part of the RNA product that is not complementary to, and not base paired to, template-strand DNA--are the same as in RPitc (Figure 2B). In contrast, in RPrtc, the conformation and interactions of the part of the DNA

template strand that is not complementary to, and not base paired to, RNA differ from those in RPitc (Figure 2B). Inspection of structures of RPrtc and RPitc indicates that RNAP makes numerous interactions with RNA, and few interactions with DNA, in the RNAP hybrid-binding region, accounting for our observation that RNA conformation, rather than DNA conformation, is maintained upon loss of RNA-DNA complementarity and base pairing.

In our crystal structures of RPitc and RPrtc, the 5' end of the RNA product is in contact with the "σ finger" (also referred to as σ region 3.2), which enters the RNAP-active-center cleft, and obstructs the path of the RNA product (Figure 1, black asterisk; see, Basu et al., 2014; Li et al., 2020; Mazumder and Kapanidis, 2019; Mekler et al., 2002; Murakami et al., 2002; Vassylyev et al., 2002; Winkelman et al., 2021; Zhang et al., 2012). In solution, and in some crystal forms under some conditions, extension of the RNA product beyond a length of 5 nt can drive stepwise displacement of the σ finger (Li et al., 2020). In contrast, with the crystal forms and conditions of this work, RNA extension does not drive stepwise displacement of the σ finger--presumably due to crystal-lattice constraints on conformational change in σ--and thus RNA products are limited to a length of 5 nt.

Mapping of RNAP leading-edge and trailing-edge positions in RPrtc: RNA extension without DNA scrunching

DNA scrunching by RNAP has two biochemically detectable hallmarks: (i) downstream movement of the RNAP leading edge, but not the RNAP trailing edge, relative to DNA (Kapanidis et al., 2006; Winkelman et al., 2016a; Winkelman and Gourse, 2017; Winkelman et al., 2016b; Yu et al., 2017), and (ii) expansion of the transcription bubble (Revyakin et al., 2006; Yu et al., 2017). In a preceding section, we proposed that RNA extension in reiterative transcription does not involve DNA scrunching, except for 1 bp of DNA scrunching to position the 3' end of the initial RNA product in the RNAP active-center P site. As a first approach to test this proposal, we assessed positions of the RNAP leading edge and trailing edge relative to DNA during reiterative transcription initiation (Figure 3).

We used unnatural amino-acid mutagenesis to incorporate the photoactivatable amino acid p-benzoyl-L-phenylalanine (Bpa) at the RNAP leading edge and training edge, and we used protein-DNA photocrosslinking to define positions of the RNAP leading edge and trailing edge relative to DNA (methods essentially as in Winkelman et al., 2016b; Yu et al., 2017). We analyzed reiterative transcription initiation at derivatives of the bacteriophage N25 promoter containing either a template-strand $G_{+1}G_{+2}G_{+3}$ homopolymeric sequence (RPo [$G_{+1}G_{+2}G_{+3}$] and RPrtc [$G_{+1}G_{+2}G_{+3}$]) or a template-strand $C_{+1}C_{+2}C_{+3}$ homopolymeric sequence (RPo [$C_{+1}C_{+2}C_{+3}$] and RPrtc [$C_{+1}C_{+2}C_{+3}$]). For reference, we analyzed standard transcription initiation at the wild-type N25 promoter (RPo WT and RPitc WT). *In vitro* transcription experiments, carried out carried using the same reaction conditions, show production of long, up to at least 50-nt, reiterative transcripts with the $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ promoters in the presence of CTP and GTP, respectively, and show production of short, up to ~8 nt, standard transcripts with the WT promoter in the presence of ATP and UTP (Hsu et al., 2003; Revyakin et al., 2006; Figure S1).

The photocrosslinking results indicate that RPrtc exhibits an RNAP trailing-edge position that is unchanged as compared to RPo, exhibits an RNAP leading-edge position that is shifted downstream by 1 bp as compared to RPo of only 1 bp, and thus exhibits an RNAP trailing-edge/leading-edge distance that is increased by 1 bp as compared to RPo (Figure 3A). In contrast, RPitc exhibits an RNAP trailing-edge position that is unchanged as compared to RPo, exhibits an RNAP leading-edge position that is shifted downstream by up to 6 bp (range = 1-6 bp; mode = 3 bp) as compared to RPo, and thus exhibits an RNAP trailing-edge/leading-edge distance that is increased by up to 6 bp as compared to RPo (range of increase = 1-6 bp; modal increase = 3 bp; Figure 3A). An increase in RNAP trailing-edge/leading-edge distance is a defining hallmark of DNA scrunching (Kapanidis et al., 2006; Winkelman et al., 2016b; Yu et al., 2017). Thus, the results indicate that RPrtc engaged in synthesis of RNA products up to at least 50 nt in length exhibits only 1 bp of DNA scrunching, whereas RPitc engaged in synthesis of RNA up to 8 nt in length exhibits up to 6 bp of DNA scrunching. We conclude, consistent with the crystal structures of the preceding sections, that in reiterative transcription--following synthesis of a 2-nt initial RNA product and

1 bp of DNA scrunching to position the 3' end of the initial RNA product in the RNAP active-center P site--RNA extension does not involve DNA scrunching (Figure 3B).

Measurement of transcription bubble size in RPrtc: RNA extension without DNA scrunching

As a second approach to test the proposal that RNA extension in reiterative transcription does not involve DNA scrunching--except for 1 bp of DNA scrunching to position the 3' end of the initial RNA product in the RNAP active-center P site--we assessed transcription-bubble expansion during reiterative transcription initiation (Figure 4).

We used magnetic-tweezers single-molecule DNA-nanomanipulation to assess reiterative transcription (Figure 4A-B), analyzing the same promoter derivatives as in the preceding section (Figures 3, S1). The resulting transition amplitudes, transition-amplitude histograms, and RNAP-dependent DNA unwinding values show that RNAP-dependent DNA unwinding is greater by 1±0.4 bp in RPrtc engaged in synthesis of RNA products up to at least 50 nt in length than in RPo (Figure 4C-E). In contrast, RNAP-dependent DNA unwinding is greater by 5±0.4 bp in RPitc engaged in synthesis of RNA products up to ~8 nt than in RPo (Figure 4C-E). An increase in transcription-bubble size is a defining hallmark of DNA scrunching (Revyakin et al., 2006; Yu et al., 2017). Thus, the results from single-molecule DNA nanomanipulation indicate that RPrtc engaged in synthesis of RNA products up to at least 50 nt in length exhibits only 1 bp of DNA scrunching. We conclude, consistent with the conclusions of the preceding sections, that, in reiterative transcription--following synthesis of a 2-nt initial RNA product and 1 bp of DNA scrunching to position the 3' end of the initial RNA product in the RNAP active-center P site--RNA extension does not involve DNA scrunching.

Mechanistic implications

Taken together, our results from x-ray crystallography (Figures 1-2), site-specific protein-DNA photocrosslinking (Figure 3), and DNA single-molecule nanomanipulation (Figure 4) establish that, whereas standard transcription initiation involves 1 bp of DNA scrunching for each step of RNA

extension, reiterative transcription initiation at promoters containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences involves only 1 bp of DNA scrunching, irrespective of the number of steps of RNA extension. We conclude that, in reiterative transcription initiation at promoters containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences, following synthesis of the initial 2 nt RNA product (Figure 5, left, lines 1-2), 1 bp of DNA scrunching occurs to place the RNA 3' end in the RNAP active-center P site (Figure 5, left, lines 2-3), and no additional DNA scrunching occurs in RNA extension (Figure 5, right). We infer that, in reiterative transcription initiation, RNA extension does not involve movement of DNA relative to the RNAP active center, but, instead, involves RNA slipping, in which, in each step of RNA extension, RNA slips upstream by 1 nt relative to both template-strand DNA and the RNAP active center to place the RNA 3' end in the RNAP active-center P site (Figure 5, right). Thus, we infer that, in RNA extension in reiterative transcription initiation, the nucleotide addition cycle consists of RNA slipping to convert a pre-translocated state having a 3 bp RNA-DNA hybrid and having the RNA 3' end in the RNAP-active-center A site ("pre-slipped" state; Figure 5, left, RPitc,3 pre and Figure 5, right, RPrtc,n pre) into a post-translocated state having a 2 bp RNA-DNA hybrid and having the RNA 3' end in the RNAP-active-center P site ("post-slipped" state; Figure 5, right, RPrtc,n post), followed by NTP binding, phosphodiester bond formation, and pyrophosphate release (Figure 5, right). Thus, in contrast to standard transcription initiation, in which the RNA-DNA hybrid increases in length by 1 bp in each step of RNA extension, up to a length of ~10 bp, at which point promoter escape ensues (Hsu et al., 2003; Mazumder and Kapanidis, 2019; Ruff et al., 2015; Winkelman et al., 2021; Figure 5, left), in reiterative transcription initiation, the RNA-DNA hybrid does not extend beyond a 3 bp state, and, instead, alternates between a 3 bp pre-slipped state and a 2 bp post-slipped state in each step of RNA extension (Figure 5, right). RNA slipping by 1 nt breaks 1 bp of the RNA-DNA hybrid--specifically the upstream-most base pair of the RNA-DNA hybrid (Figure 5, right); breakage of the upstream-most base pair of the RNA-DNA hybrid occurs because slipping moves the RNA nucleotide that had been base paired to the upstream-most nucleotide of the template-strand homopolymeric sequence into alignment with a non-complementary DNA nucleotide upstream of the

template-strand homopolymeric sequence (Figure 5, right). According to this mechanism, the "branching point," or "decision point," between the standard-transcription-initiation and reiterative-transcription-initiation pathways occurs upon formation of RPitc,3 pre (Figure 5, lines 4-5; at this decision point, DNA scrunching followed by RNA extension yields standard transcription initiation, whereas RNA slipping followed by RNA extension yields reiterative transcription initiation (Figure 5, lines 4-5).

The mechanism defined in this work and set forth in Figure 5, requires a template-strand homopolymeric sequence at least 3 nt in length. This aspect of the mechanism is consistent with, and supported by, the observation from previous work that reiterative transcription initiation is efficient only at promoters containing template-strand homopolymeric sequences at least 3 nt in length (Cheng et al., 2001; Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015; Xiong and Reznikoff, 1993).

Prospect

The results and mechanism of this work, pertain to reiterative transcription initiation at promoters containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences (Figures 1-5). Previous work shows that reiterative transcription initiation also occurs efficiently at promoters containing template-strand $A_{+1}A_{+2}A_{+3}$ and $T_{+1}T_{+2}T_{+3}$ homopolymeric sequences (Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015) and at promoters containing $\geq$ 4 nt homopolymeric sequences (Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015). We consider it likely that the mechanism of Figure 5 also applies to these promoters. Previous work also shows that reiterative transcription initiation can occur at promoters containing more complex, non-homopolymeric repeat sequences (Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015). We consider it likely that a mechanism related to the mechanism of Figure 5, but with different extents of DNA scrunching and RNA slipping, could apply at these promoters. We note that the crystal-structure

determination, site-specific protein-DNA photocrosslinking, and single-molecule nanomanipulation procedures of this report could be used to define mechanisms of reiterative transcription at any promoter.

