**Title**
Single-molecule studies of different steps in human RNA polymerase II and bacterial RNA polymerase transcription



**Author**
Alhadid, Yazan Khalaf

**Publication Date**
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Single-molecule studies of different steps in human RNA polymerase II and bacterial RNA polymerase transcription

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

Doctor of Philosophy in Molecular, Cellular, and Integrative Physiology

by

Yazan Khalaf Alhadid

2018

ABSTRACT OF THE DISSERTATION

Single-molecule studies of different steps in human RNA polymerase II and bacterial RNA polymerase transcription

by

Yazan Khalaf Alhadid

Doctor of Philosophy in Molecular, Cellular, and Integrative Physiology

University of California, Los Angeles, 2018

Professor Shimon Weiss, Chair

Transcription of genomic DNA of all organisms is carried out by members of the multi-subunit RNA polymerase family. Regulation of RNA polymerase localization and activity underlies cellular homeostasis, division, and response to environmental cues. The catalytic mechanism, overall architecture, and many sequence and structural features of bacterial RNA polymerase are conserved in its Archaeal and Eukaryotic counterparts. The human RNA polymerase II (Pol II) is responsible for transcription of all protein-coding and many non-coding genes. The majority of current knowledge on RNA polymerases and their mechanism at

different steps in transcription derives from extensive work done using classical biochemical, genetic and structural biology methods. However, the use of single-molecule approaches addressed crucial questions on the function and mechanism of RNA polymerases during transcription, which were not possible to answer with ensemble-based approaches due to averaging effects. A useful fluorescence-based single-molecule technique to measure distances on the molecular scale and monitor dynamics is Förster resonance energy transfer (FRET). Here, I report on the development of diffusion-based single-molecule FRET (smFRET) methods to investigate different steps in transcription by the *in vitro* reconstituted human Pol II system. Using an assay that monitors the FRET changes between fluorescent dyes in the unwound region of promoter DNA (transcription bubble), I demonstrated the effect of certain components of the reconstituted system on the relative size of the transcription bubble. I also detail the optimizations done to enhance the affinity of single-stranded DNA (ssDNA) FRET probes to complementary target sequences. These ssDNA FRET probes were used to investigate the effect of certain components of the reconstituted system on Pol II activity by measuring the relative levels of RNA product. In addition to studies on the Pol II system, I report on the effect of the 5'-group of nascent RNA on the stability of the *Escherichia coli* RNA polymerase (RNAP) transcription bubble. I show how the presence of a 5'-monophosphate appears to destabilize the open bubble while a 5'-hydroxyl has no effect. Finally, I describe the work done on a project I took part in that identified a previously uncharacterized RNAP paused complex in initiation. We demonstrate that RNAP complexes undergoing initial transcription can enter the inactive paused state by backtracking. I also demonstrate how the presence of a 5'-triphosphate rapidly enhances entrance of RNAP complexes undergoing initial transcription into an inactive paused complex.

The dissertation of Yazan Khalaf Alhadid is approved.

Arnold J. Berk

Gregory A. Brent

Thomas M. Vondriska

Shimon Weiss, Committee Chair

University of California, Los Angeles

2018

*This dissertation is dedicated to my father Khalaf and mother Janet, who have always believed in my capabilities and encouraged me to be the best that I can be. Your support carried me through difficult times to reach this point. To my brother Qais, your support and advice helped me overcome many hurdles throughout my Ph. D.*

TABLE OF CONTENTS

Abstract of the dissertation……ii

Committee Page……iv

Dedication Page……v

Table of Contents……..vi

List of Figures……viii

Acknowledgments……x

Curriculum Vitae……...xii

Chapter 1: Plan of the dissertation……1

Chapter 2: Contributions of fluorescence-based single-molecule techniques to the study of transcription by bacterial RNA polymerase and human RNA polymerase II……...5

References……..31

Chapter 3: Development of diffusion-based single-molecule FRET methodologies to study different steps in human RNA polymerase II transcription…..…….46

References……..92

Chapter 4: Investigating the effect of different 5'-RNA moieties on *E. coli* RNA polymerase transcription initiation using single-molecule FRET……103