Reiterative transcription initiation can generate RNA products up to at least 50 nt in length (Turnbough and Switzer, 2008; Turnbough Jr, 2011; Figure S1). Because the volume of the RNAP active center cleft in RPrtc cannot accommodate more than ~12 nt of RNA, long RNA products generated by reiterative transcription initiation must exit from, and extend outside, the RNAP active-center cleft (Murakami et al., 2017). A key unresolved question is where long RNA products generated in reiterative transcription initiation exit the RNAP active-center cleft. In the crystal structures of this work, RNA products generated by reiterative transcription initiation were limited to 4-5 nt in length, because further RNA extension was blocked by the presence of the σ finger in the RNAP active-center cleft and by crystal-lattice constraints that prevented displacement of the σ finger from the RNAP active-center cleft (Figure 1). One hypothesis is that, in solution, complete displacement of the σ finger from the RNAP-active-center cleft channel could allow long RNA products generated in reiterative transcription initiation to exit the RNAP-active-center cleft through the RNAP RNA-exit channel, the same exit route used by RNA in standard transcription (Figure 5, right). Another hypothesis is that, in solution, a smaller conformational change in σ and/or RNAP could allow long RNA products generated in reiterative transcription initiation to exit the RNAP active-center cleft through a different route (Figure S2, left). Consistent with the first hypothesis, our crystal structures show that the 3' region of RNA products generated in reiterative transcription follow the same path relative to RNAP as in standard transcription initiation (Figure 2). However, arguing against the first hypothesis, it is unclear how, with only 1 bp of DNA scrunching (Figures 1, 3-5), the system could acquire the energy needed to drive complete displacement of the σ finger from the RNAP hybrid binding site and displacement of the σ region-3/region-4 linker from the RNAP RNA-exit channel (free energy that, in standard transcription initiation, is thought to be provided by ~8-10 bp of DNA scrunching; Kapanidis et al., 2006; Revyakin et al., 2006), and it is unclear how the σ region-3/region-4 linker could be completely displaced without triggering

promoter escape (which, in standard transcription initiation, is thought to be triggered by displacement of the σ region-3/region-4 linker; Li et al., 2020; Mekler et al., 2002; Murakami et al., 2002; Vassylyev et al., 2002). Cryo-EM structure determination, analyzing reiteratively transcribing complexes prepared in solution and flash frozen, may provide a means to resolve these questions.

Previous work has shown that complexes engaged in reiterative transcription initiation synthesizing RNA products up to at least 8 nt in length can switch from reiterative transcription initiation to standard transcription initiation, yielding productive complexes that escape the promoter and synthesize full-length RNA products comprising non-templated, reiterative-transcription-dependent-nucleotides at their 5' ends followed by templated, standard-transcription dependent nucleotides (Turnbough and Switzer, 2008; Turnbough Jr, 2011; Vvedenskaya et al., 2015). Key unresolved questions include how the switching occurs and where the RNA products that result from switching exit the RNAP active-center cleft. One hypothesis is that RNA products that result from switching exit the RNAP active-center cleft through the RNAP RNA-exit channel, with the templated RNA segment generated after switching to standard transcription initiation proceeding into, and through, the RNAP RNA-exit channel as in standard transcription, and pulling the 5' non-templated RNA segment behind it. According to this hypothesis the RNA product would proceed into, and through, the RNAP RNA-exit channel as an RNA loop (Figure S2, right). Cryo-EM structures of transcription complexes containing double-stranded RNA in the RNAP RNA-exit channel verify that the dimensions of the RNAP RNA-exit channel could accommodate such an RNA loop (Guo et al., 2018; Kang et al., 2018). Another hypothesis is that RNA products that result from switching exit the RNAP active-center cleft through a different route. According to this hypothesis, the RNA products that result from switching may diverge from the path of the RNA relative to RNAP in standard transcription initiation, and the RNA-DNA hybrid could be shorter than in standard transcription, potentially as short as the 2 bp to 3 bp RNA-DNA hybrids in reiterative transcription initiation. Cryo-EM structure determination, analyzing transcription elongation complexes containing 5' non-templated RNA segments

generated by switching from reiterative transcription initiation to standard transcription initiation, may provide a means to resolve this question.

Materials and Methods

Proteins

T. thermophilus RNAP core enzyme was prepared from *T. thermophilus strain* H8 (DSM 579; German Collection of Microorganisms and Cell Cultures GmbH), using cell lysis, polyethylenimine precipitation, ammonium sulfate precipitation, and ion-exchange chromatography on SP Sepharose FF (GE Healthcare Life Sciences, Inc.), Mono Q (GE Healthcare Life Sciences, Inc.), and Mono S HR (GE Healthcare Life Sciences, Inc.), as in (Maffioli et al., 2017; Zhang et al., 2012).

T. thermophilus σ^A was prepared by expression of a gene for N-terminally hexahistidine-tagged *T. thermophilus* σ^A in *E. coli*, followed by cell lysis, immobilized-metal-ion affinity chromatography on Ni-NTA agarose (Qiagen, Inc.), and anion-exchange chromatography on Mono Q (GE Healthcare Life Sciences, Inc.), as in (Zhang et al., 2012).

T. thermophilus RNAP σ^A holoenzyme was prepared by incubating 15 μM *T. thermophilus* RNAP core enzyme and 60 μM *T. thermophilus* σ^A in 2 ml 20 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 1% glycerol for 12 h at 4°C, followed by size-exclusion chromatography on Superdex 200 (GE Healthcare Life Sciences, Inc.), as in (Zhang et al., 2012).

Bpa-containing *E. coli* RNAP core enzyme derivatives RNAP-β'^{R1148Bpa} and RNAP-β'^{T48Bpa} were prepared from *E. coli* strain NiCo21(DE3) (New England Biolabs, Inc.) transformed with plasmid pEVOL-pBpF (Chin et al., 2002) and either plasmid pIA900-RNAP-β'^{R1148Bpa} (Winkelman et al., 2015) or plasmid pIA900-RNAP-β'^{T48Bpa} (Winkelman et al., 2015), using procedures as in (Winkelman et al., 2015). *E. coli* RNAP core enzyme without Bpa was prepared from *E. coli* strain NiCo21(DE3) (New England Biolabs, Inc.) transformed with plasmid pIA900 (Svetlov and Artsimovitch, 2015), using procedures as in (Winkelman et al., 2015).

E. coli σ^{70} was prepared from *E. coli* strain NiCo21(DE3) transformed with plasmid pσ^{70}-His (gift of J. Roberts; Marr and Roberts, 1997), using procedures as in (Marr and Roberts, 1997).

E. coli RNAP σ^{70} holoenzyme was prepared by incubating 1 μM *E. coli* RNAP core enzyme and 5 μM *E. coli* σ^{70} in 10 mM Tris-Cl (pH 8.0), 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol for 30 min at 25°C.

Oligonucleotides

Oligodeoxyribonucleotides (Table S2) were purchased from IDT, Inc.

Nucleic-acid scaffolds for structure determination

Nucleic-acid scaffolds for crystallization were prepared by incubating 20 nmol nontemplate-strand oligodeoxyribonucleotide and 22 nmol template-strand oligodeoxyribonucleotide in 50 μl 5 mM Tris-HCl (pH 7.9), 200 mM NaCl, and 10 mM $MgCl_2$ for 5 min at 95°C, followed by cooling to 25°C in 2°C steps with 1 min per step using a thermal cycler (Applied Biosystems, Inc.), and were stored at -80°C.

Structure determination: crystal growth and in crystallo RNA synthesis

RPo was prepared by mixing 20 μl 18 μM *T. thermophilus* RNAP σ^{A} holoenzyme in 20 mM Tris-HCl (pH 7.9), 100 mM NaCl, and 1% glycerol with 1.1 μl 0.4 mM nucleic-acid scaffold in 5 mM Tris-HCl (pH 7.9), 200 mM NaCl, and 10 mM $MgCl_2$, and incubating 1 h at 25°C.

Crystallization was performed essentially as in (Zhang et al., 2012). Crystallization drops containing 1 μl 20 μM RPo in 20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 1% glycerol and 1 μl reservoir solution [0.1 M Tris-HCl (pH 8.4), 180 mM KCl, 50 mM $MgCl_2$, and 9% (m/v) PEG 4000] were equilibrated with 400 μl reservoir solution in sealed hanging-drop plates. Crystals were grown 5 days at 22°C. Micro-seeding was performed as needed using the same reservoir solution.

In crystallo RNA synthesis was performed by adding 0.2 μl 10 mM each NTP (ATP, UTP, and CTP for RPitc,5; CTP for RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]; GTP for RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]) in reservoir solution

directly into crystallization drops at 22°C, incubating 2 min at 22°C, transferring crystals to 2 µl reservoir solution containing 18% (v/v) (2R, 3R)-(-)-2,3-butanediol (Sigma-Aldrich, Inc.) at 22°C, and flash-cooling in liquid nitrogen.

Structure determination: data collection and data reduction

Diffraction data were collected from cryo-cooled crystals at Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2 (RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] and RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]; Table S2) and at Argonne Photon Source (APS) beamline 19-ID (RPitc,5; Table S2). Data were processed using HKL3000 (Minor et al., 2006). Structures were solved by molecular replacement using the crystal structure of *T. thermophilus* RPo (PDB 4G7H) as search model. Iterative cycles of reciprocal refinement using Phenix (Liebschner et al., 2019) and real-space model building using Coot (Emsley et al., 2010) were performed. Improvement of the coordinate model resulted in improvement of phasing, and electron density maps for nucleic acids, which were not included in models at this stage of model building and refinement, improved over successive cycles. Nucleic acids then were built into the model and refined in stepwise fashion. The RPitc,5 model refined to R_{work} = 0.23 and R_{free} = 0.27, the RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] model refined to R_{work} = 0.24 and R_{free} = 0.29, and the RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] refined to R_{work} = 0.26 and R_{free} = 0.28. The final atomic models and structure factors were deposited in the Protein Data Bank (PDB) with accession codes RPitc,5 (7MLB), RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] (7MLI), and RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] (7MLJ; Table S2).

DNA templates for in vitro transcription and crosslinking assays

Linear DNA templates used for *in vitro* transcription (Figure S1) and *in vitro* photo-crosslinking assays (Figure 3) contain sequences from positions -72 to +71 of N25 or positions -72 to +71 of N25 promoter containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences (N25 [$G_{+1}G_{+2}G_{+3}$] and N25 [$C_{+1}C_{+2}C_{+3}$], respectively). Templates were generated by PCR in reactions containing 1 X Phusion HF Master Mix (Thermo Fisher Scientific, Inc.), 0.8 µM primer N25B, 0.8 µM

primer N25T, and ~1 pg of oligos JW173 (N25), JW169 (N25 [$G_{+1}G_{+2}G_{+3}$]), or JW170 (N25 [$C_{+1}C_{+2}C_{+3}$]). Reaction products were purified using a PCR purification kit (Qiagen, Inc.).

Nontemplate-strand sequence from positions -72 to +71 of N25:

5'-GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGT ATAATAGATTCATAATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAA TAGCCATCCCAATCGAACAG - 3'

Nontemplate-strand sequence from positions -72 to +71 of N25 [$G_{+1}G_{+2}G_{+3}$]:

5'-GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGT ATAATAGATTCCCCATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAA TAGCCATCCCAATCGAACAG-3'

Nontemplate-strand sequence from positions -72 to +71 of N25 [$C_{+1}C_{+2}C_{+3}$]:

5'-GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGT ATAATAGATTCGGGATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAA TAGCCATCCCAATCGAACAG-3'

In vitro transcription

Experiments in Figure S1 were performed using reaction mixtures (10 µl) containing 40 nM RNAP holoenzyme derivative and 4 nM linear promoter-containing template in 10 mM Tris-Cl (pH 8.0), 100 mM KCl, 10 mM $MgCl_2$, and 0.1 mg/ml bovine serum albumin (BSA). For reactions with N25, 200 µM ATP and 200 µM UTP along with 1µl 200 Bq/fmol [α^{32}P]-UTP (PerkinElmer, Inc.) were added to initiate RNA synthesis. For reactions with N25 [$G_{+1}G_{+2}G_{+3}$], 200 µM CTP along with 1µl 200 Bq/fmol [α^{32}P]-CTP (PerkinElmer, Inc.) was added to initiate RNA synthesis. For reactions with N25 [$C_{+1}C_{+2}C_{+3}$], 200 µM CTP along with 1µl 200 Bq/fmol [α^{32}P]-CTP (PerkinElmer, Inc.) was added to initiate RNA synthesis. After 20 min at 37°C, reactions were terminated by addition of 100 µl 10 mM EDTA (pH 8.0) and 1 mg/ml glycogen. Samples were extracted with acid phenol:chloroform (Ambion, Inc.), and RNA products were recovered by ethanol precipitation and re-suspended in 6.5 µl water. Samples were then

diluted 2-fold in urea loading dye (1 X TBE, 8 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue) and samples were analyzed by electrophoresis on 8 M urea, 1 X TBE polyacrylamide gels UreaGel System (National Diagnostics, Inc.), followed by storage-phosphor imaging (Typhoon 9400 variable-mode imager; GE Healthcare Life Sciences, Inc.). RNA sizes were determined by comparison to radiolabeled Decade Marker (ThermoFisher Scientific, Inc.).

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

In vitro photo-crosslinking and crosslink mapping experiments were done using procedures described in (Yu et al., 2017). For the experiments in Figure 3, 50 µl reactions containing 40 nM RNAP holoenzyme, 4 nM template, 100 µM ApU dinucleotide (for N25 RPo and RPitc samples), 100 µM CpC dinucleotide (for N25 [$G_{+1}G_{+2}G_{+3}$] RPo and RPrtc samples), or CpG dinucleotide (for N25 [$C_{+1}C_{+2}C_{+3}$] RPo and RPrtc samples) and 200 µM ATP and UTP (for for N25 RPitc samples), 200 µM CTP (for N25 [$G_{+1}G_{+2}G_{+3}$] RPrtc samples), or 200 µM GTP (for N25 [$C_{+1}C_{+2}C_{+3}$] RPrtc samples) and 1 X RB [10 mM Tris-Cl (pH 8.0), 70 mM NaCl, 10 mM $MgCl_2$, and 0.1 mg/ml BSA] were incubated for 5 min at 25°C, and UV irradiated for 2 min at 25°C in a Rayonet RPR-100 photochemical reactor equipped with 16 x 350 nm tubes (Southern New England Ultraviolet, Inc.).

To denature RNAP-DNA complexes, reactions were mixed with 15 µl 5 M NaCl and 6 µl 100 µg/ml heparin, incubated for 5 min at 95°C, and cooled to 4°C. Crosslinked RNAP-DNA complexes were isolated by adding 20 µl MagneHis Ni-particles (Promega, Inc.) equilibrated and suspended in 10 mM Tris-Cl (pH 8.0), 1.2 M NaCl, 10 mM $MgCl_2$, 10 µg/ml heparin, and 0.1 mg/ml BSA; MagneHis Ni-particles were collected using a magnetic microfuge tube rack; particles were washed with 50 µl 10 mM Tris-Cl (pH 8.0), 1.2 M NaCl, 10 mM $MgCl_2$, 10 µg/ml heparin, and 0.1 mg/ml BSA, washed twice with 50 µl 1 X *Taq* DNA polymerase buffer (New England Biolabs, Inc.), and particles (which contained bound RNAP-DNA complexes) were resuspended in 10 µl 1 X *Taq* DNA polymerase buffer. For leading-edge crosslink mapping, primer extension reactions (12.5 µl) were performed by combining 2 µl

of the recovered RNAP-DNA complexes, 1 µl of 1 µM ^{32}P-5'-end-labeled primer N25B [200 Bq/fmol; prepared using [γ^{32}P]-ATP (PerkinElmer, Inc.)] and T4 polynucleotide kinase (New England Biolabs, Inc.) as described in (Sambrook and Russell, 2006)], 1 µl 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM TTP, 0.5 µl 5 U/µl Taq DNA polymerase (New England Biolabs, Inc.), 5 µl 5 M betaine, 0.625 µl 100% dimethyl sulfoxide, and 1.25 µl 10 X *Taq* DNA polymerase buffer; 40 cycles of 30 s at 95°C, 30 s at 53°C, and 30 s at 72°C. For trailing-edge crosslink mapping, primer extension reactions (12.5 µl) were performed by combining 2 µl of the recovered RNAP-DNA complexes, 1 µl of 1 µM ^{32}P-5'-end-labeled primer N25T [200 Bq/fmol; prepared using [γ^{32}P]-ATP (PerkinElmer, Inc.)] and T4 polynucleotide kinase (New England Biolabs, Inc.) as described in (Sambrook and Russell, 2006)], 1 µl 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM TTP; GE Healthcare Life Sciences, Inc.), 0.5 µl 5 U/µl Taq DNA polymerase (New England Biolabs, Inc.), 5 µl 5 M betaine, 0.625 µl 100% dimethyl sulfoxide, and 1.25 µl 10 X *Taq* DNA polymerase buffer; 40 cycles of 30 s at 95°C, 30 s at 53°C, and 30 s at 72°C. Reactions were stopped by addition of 12.5 µl 1 X TBE, 8 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue; radiolabeled products were separated by electrophoresis on 10% 8M urea slab gels (equilibrated and run in 1 X TBE) and visualized by storage-phosphor imaging (Typhoon 9400 variable-mode imager; GE Healthcare Life Sciences, Inc.). Positions of RNAP-DNA crosslinking were determined by comparison to products of a DNA-nucleotide sequencing reaction generated using oligo N25B (when mapping trailing-edge crosslinks) or N25T (when mapping leading edge crosslinks) and a linear DNA template containing sequences from positions -72 to +71 of N25 (Thermo Sequenase Cycle Sequencing Kit; Affymetrix Inc.).

DNA templates for single-molecule DNA nanomanipulation

Plasmid pUC18-T20-N25-WT, which was used for experiments with the N25 promoter, was prepared as follows: The pUC18-T20 backbone was amplified by performing PCR using 1 ng of pUC18-T20 (Yu et al., 2017) in a 50 µl reaction containing 1 X Phusion HF Master Mix, 0.8 µM oligo

JW151 and 0.8 μM oligo JW152, and amplicons were digested with 30 units of DpnI (New England Biolabs, Inc.) for 30 min at 37°C. The N25 promoter was amplified by PCR using 1 ng JW173 in 50 μl reactions containing 1 X Phusion HF Master Mix, 0.8 μM oligo JW179, and 0.8 μM oligo JW173, and the amplicon was isolated using a PCR purification kit (Qiagen, Inc.). Gibson assembly was performed by combining 100 ng of DpnI-digested pUC18-T20 amplicon with 25 ng of the N25 promoter amplicon in 5 μl 2 X Gibson Assembly Master Mix (New England Biolabs, Inc.) in 10 μl reactions and incubating 15 min at 50°C. The ligation mixture was diluted 2-fold with ultrapure water and was used to transform NEB 5-alpha electrocompetent cells (New England Biolabs, Inc.). Transformants were plated on LB agar plates containing 100 μg/ml carbenicillin, and recombinant plasmid DNA was isolated from individual transformants.

Plasmids pUC18-T20-N25[$G_{+1}G_{+2}G_{+3}$]and pUC18-T20-N25[$C_{+1}C_{+2}C_{+3}$], which were used for experiments with the N25 [$G_{+1}G_{+2}G_{+3}$] promoter and the N25 [$C_{+1}C_{+2}C_{+3}$] promoter, were prepared in the same manner, but using JW169 and JW170, respectively, in place of JW173

2.0 kb DNA fragments carrying single centrally located N25, N25 [$G_{+1}G_{+2}G_{+3}$], or N25 [$C_{+1}C_{+2}C_{+3}$] promoters were prepared by digesting plasmid pUC18-T20C2-N25-WT, plasmid pUC18-T20C2-N25 [$G_{+1}G_{+2}G_{+3}$], or plasmid pUC18-T20C2-N25 [$C_{+1}C_{+2}C_{+3}$] with XbaI and SbfI-HF (New England Biolabs, Inc.), followed by agarose gel electrophoresis and isolation of DNA from the gel using a gel extraction kit (Qiagen, Inc.).

Starting from the above 2.0 kb pUC18 DNA segments, constructs for single-molecule DNA nanomanipulation were prepared via ligation, at the XbaI end, to a 1.0 kb DNA fragment carrying multiple biotin residues on both strands [generated by PCR amplification of plasmid pARTaqRPOC-*lac*CONS) (Revyakin et al., 2003) with primers RPOC3140 and XbaRPOC4050 (Table S2) as described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006), followed by digestion with XbaI (New England Biolabs, Inc.)] and, then, at the SbfI end, to a 1.0 kb DNA fragment carrying multiple digoxigenin residues on both strands [generated by PCR amplification of plasmid pARTaqRPOC-*lac*CONS (Yu et al., 2017) with primers RPOC820 and SbfRPOC50 (Table S2) as

described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006), followed by digestion with SbfI-HF (New England Biolabs, Inc.)]. The resulting constructs were linked to streptavidin-coated magnetic beads (1 µm diameter, MyOne Streptavidin C1; Life Technologies, Inc.), immobilized on a glass surface coated with anti-digoxigenin (prepared as in (Lionnet et al., 2012)) and calibrated as described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006).