References……..120

Chapter 5: Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase……..122

References……..131

LIST OF FIGURES

Figure 2-1. High-resolution structure of Thermus aquaticus core RNAP……...28

Figure 2-2. Schematic of transcription cycle and σ70 interactions with promoter DNA……29

Figure 3-1. Effect of ssDNA structure on FRET probe hybridization…….69

Figure 3-2. Effect of different salts on RP1 probe hybridization to ssDNA……71

Figure 3-3. Effect of $MgCl_2$/heat denaturation on probe-target hybridization analyzed using EMSA……..73

Figure 3-4. Three proposed states in initiation for labeled promoter DNA…….74

Figure 3-5. More open state of transcription bubble is enhanced by TFIIH and by pre-incubating Pol II in buffer……..76

Figure 3-6. TFIIH appears to maintain a more open promoter; DSIF appears to promote re-annealing……..78

Figure 3-7. 6-mer RNA primer enhances transcriptional activity; MF and LF subpopulations exhibit millisecond dynamics……...80

Figure 3-8. *In vitro* smFRET-ALEX activity assay schematic…….82

Figure 3-9. Heteroduplex templates with super core promoter used for smFRET-ALEX activity assay…….84

Figure 3-10. Optimizing the activity of the minimal Pol II system; TFIIH enhances promoter escape……86

Figure 3-11. Development of smFRET-based Pol II pausing/pause release assay……..88

Figure 3-12. Studying the structural dynamics of the human Pol II transcription bubble in initiation……90

Figure 4-1. Unlabeled and Labeled DNA templates used for smFRET-ALEX assays……...112

Figure 4-2. Entrance kinetics into the RNAP-paused backtracked state is enhanced for 5'-triphosphate RNAs relative to 5'-OH RNAs……...114

Figure 4-3. 1D-FRET histograms showing the effect of 5'-monophosphate vs 5'-hydroxyl linear dinucleotides on the population of open complexes at different temperatures……116

Figure 4-4. 1D-FRET histograms showing the effect of 5'-monophosphate linear dinucleotides on the population of open complexes at different temperatures……..117

Figure 4-5. Fraction of open complexes decreases over time in the presence of 4 mM pApA at different temperatures; pApA results in dissociation of the open complex…….118

Figure 5-1. Single-round transcription quenched kinetics assay…….124

Figure 5-2. Quenched kinetics transcription results identify an initiation-related stalled state...125

Figure 5-3. GreA suppresses the kinetic delay in transcription initiation……126

Figure 5-4. Backtracking in initiation correlates with RNAP pausing in the presence of equimolar NTPs……127

Figure 5-5. A modified transcription initiation model…….128

# ACKNOWLEDGMENTS

I have been very fortunate as a Ph. D. student to have worked for and with many great individuals, to whom I am indebted, who have imparted numerous qualities that shaped my development into an independent researcher. First and foremost, I would like to express my sincere and profound gratitude to my mentor Dr. Shimon Weiss for inviting me to join his research group and entrusting me to lead an ambitious project. The door to Shimon's office was always open for me and other lab members to seek his advice or support on scientific and personal matters. He has instilled in me many lessons that I will carry with me for the rest of my life. The research I performed would not have been possible without the tremendous support of our collaborator Dr. Dylan Taatjes. Every aspect of my scientific development has been profoundly shaped by the example and mentorship of both Shimon and Dylan. I would like to thank our collaborator Dr. Sergei Borukhov for everything he has taught me. I would like to also give special thanks to members of my committee; Dr. Arnold J. Berk, Dr. Gregory A. Brent, and Dr. Thomas M. Vondriska for all their guidance and instruction. Finally, I would like to express my sincere gratitude to my co-workers and close friends, Dr. Sang Yoon Chung and Dr. Antonino Ingargiola, who have assisted me throughout many challenges as a graduate student.