Measurement of RNAP-dependent DNA unwinding by single-molecule DNA nanomanipulation: data collection

Experiments were performed essentially as described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006). We tested a 2.0 kb, mechanically extended, torsionally constrained, DNA molecule carrying N25 WT, N25 [$C_{+1}C_{+2}C_{+3}$], or N25 [$G_{+1}G_{+2}G_{+3}$] (superhelical density = 0.021 for experiments with positively supercoiled DNA; superhelical density = -0.021 for experiments with negatively supercoiled DNA; extension force = 0.3 pN). *E. coli* RNAP σ^{70} holoenzyme was used at 10 nM in experiments with positively supercoiled DNA and at 0.5 nM for experiments with negatively supercoiled DNA. The assay buffer contained 25 mM Na-HEPES (pH 7.9), 75 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mg/mL BSA, and 0.1% Tween-20. NTPs (ATP, GTP, UTP, and CTP; GE Healthcare Life Sciences, Inc.) were used at 0.5 mM. Experiments were performed at 30°C. For each promoter, all data were collected on a single DNA molecule. For experiments with negatively supercoiled DNA, which has $\bar{\tau}_{unwound}$ > 1 h (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006), the unwound DNA was mechanically disrupted after 1 min as described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006), by rotating the magnets 8 turns counterclockwise to introduce positive superhelical turns (superhelical density = 0.021), waiting for 1 min, and rotating magnets 8 turns clockwise to re-introduce negative superhelical turns (superhelical density = –0.021) and reform unwound complexes.

Measurement of RNAP-dependent DNA unwinding by single-molecule DNA nanomanipulation: data reduction

Raw time traces were analyzed to give DNA extension (l) as described (Revyakin et al., 2003; Revyakin et al., 2004, 2005; Revyakin et al., 2006). Changes in l resulting from DNA unwinding (Δl_u) were calculated as $\Delta l_u = (\Delta l_{obs,neg} + \Delta l_{obs,pos})/2$, where $\Delta l_{obs,pos}$ is the observed change in l in assays with positively supercoiled DNA, and $\Delta l_{obs,neg}$ is the observed change in l with negatively supercoiled DNA. Extents of DNA unwinding (in base pairs) were calculated as $(\Delta l_u / \delta)$ 10.4, where δ (55 nm) is the change in l per turn at the superhelical densities of this work (+/– 0.021), and 10.4 is the mean number of base pairs in each turn of B-DNA (Wang, 1979).

Statistical analysis

Data for single-molecule DNA nanomanipulation are means ± SEM of at least 50 technical replicates on a single, positively supercoiled DNA molecule, or at least 40 technical replicates on a single, negatively supercoiled DNA molecule.

Acknowledgements

Work was supported by National Institutes of Health grants GM118059 (BEN) and GM041376 (RHE) and National Natural Science Foundation of China grant 31822001 (YZ).

Competing interests

The authors declare no competing interests.

Author Contributions

Conceptualization, Y.L., J.T.W., L.Y., Y.Z., B.E.N., and R.H.E.; Investigation, Y.L., J.T.W., L.Y., C.P., and Y.Z.; Formal Analysis, Y.L., J.T.W., L.Y., and Y.Z.; Writing, Y.L., J.T.W., B.E.N., and R.H.E.; Visualization, Y.L., J.T.W., B.E.N. ,and R.H.E.; Supervision, B.E.N. and R.H.E.; Project Administration, B.E.N. and R.H.E.; Funding Acquisition, Y.Z., B.E.N. and R.H.E.

References

Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. J Biol Chem *289*, 24549-24559.

Chamberlin, M., and Berg, P. (1962). Deoxyribo ucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. Proc Natl Acad Sci USA *48*, 81-94.

Chamberlin, M., and Berg, P. (1964). Mechanism of Rna Polymerase Action: Characterization of the DNA-Dependent Synthesis of Polyadenylic Acid. J Mol Biol *8*, 708-726.

Chen, J., Boyaci, H., and Campbell, E.A. (2021). Diverse and unified mechanisms of transcription initiation in bacteria. Nat Rev Microbiol *19*, 95-109.

Cheng, Y., Dylla, S.M., and Turnbough, C.L., Jr. (2001). A long T•A tract in the *upp* initially transcribed region is required for regulation of *upp* expression by UTP-dependent reiterative transcription in *Escherichia coli*. J Bacteriol *183*, 221-228.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli.* Proc Natl Acad Sci USA *99*, 11020-11024.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

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.

Guo, H.C., and Roberts, J.W. (1990). Heterogeneous initiation due to slippage at the bacteriophage 82 late gene promoter *in vitro*. Biochemistry *29*, 10702-10709.

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

Han, X., and Turnbough, C.L., Jr. (1998). Regulation of *carAB* expression in *Escherichia coli* occurs in part through UTP-sensitive reiterative transcription. J Bacteriol *180*, 705-713.

Hsu, L.M., Vo, N.V., Kane, C.M., and Chamberlin, M.J. (2003). *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. Biochemistry *42*, 3777-3786.

Jacques, J.P., and Kolakofsky, D. (1991). Pseudo-templated transcription in prokaryotic and eukaryotic organisms. Genes Dev *5*, 707-713.

Jensen-MacAllister, I.E., Meng, Q., and Switzer, R.L. (2007). Regulation of *pyrG* expression in *Bacillus subtilis*: CTP-regulated antitermination and reiterative transcription with *pyrG* templates *in vitro*. Mol Microbiol *63*, 1440-1452.

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

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., 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 USA *117*, 5801-5809.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J.*, et al.* (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol *75*, 861-877.

Lionnet, T., Allemand, J.F., Revyakin, A., Strick, T.R., Saleh, O.A., Bensimon, D., and Croquette, V. (2012). Magnetic trap construction. Cold Spring Harb Protoc *2012*, 133-138.

Liu, C., Heath, L.S., and Turnbough, C.L., Jr. (1994). Regulation of *pyrBI* operon expression in *Escherichia coli* by UTP-sensitive reiterative RNA synthesis during transcriptional initiation. Genes Dev *8*, 2904-2912.

Macdonald, L.E., Zhou, Y., and McAllister, W.T. (1993). Termination and slippage by bacteriophage T7 RNA polymerase. J Mol Biol *232*, 1030-1047.

Maffioli, S.I., Zhang, Y., Degen, D., Carzaniga, T., Del Gatto, G., Serina, S., Monciardini, P., Mazzetti, C., Guglierame, P., Candiani, G.*, et al.* (2017). Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. Cell *169*, 1240-1248 e1223.

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

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

Mazumder, A., and Kapanidis, A.N. (2019). Recent Advances in Understanding σ^{70}-Dependent Transcription Initiation Mechanisms. J Mol Biol *431*, 3947-3959.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell *108*, 599-614.

Meng, Q., Turnbough, C.L., Jr., and Switzer, R.L. (2004). Attenuation control of *pyrG* expression in *Bacillus subtilis* is mediated by CTP-sensitive reiterative transcription. Proc Natl Acad Sci USA *101*, 10943-10948.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr *62*, 859-866.

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

Murakami, K.S., Shin, Y., Turnbough, C.L., Jr., and Molodtsov, V. (2017). X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway. Proc Natl Acad Sci USA *114*, 8211-8216.

Qi, F., and Turnbough, C.L., Jr. (1995). Regulation of *codBA* operon expression in *Escherichia coli* by UTP-dependent reiterative transcription and UTP-sensitive transcriptional start site switching. J Mol Biol *254*, 552-565.

Revyakin, A., Allemand, J.F., Croquette, V., Ebright, R.H., and Strick, T.R. (2003). Single-molecule DNA nanomanipulation: detection of promoter-unwinding events by RNA polymerase. Methods Enzymol *370*, 577-598.

Revyakin, A., Ebright, R.H., and Strick, T.R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. Proc Natl Acad Sci USA *101*, 4776-4780.

Revyakin, A., Ebright, R.H., and Strick, T.R. (2005). Single-molecule DNA nanomanipulation: improved resolution through use of shorter DNA fragments. Nat Methods *2*, 127-138.

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.

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

Sambrook, J., and Russell, D.W. (2006). The condensed protocols from Molecular cloning : a laboratory manual (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Severinov, K., and Goldfarb, A. (1994). Topology of the product binding site in RNA polymerase revealed by transcript slippage at the phage lambda PL promoter. J Biol Chem *269*, 31701-31705.

Shin, Y., Hedglin, M., and Murakami, K.S. (2020). Structural basis of reiterative transcription from the *pyrG* and *pyrBI* promoters by bacterial RNA polymerase. Nucleic Acids Res *48*, 2144-2155.

Svetlov, V., and Artsimovitch, I. (2015). Purification of bacterial RNA polymerase: tools and protocols. Methods Mol Biol *1276*, 13-29.

Tu, A.H., and Turnbough, C.L., Jr. (1997). Regulation of *upp* expression in *Escherichia coli* by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription. J Bacteriol *179*, 6665-6673.

Turnbough, C.L., Jr. (2019). Regulation of Bacterial Gene Expression by Transcription Attenuation. Microbiol Mol Biol Rev *83*.

Turnbough, C.L., and Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. Microbiol Mol Biol Rev *72*, 266-300.

Turnbough Jr, C.L. (2011). Regulation of gene expression by reiterative transcription. Curr Opin Microbiol *14*, 142-147.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature *417*, 712-719.

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.

Wagner, L.A., Weiss, R.B., Driscoll, R., Dunn, D.S., and Gesteland, R.F. (1990). Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. Nucleic Acids Res *18*, 3529-3535.

Wang, J.C. (1979). Helical repeat of DNA in solution. Proc Natl Acad Sci USA *76*, 200-203.

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 USA *113*, E1787-E1795.

Winkelman, J.T., and Gourse, R.L. (2017). Open complex DNA scrunching: a key to transcription start site selection and promoter escape. BioEssays *39*, 1600193.

Winkelman, J.T., Nickels, B.E., and Ebright, R.H. (2021). The transition from transcription initiation to transcription elongation: start-site selection, initial transcription, and promoter escape. In RNA Polymerase as a Molecular Motor, Second Edition, R. Landick, J. Wang, and T.R. Strick, eds. (Cambridge, UK: RSC Publishing).

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

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

Xiong, X.F., and Reznikoff, W.S. (1993). Transcriptional slippage during the transcription initiation process at a mutant *lac* promoter *in vivo*. J Mol Biol *231*, 569-580.

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

Zhang, 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.

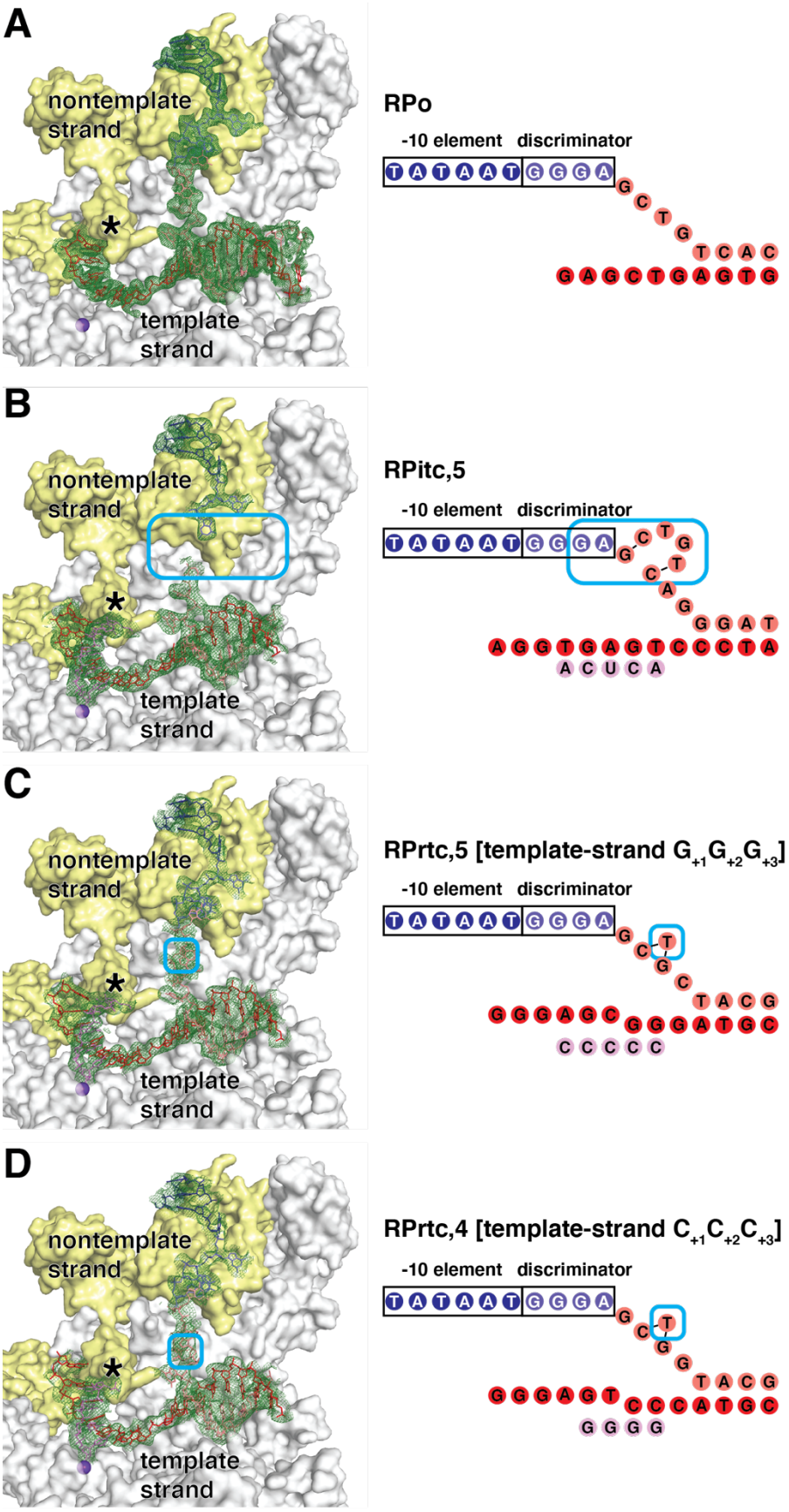

A
nontemplate strand
template strand
RPo
-10 element
discriminator
B
nontemplate strand
template strand
RPitc,5
-10 element
discriminator
C
nontemplate strand
template strand
RPrtc,5 [template-strand $G_{+1}G_{+2}G_{+3}$]
-10 element
discriminator
D
nontemplate strand
template strand
RPrtc,4 [template-strand $C_{+1}C_{+2}C_{+3}$]
-10 element
discriminator

Figure 1. Structures of RPrtc: RNA extension through RNA slipping without DNA scrunching

Crystal structures of transcription initiation complexes engaged in standard transcription initiation and reiterative transcription initiation. Left, experimental electron density (mFo-DFc; contoured at 2.0σ in A and 1.5σ in B-D) and atomic model, showing interactions of RNAP and σ with transcription-bubble nontemplate strand, transcription-bubble template strand, and downstream double-stranded DNA (RNAP β subunit and β' nonconserved domain omitted for clarity). Right, nucleic-acid scaffold. RNAP, gray; RNAP active-center catalytic Mg^{2+} ion, violet sphere; σ, yellow; σ finger, asterisk; -10 element of DNA nontemplate strand, dark blue; discriminator element of DNA nontemplate strand, light blue; rest of DNA nontemplate strand, pink; DNA template strand, red; RNA product, magenta. Cyan boxes in left and right subpanels, and bulged out nucleotides in right subpanels, indicate disordered nucleotides corresponding to scrunched, bulged out, nucleotides. Raised template-strand nucleotides in panels C-D indicate non-base-paired nucleotides.

(A) RNAP-promoter open complex (RPo; PDB 4G7H; (Zhang et al., 2012)).

(B) RNAP-promoter initial transcribing complex containing 5 nt RNA product generated by *in crystallo* standard transcription initiation (RPitc,5).

(C) RNAP-promoter reiteratively transcribing complex containing 5 nt RNA product generated by *in crystallo* reiterative transcription initiation on nucleic-acid scaffold having a template-strand $G_{+1}G_{+2}G_{+3}$ homopolymeric sequence (RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]).

(D) RNAP-promoter reiteratively transcribing complex containing 4 nt RNA product generated by *in crystallo* reiterative transcription initiation on nucleic-acid scaffold having a template-strand $C_{+1}C_{+2}C_{+3}$ homopolymeric sequence (RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]).

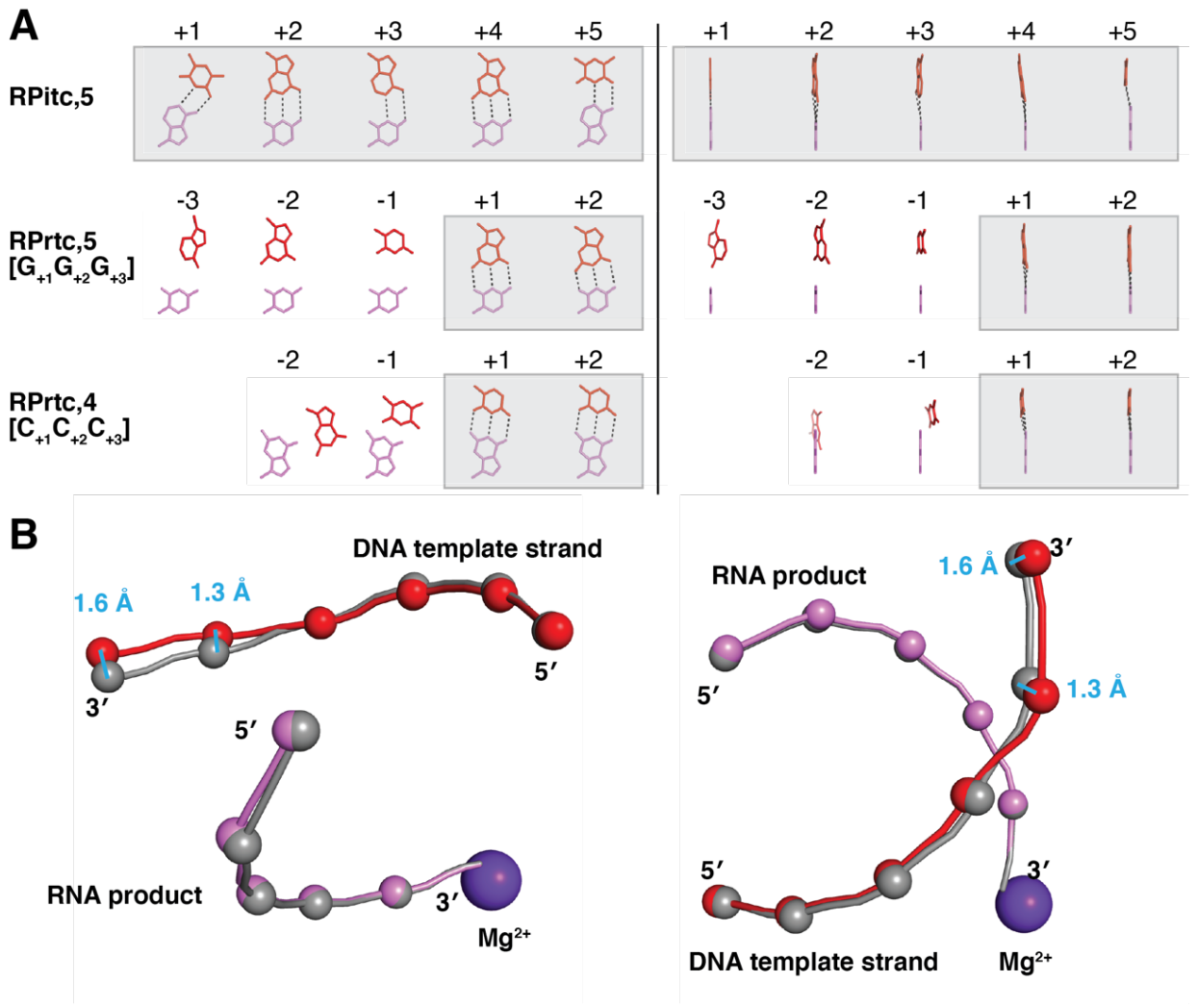


Figure 2. Structures of RPrtc: short RNA-DNA hybrid

(A) RNA-DNA base pairing in crystal structures of transcription initiation complexes engaged in standard transcription initiation (RPitc,5) and reiterative transcription initiation (RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] and RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]). Left, template-strand DNA bases (red) and corresponding RNA bases (magenta) in view orientation parallel to RNA-DNA-hybrid helix axis. Right, template-strand DNA bases (red) and corresponding RNA bases (magenta) in view orientation perpendicular to RNA-DNA-hybrid helix axis. Positions are numbered relative to the transcription start site, defined as position +1. Dashed lines indicate Watson-Crick hydrogen bonds. Shaded rectangles indicate RNA-DNA hybrids. At positions -1, -2, and -3

of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$], and at positions -1 and -2 of RPrtc,4 [$C_{+1}C_{+2}C_{+3}$], template-strand DNA bases are displaced relative to their locations in RPitc,5, and no base pairing occurs.

(B) Superimposition of DNA template strand and RNA of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] (red spheres, DNA phosphates; magenta spheres, RNA phosphates; violet sphere, RNAP active-center catalytic Mg^{2+} ion) on DNA template strand and RNA of RPitc,5 (gray spheres, DNA and RNA phosphates). Left, view orientation parallel to RNA-DNA-hybrid helix axis; right, view orientation perpendicular to RNA-DNA-hybrid helix axis. Distances in cyan, displacement of template-strand DNA nucleotides at positions -2 and -3 of RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] relative to their locations in RPitc,5.