Chapter 2 is primarily a reprint of the original review article published in Protein Science in July 2017 entitled "Studying transcription initiation by RNA polymerase with diffusion-based single-molecule fluorescence" (Volume 26, pages 1278-1290). As the first author of the review, the text and figures used are in compliance with the terms and conditions listed by Wiley author services. I would like to acknowledge the contributions of Sang Yoon Chung, Eitan Lerner, Dylan Taatjes, Sergei Burokhov, and Shimon Weiss for their help in writing and editing the manuscript. Additional contributions are mentioned in the published article.

Chapter 3 of this dissertation is unpublished work on the use of single-molecule fluorescence techniques to study the reconstituted human RNA polymerase II system. This work is done in collaboration with the Dylan Taatjes group at the University of Colorado, Boulder. I would like to acknowledge the work done by Benjamin Allen and Charli Fant in purifying all components of the reconstituted human transcription system and on providing technical advice on experiments. I would also like to acknowledge Sang Yoon Chung for his advice and assistance in performing experiments.

Chapter 4 details unpublished work investigating the effect of the moiety present at the 5'-end of RNA on the stability of the open complex and on entrance into a paused state in initiation during initial transcription. This work was done in consultation with Sergei Burokhov.

Chapter 5 is a reprint of the article published in PNAS in October 2016 entitled "Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase" (Volume 113, pages E6562-E6571). This project was spearheaded by the Eitan Lerner and Sang Yoon Chung who performed most of the experiments and analysis measuring activity from *E. coli* RNA polymerase. I assisted in performing experiments and in optimizing conditions for detecting transcripts with ssDNA FRET probes. Shuang Wang performed all experiments and analyzed magnetic tweezer data. Benjamin Allen performed assays monitoring the profile of abortive RNA products. Jookyung Lee purified GreA and performed experiments. Additional author contributions are mentioned in the article.

Part of the work in this dissertation was provided by the Jennifer S. Buchwald Graduate Fellowship in Physiology. I would also like to thank my undergraduate students Crystal Joh and Archana Sivanandam.

**Curriculum Vitae**
**Yazan Alhadid, Ph. D.**

## Research Experience

*University of California, Los Angeles (Los Angeles, CA)* June 2012 –
Graduate Student Researcher – Dr. Shimon Weiss, Molecular, Cellular & Integrative Physiology

*University of California, Los Angeles (Los Angeles, CA)* April 2010 – December 2011
Research Associate – Dr. Jeff Abramson, Dept. of Physiology

*University of California, Irvine (Irvine, CA)* September 2007 – December 2008
Graduate Student Researcher – Dr. Thomas Poulos, Dr. Shiou-Chuan Tsai, Dr. George Chandy, Biological Sciences

*Oregon Health and Science University (Portland, OR)* June 2006 – June 2007
Undergraduate Student Researcher – Dr. Nabil Alkayed, Department of Anesthesiology

*Portland State University (Portland, OR)* June 2004 – December 2004
Undergraduate Student Researcher – Dr. Dirk Iwata-Reuyl, Department of Chemistry

## Education

**University of California, Los Angeles**
Molecular, Cellular and Integrative Physiology Program (Ph. D.) 2018 (Expected)

**University of California, Irvine**
Biological Sciences (Master's degree) 2008

**Portland State University**
Chemistry: emphasis Biochemistry (Bachelor of Science) 2007

**Publication**

1. The effect of macromolecular crowding on single-round transcription by E. coli RNA polymerase. SangYoon Chung, Eitan Lerner, Yan Jin, Soohong Kim, **Yazan Alhadid**, Logan Wilson Grimaud, Irina X. Zhang, Charles M. Knobler, William M. Gelbart, Shimon Weiss, Nucleic Acids Research. (Accepted)

2. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. Eitan Lerner, Thorben Cordes, Antonino Ingargiola, **Yazan Alhadid**, SangYoon Chung, Xavier Michalet, Shimon Weiss. *Science* 359(288):1-12 **(2018)**

3. Studying transcription initiation by Escherichia coli RNA polymerase with single molecule fluorescence. **Yazan Alhadid**, SangYoon Chung, Eitan Lerner, Dylan J. Taatjes, Sergei Borukhov, Shimon Weiss. *Protein Science*. 26(7):1278-1290 **(2017)**

4. Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. Eitan Lerner, SangYoon Chung, Benjamin L. Allen, Shuang Wang, Jookyung Lee, Shijia W. Lu, Logan W. Grimaud, Antonino Ingargiola, Xavier Michalet, **Yazan Alhadid**, Sergei Borukhov, Terence R. Strick, Dylan J. Taatjes, Shimon Weiss. *Proc. Natl. Acad. Sci.* 113(43):E6562-E6571 **(2016)**

5. Soluble Epoxide Hydrolase Gene Deletion Is Protective Against Experimental Cerebral Ischemia. Wenri Zhang, Takashi Otsuka, Nobuo Sugo, Ardi Ardeshiri, **Yazan K. Alhadid**, Jeffrey J. Iliff, Andrea E. DeBarber, Dennis R. Koop, and Nabil J. Alkayed . *Stroke* 39:2073-2078 **(2008)**

**Scientific Conferences**

- FASEB Science Research Conferences "Mechanism and Regulation of Prokaryotic Transcription (2017, Saxtons River, VT)
- 29th Annual Symposium of The Protein Society (2015, Barcelona, Spain)
- Biophysical Society Conference (2016, Los Angeles, CA)

**Awards and Honors**

- Jennifer S. Buchwald Graduate Fellowship in Physiology, Summer 2012
- Diversity Recognition Scholarship, Portland State University
- Dean's List, Portland State University, Fall 2004 and Winter 2005

# Chapter 1:

Plan of the dissertation

The topics presented in this dissertation consist of the application of a single-molecule fluorescence based spectroscopy technique to the study of the *Escherichia (E.) coli* and human transcription systems reconstituted *in vitro*. The primary method used is based on the phenomenon of Förster resonance energy transfer (or FRET), which is a widely used technique at both the ensemble and single-molecule (smFRET) scales. Gene transcription is a fundamental process in every domain of life that uses a multi-subunit RNA polymerase enzyme to generate RNAs that are identical in sequence to a transcribed DNA region or gene. The evolutionarily related *E. coli* RNA polymerase (RNAP) and human RNA polymerase II (Pol II) are both responsible for the transcription of all protein-encoding genes and many non-coding genes as well for Pol II.

In chapter 2, I briefly highlight certain contributions of smFRET to the study of different steps in RNAP and Pol II transcription.

Chapter 3 details the development of new smFRET methods for studying different steps in transcription by the reconstituted human Pol II system. This work was done in collaboration with Dr. Benjamin Allen and Charli Fant from the laboratory of Dr. Dylan Taatjes (University of Colorado, Boulder). I describe the optimization of the affinity of single-strand DNA (ssDNA) FRET probes to complementary ssDNA and RNA target by reducing secondary structures in probe and target and by measuring probe hybridization in buffer containing different concentrations and types of salts. Also, using the successfully reconstituted Pol II system, we identified distinct conformations in the transcription bubble at different steps in transcription initiation. We then were able to detect transcripts generated by a minimal set of factors that form the pre-initiation complex using a ssDNA FRET probe. In both the DNA conformation assay and the RNA detection assay, TFIIH enhanced the activity when ribonucleotide triphosphate

(NTPs) substrates were added. Finally, the progress towards measuring Pol II pausing and pause escape using smFRET are mentioned. I also provide my thoughts on the future experiments that could be done.

The focus of chapter 4 is on studying the effects of different 5' RNA moieties on RNAP transcription initiation. The propensity for RNAP to pause and backtrack during initial transcription was evaluated for transcripts initiated with dinucleotides containing either a 5'-triphosphate or a 5'-hydroxyl. I show that the kinetics of entering into the paused state is significantly faster with the 5'-triphosphate. I then demonstrate the effect of 5'-monophosphate dinucleotides (pNpNs) on the stability of RNAP open complexes by monitoring the conformation of the transcription bubble with smFRET. The fraction of RNAP open complexes decreases in the presence of the pNpNs and is more pronounced at an elevated temperature of 37 °C, while no such effect was observed for dinucleotides with a 5'-hydroxyl. This work was done in consultation with Dr. Sergei Borukhov's (Rowan University), who performed some analogous experiments using the permanganate footprinting assay to monitor the fraction of open complexes.