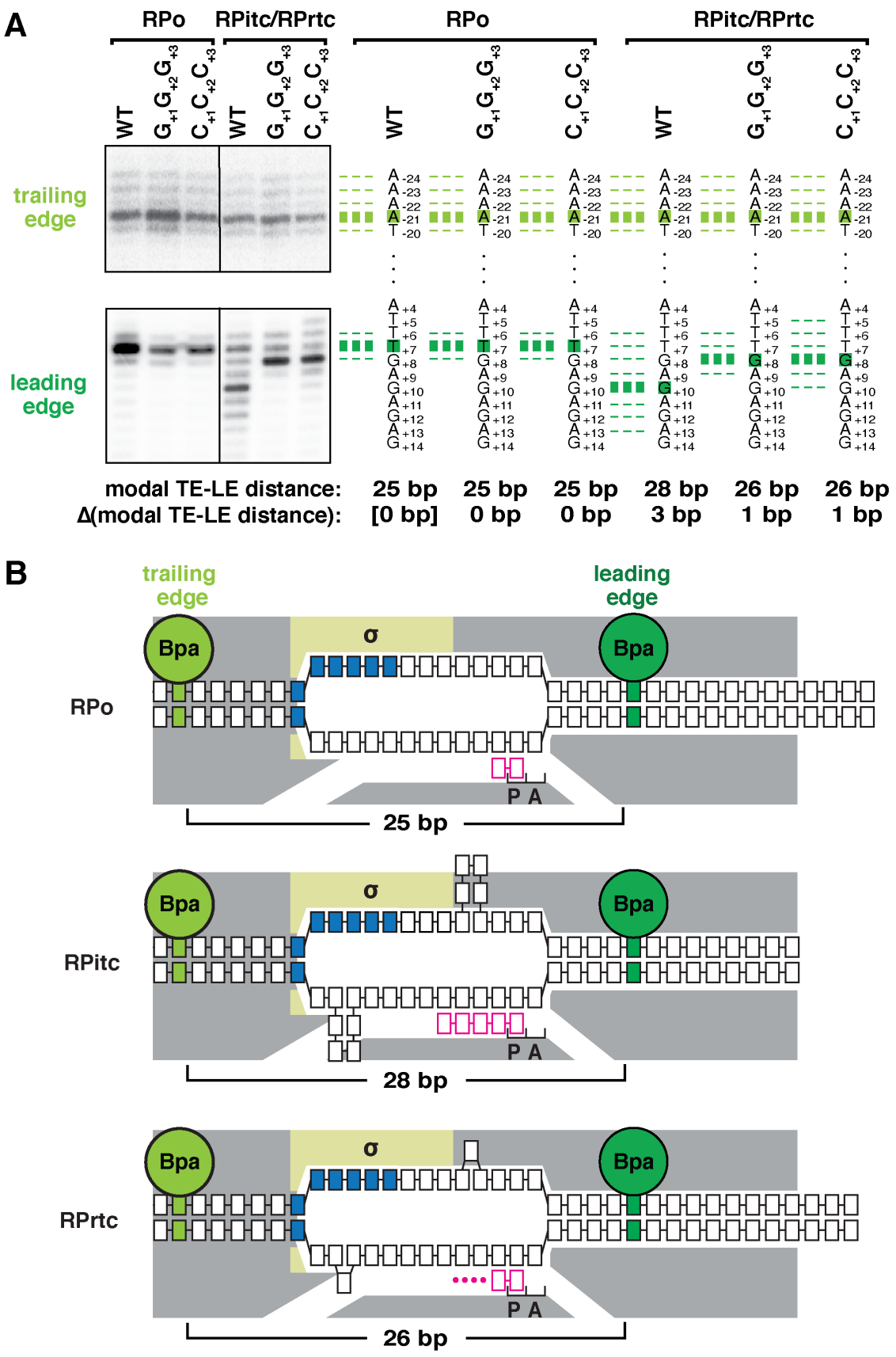
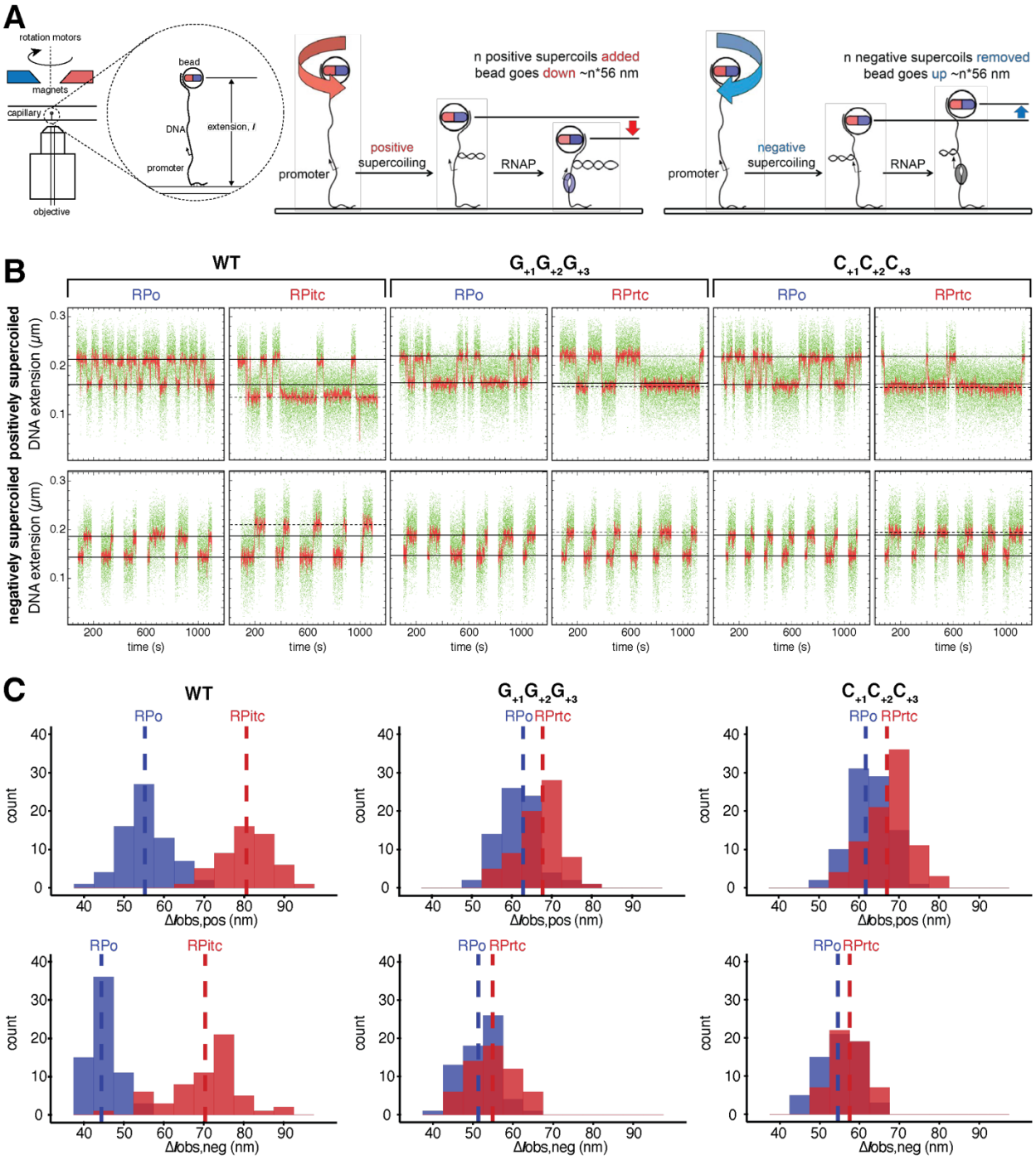

A
RPo
RPitc/RPrtc
RPo
RPitc/RPrtc
WT
G+1G+2G+3
C+1C+2C+3
trailing edge
leading edge
modal TE-LE distance: 25 bp 25 bp 25 bp 28 bp 26 bp 26 bp
Δ(modal TE-LE distance): [0 bp] 0 bp 0 bp 3 bp 1 bp 1 bp
B
trailing edge
leading edge
Bpa
σ
RPo
RPitc
RPrtc
P A
25 bp
28 bp
26 bp

Figure 3. Mapping of RNAP leading-edge and trailing-edge positions in RPrtc by use of protein-DNA photocrosslinking: RNA extension without DNA scrunching

(A) RNAP trailing-edge and leading-edge positions in transcription initiation complexes at the N25 promoter (WT) and derivatives of the N25 promoter containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences ($G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$). First bracketed subpanel, protein-DNA photocrosslinking data for RNAP-promoter open complex (RPo); second bracketed subpanel, protein-DNA photocrosslinking data for transcription initiation complexes engaged in standard transcription initiation (RPitc) and reiterative transcription initiation (RPrtc); third bracketed subpanel, interpretation of protein-DNA photocrosslinking data for RPo; fourth bracketed subpanel, interpretation of protein-DNA photocrosslinking data for RPitc and RPrtc. In third and fourth subpanels, promoter sequences are shown with positions numbered relative to the transcription start site; thick and thin light green lines indicate strong and weak RNAP-trailing-edge crosslinks (major crosslinking position marked in sequence with light green box); and thick and thin dark green lines indicate strong and weak RNAP-leading-edge crosslinks (major crosslinking position marked in sequence with dark green box). Bottom of panel, observed modal trailing-edge/leading-edge distances (modal TE-LE distance) and differences in modal TE-LE distance relative to modal TE-LE distance in RPo at wild-type N25 promoter [Δ(modal TE-LE distance)].

(B) Mechanistic interpretation of data in panel A. Three states are shown: RPo, RPitc [specifically RPitc having a 5 nt RNA product in a post-translocated state (RPitc,5 post), corresponding to the major crosslink in panel A], and RPrtc. Gray, RNAP; yellow, σ; light green, trailing-edge Bpa and nucleotide crosslink to trailing-edge Bpa; dark green, leading-edge Bpa and nucleotide crosslink to leading-edge Bpa; blue boxes, -10 element nucleotides; black boxes, DNA nucleotides (nontemplate-strand nucleotides above template-strand nucleotides; nucleotides downstream of -10 element numbered); magenta boxes, RNA nucleotides; magenta ellipsis, non-base-paired 5' RNA nucleotides; P and A, RNAP active-center product and addition sites. Scrunching of nontemplate and template DNA strands is indicated by bulged-out nucleotides. Initial-product formation in both standard transcription initiation and reiterative

transcription initiation involves one step of DNA scrunching. RNA extension in standard transcription initiation involves additional DNA scrunching, but RNA extension in reiterative transcription does not.



D

	RPo	RPitc (WT)	RPrtc ($G_{+1}G_{+2}G_{+3}$)	RPrtc ($C_{+1}C_{+2}C_{+3}$)
ΔΔlobs,pos (nm)	[0]	25±2	5±2	5±2
Δunwinding (bp)	[0]	5±0.4	1±0.4	1±0.4

Figure 4. Measurement of transcription bubble size in RPrtc by use of single-molecule DNA nanomanipulation: RNA extension without DNA scrunching

(A) Experimental approach (Revyakin et al., 2006; Yu et al., 2017). Left, apparatus. Middle and right, end-to-end extension (l) of a mechanically stretched, positively supercoiled (middle), or negatively supercoiled (right), DNA molecule is monitored. Unwinding of n turns of DNA by RNAP results in compensatory gain of n positive supercoils or loss of n negative supercoils, and movement of the bead by n*56 nm.

(B) Single-molecule time traces for RPo and RPitc at the N25 promoter (WT; left), and for RPo and RPrtc at derivatives of the N25 promoter containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences ($G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$; middle and right). Upper subpanels, positively supercoiled DNA; lower subpanels, negatively supercoiled DNA. Green points, raw data (30 frames/s); red points, averaged data (1 s window); horizontal black lines, unbound and RPo states; dashed horizontal black lines, RPitc and RPrtc states (with the difference in Δlobs between RPo and RPitc being substantially greater than the difference in Δl_{obs} between RPo and RPrtc).

(C) Single-molecule transition-amplitude histograms for RPo and RPitc at the N25 promoter (WT; left), and for RPo and RPrtc at derivatives of the N25 promoter containing template-strand $G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$ homopolymeric sequences ($G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$; middle and right). Upper subpanels, positively supercoiled DNA; lower subpanels, negatively supercoiled DNA. Vertical dashed lines, means; Δlobs,pos, transition amplitudes with positively supercoiled DNA; Δlobs,neg, transition amplitudes with negatively supercoiled DNA.

(D) Differences in $\Delta l_{obs,pos}$ and DNA unwinding relative to those in RPo at wild-type N25 promoter (means ± 2SEM).

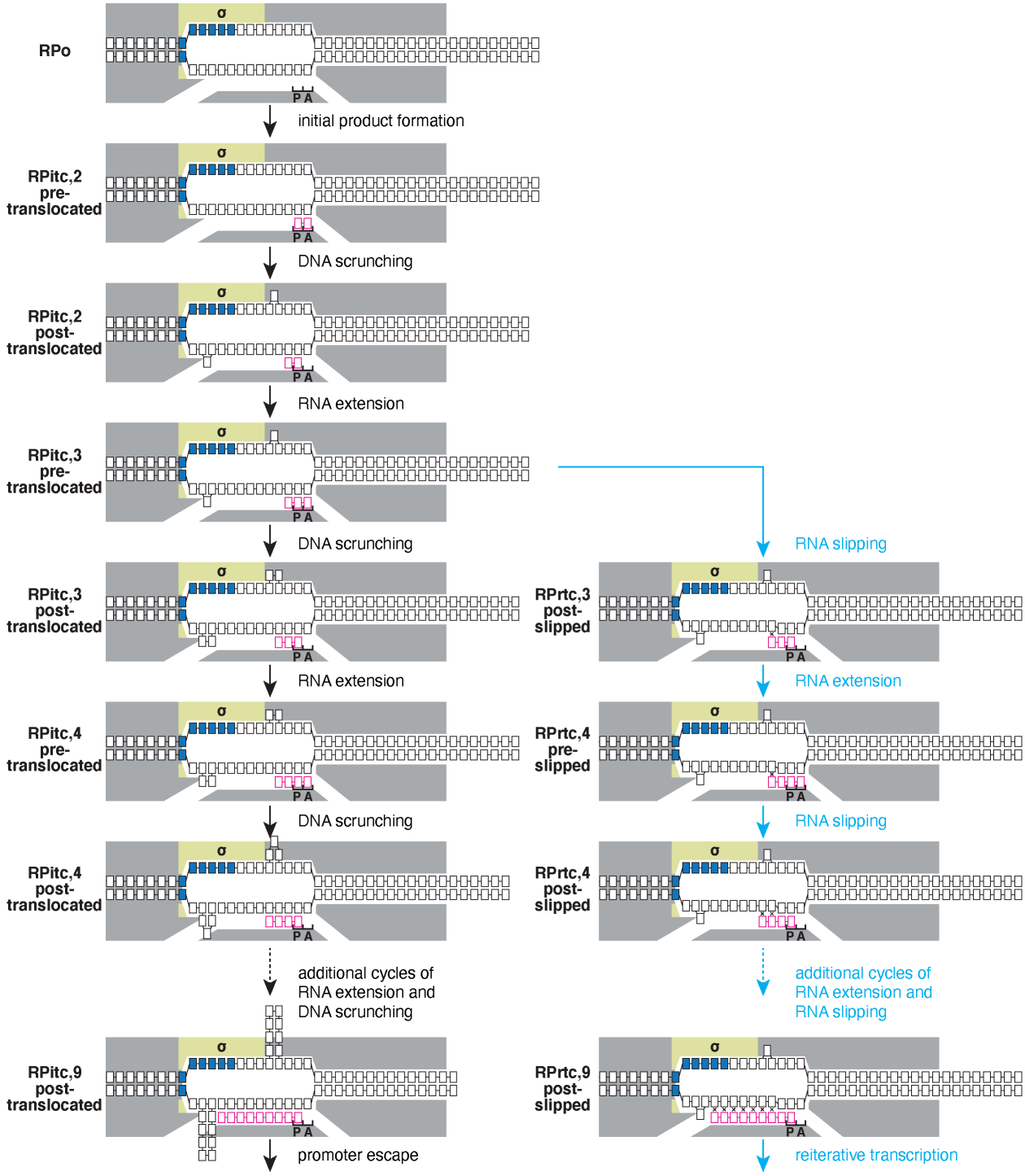


Figure 5. Mechanisms of standard transcription initiation and reiterative transcription initiation. Standard transcription initiation (left column) and reiterative transcription initiation (first four panels of left column followed by panels of right column). Cyan, reactions present only in reiterative transcription: i.e., cycles of RNA extension and slippage. Raised transcription bubble template-strand nucleotides and

Xs in right column indicate non-base-paired nucleotides. Other colors and symbols are as in Figure 3. Scrunching is indicated by bulged-out nucleotides (~8-10 scrunched base pairs prior to promoter escape in the standard transcription initiation pathway; 1 scrunched base pair in the reiterative transcription initiation pathway). Scrunched nucleotides of nontemplate and template DNA strands during initial transcription are accommodated as bulges within the unwound transcription bubble. Exact positions and conformations of scrunched nucleotides remain to be determined, and structural considerations suggest that exact positions and conformations of scrunched nucleotides of the template DNA strand change as RNA length increases during initial transcription. Here, for clarity, scrunched nucleotides are shown in arbitrary positions. Exact positions of 5' RNA nucleotides in RPrtc,9 post-slipped have not been established. Here, for clarity, 5' RNA nucleotides in RPrtc,9 post-slipped are shown in arbitrary positions. The σ finger partly occupies the RNAP hybrid binding site in initiation complexes, and the σ region-3/region-4 linker occupies the RNAP hybrid binding site in initiation complexes. Here, for clarity, the σ finger and the σ region-3/region-4 linker are omitted.

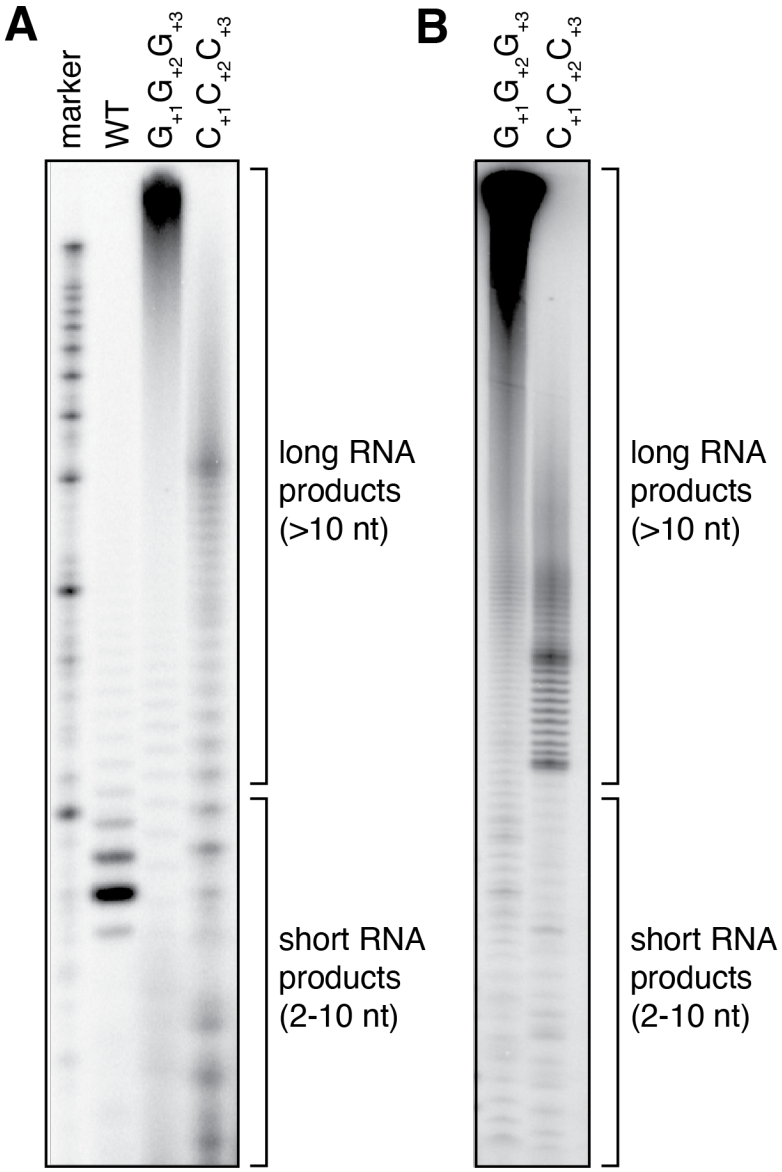


Figure S1. Standard transcription initiation and reiterative transcription initiation.

In vitro transcription experiments with promoter derivatives of this work.

(A) Bacteriophage T5 N25 promoter (WT; transcription with ATP and UTP; standard transcription initiation with ATP and UTP) and derivatives thereof having template-strand homopolymeric sequences ($G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$; transcription with CTP and GTP, respectively; reiterative transcription initiation) used in experiments of Figures 3-4. RNA products 2-10 nt in length are products of standard

transcription initiation or reiterative transcription initiation; RNA products greater than 10 nt in length are products of reiterative transcription initiation.

(B) Nucleic-acid scaffolds having template-strand homopolymeric sequences ($G_{+1}G_{+2}G_{+3}$ and $C_{+1}C_{+2}C_{+3}$; transcription with CTP and GTP, respectively; reiterative transcription initiation) used in structure determination of Figures 1-2.

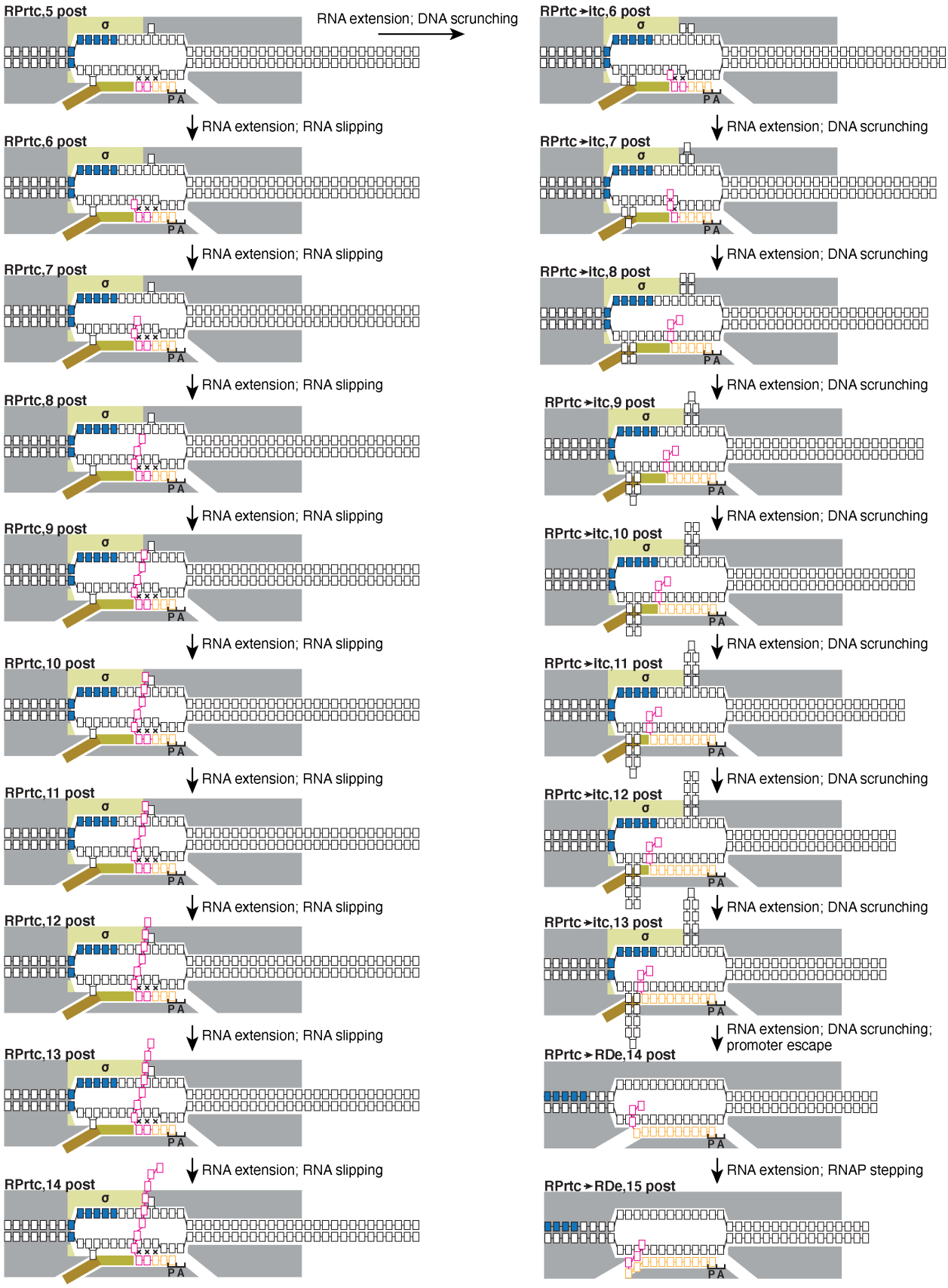

RPrtc,5 post
RNA extension; DNA scrunching
RPrtc→itc,6 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,6 post
RPrtc→itc,7 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,7 post
RPrtc→itc,8 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,8 post
RPrtc→itc,9 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,9 post
RPrtc→itc,10 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,10 post
RPrtc→itc,11 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,11 post
RPrtc→itc,12 post
RNA extension; RNA slipping
RNA extension; DNA scrunching
RPrtc,12 post
RPrtc→itc,13 post
RNA extension; RNA slipping
RNA extension; DNA scrunching;
promoter escape
RPrtc,13 post
RPrtc→RDe,14 post
RNA extension; RNA slipping
RNA extension; RNAP stepping
RPrtc,14 post
RPrtc→RDe,15 post
σ
P A

Figure S2. Hypothesized mechanisms for synthesis of long RNA products (>10 nt) through reiterative transcription initiation.

The part of the σ finger that occupies the RNAP hybrid binding site is colored dark yellow; the part of the σ region-3/region-4 linker that occupies the RNAP RNA-exit channel is colored brown. Non-templated, reiterative-transcription-dependent RNA nucleotides are colored magenta; templated, standard-transcription-dependent RNA nucleotides are colored orange. Other symbols and colors are as in Figure 5.

(Left) Hypothesized mechanism for synthesis of long RNA products comprising only non-templated, reiterative-transcription-dependent nucleotides ("non-productive" reiterative transcription initiation by RPrtc, without displacement of the σ finger, displacement of the σ region-3/region-4 linker, and promoter escape).

(Right) Hypothesized mechanism for synthesis of long RNA products comprising non-templated, reiterative-transcription-dependent nucleotides at their 5' ends followed by templated, standard-transcription-dependent nucleotides ("productive" reiterative transcription initiation, entailing switching from reiterative transcription initiation by RPitc to standard transcription initiation by RPrtc→itc, resulting in displacement of the σ finger, displacement of the σ region-3/region-4 linker, promoter escape, optional σ release, and transcription elongation using an RNAP-stepping mechanism).

Table S1. Data collection and refinement statistics.

Complex	**RPitc,5**	**RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]**	**RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]**
PDB code	7MLB	7MLI	7MLJ
Beamline	APS19-ID	SSRL-9-2	SSRL-9-2
Resolution range	50.00 - 3.60 (3.66 - 3.60)	50.00 - 3.60 (3.66 - 3.60)	50.00 - 3.75 (3.81 - 3.75)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	186.17 104.27 299.02 90 97.95 90	187.08 103.88 298.74 90 97.95 90	186.31 104.66 298.41 90 98.12 90
Unique reflections	66110 (3055)	64844 (3172)	57607 (2910)
Multiplicity	4.0 (3.3)	4.2 (3.9)	4.2 (4.2)
Completeness (%)	98.5 (92.3)	96.84 (91.38)	96.97 (93.60)
Mean I/sigma(I)	10.7	10.8	14.5
Wilson B-factor	116.36	107.75	114.24
R-merge	0.116 (0.845)	0.151 (1.36)	0.103 (1.031)
CC1/2	(0.624)	(0.912)	(0.837)
Reflections used in refinement	65151 (6097)	64469 (5989)	57263 (5466)
R-work/R-free	0.2330/0.2697	0.2432/0.2918	0.2654/0.2833
Number of non-hydrogen atoms	28400	28502	28489
RMS(bonds)	0.002	0.002	0.002
RMS(angles)	0.53	0.54	0.52
Ramachandran favored (%)	98.44	98.30	97.84
Ramachandran outliers (%)	0.09	0.14	0.12
Rotamer outliers (%)	6.37	6.37	5.63
Clashscore	6.16	6.41	6.22

Statistics for the highest-resolution shell are shown in parentheses.

Table S2. Oligonucleotides

Name	Sequence (5' to 3')	Description
N25T	GAGAGAGGTACCTCGAGG	PCR primer for production of DNA templates containing N25 promoter derivatives and for primer-extension mapping of leading-edge crosslinks
N25B	CTGTTCGATTGGGATGGCTATTCGG	PCR primer for production of DNA templates containing N25 promoter derivatives and for primer-extension mapping of trailing-edge crosslinks
JW151	GGTACCTCTGCTTCTCCTCGGGGATCACG	PCR primer for amplification of pUC18-T20 backbone
JW152	GGTACCTGGAGGAAGCCGACCGGAAGC	PCR primer for amplification of pUC18-T20 backbone
JW169	GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGTATAATAGATTCCCCATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAATAGCCATCCCAATCGAACAG	N25 [$G_{+1}G_{+2}G_{+3}$] template-strand oligonucleotide, used in PCR with primers JW179 and JW180 to produce N25 [$G_{+1}G_{+2}G_{+3}$] DNA fragment to insert into pUC18-T20
JW170	GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGTATAATAGATTCGGGATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAATAGCCATCCCAATCGAACAG	N25 [$C_{+1}C_{+2}C_{+3}$] template-strand oligonucleotide, used in PCR with primers JW179 and JW180 to produce DNA fragment to insert into pUC18-T20
JW173	GAGAGAGGTACCTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGTATAATAGATTCATAATTTGAGAGAGGAGTTTAAATATGGCTGGTTCTCGCGAGAATTCCGAATAGCCATCCCAATCGAACAG	N25 used in PCR with primers JW179 and JW180 to produce DNA fragment to insert into pUC18-T20
JW179	CGAGGAGAAGCAGAGGTACCCTGTTCGATTGGGATGGCTA	PCR primer for amplification of N25 and N25 derivative DNA fragments to insert into pUC18-T20

JW180	GTCGGCTTCCTCCAG GTACCGAGAGAGGTA CCTCGAGGGA	PCR primer for amplification of N25 and N25 derivative DNA fragments to insert into pUC18-T20
XbaRPOC4050	GAGAGTTCTAGAGAC CTTCTGGATCTCGTCC ACCAGG	PCR primer for production of 1 kb biotin-labelled DNA fragments for nanomanipulation experiments
RPOC3140	CTGATGCAAAAGCCC TCGGG	PCR primer for production of 1 kb biotin-labelled DNA fragments for nanomanipulation experiments
SbfRPOC50	GAGAGACCTGCAGGG AGAAGATCCGCTCCT GGAGCTACG	PCR primer for production of 1 kb digoxigenin-labelled DNA fragments for nanomanipulation experiments
RPOC820	TCCTGGCGCAGGTAG ATGAG	PCR primer for production of 1 kb digoxigenin-labelled DNA fragments for nanomanipulation experiments
RPitc,5-T	CCTGCATCCCTGAGT GGAGGG	template-strand oligonucleotide used in nucleic-acid scaffold for RPitc,5 crystallization
RPitc,5-NT	TATAATGGGAGCTGT CAGGGATGCAGG	nontemplate-strand oligonucleotide used in nucleic-acid scaffold for RPitc,5 crystallization
RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]-T	CCTGCATCCGTAGGG CGAGG	template-strand oligonucleotide used in nucleic-acid scaffold for RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] crystallization
RPrtc,5 [$G_{+1}G_{+2}G_{+3}$]-NT	TATAATGGGAGCTGC TACGGATGCAGG	nontemplate-strand oligonucleotide used in nucleic-acid scaffold for RPrtc,5 [$G_{+1}G_{+2}G_{+3}$] crystallization
RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]-T	CCTGCATCCGTGCCCT GAGG	template-strand oligonucleotide used in nucleic-acid scaffold for RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] crystallization
RPrtc,4 [$C_{+1}C_{+2}C_{+3}$]-NT	TATAATGGGAGCTGG CACGGATGCAGG	nontemplate-strand oligonucleotide, used in nucleic-acid scaffold for RPrtc,4 [$C_{+1}C_{+2}C_{+3}$] crystallization