In chapter 5, the identification and characterization of a previously hypothesized paused RNAP complex in transcription initiation is detailed. I participated in a collaborative effort between our group and the Dr. Dylan Taatjes, Dr. Sergei Borukhov, and Dr. Terence Strick (Institut Jacques Monod) groups. I participated with Dr. SangYoon Chung, Dr. Eitan Lerner, Shijia Lu and Logan Grimaud in performing a quenched kinetics assay to measure RNA production with smFRET. I also helped in optimizing the conditions to enhance the affinity of FRET probes to targets with Dr. SangYoon Chung. Dr. Benjamin Allen from the Taatjes group demonstrated using radioactive polyacrylamide gel electrophoresis the profile of abortive RNAs

in the presence and absence of GreA (prepared by Dr. Jookyung Lee from the Borukhov group), a factor that regenerates active RNAP complexes by cleaving the extruded portion of backtracked RNAs. Finally complementary single molecule assay using magnetic tweezers done by Dr. Shuang Wang (Dr. Terenece Strick group) identified the long-lived paused initiation complex in the presence of all NTPs at physiological concentrations.

# Chapter 2:

Contributions of fluorescence-based single-molecule techniques to the study of transcription by bacterial RNA polymerase and human RNA polymerase II

**Introduction:**

Transcription is used by all living organism to convert the DNA sequence information in genes to complementary RNA molecules. This process is carried out by the multi-subunit RNA polymerase (msRNAP) family of proteins, members of the two-barrel nucleic acid polymerase superfamily (*1*), found in all domains of life. All msRNAPs have a similar overall shape, resembling a crab-claw, with many structural and sequence features, especially around the active site, conserved through evolution (Figure 1A) (*1*–*5*). The ~400 kDa bacterial RNAP (RNAP) core enzyme is composed of five subunits (αI, αII, β', β, ω) that can be grouped into three functional classes: assembly platform subunits that nucleate RNAP assembly (αI, αII), catalytic subunits (β', β), and an auxiliary function subunit (ω) (*3*, *6*). Each subunit of RNAP has an analogous evolutionarily-conserved counterpart in the larger msRNAPs, from Archaea and Eukarya (Pol I, II, III), that occupy relatively similar positions within the complex (*2*, *4*, *5*). The three different eukaryotic polymerases transcribe different groups of genes; with Pol II being responsible for all protein-coding and many non-coding genes. RNA synthesis by msRNAPs is achieved via the nucleotide addition cycle, in which each iteration incorporates an incoming ribonucleoside triphosphate (NTP). Incoming NTPs are selected based upon sequence complementarity to the DNA template. The nucleotide addition mechanism for msRNAPs employs two Mg2+ ions for catalysis. The tightly bound Mg2+ is coordinated by a conserved triad of aspartic acid residues in a highly conserved NADFDGD motif within the largest subunit (β' for RNAP, RPB1 for Pol II) while a second Mg2+ associates with the incoming NTP (*2*, *4*–*6*). The cleft and channels formed between catalytic subunits (β'and β for RNAP, RPB1 and RPB2 for Pol II) includes (i) the duplex DNA-binding clamp where downstream DNA is bound, (ii) the primary channel that holds the RNA-DNA hybrid, and (iii) the RNA exit channel

(reviewed in (*6*)) (Figure 1A). The active site for RNA polymerization is located at the base of the cleft between the catalytic subunits (pincers of the crab-claw). Several mobile elements of the largest catalytic subunit, including the bridge helix and trigger loop, divide the primary channel and form a secondary channel through which diffusing NTPs can access the active site (Figure 1B).

The msRNAP complexes are catalytically competent for transcription elongation alone; however, sequence-specific recognition of promoter DNA requires the action of specificity factors. These transcription factors are responsible for binding DNA sequence elements (or motifs), which collectively makeup the core promoter, to properly orient and position msRNAPs for accurate transcription initiation. In addition to promoter recognition, specificity factors function to (i) separate the strands of duplex DNA to generate a local region of 'melted DNA' in the msRNAP cleft, (ii) load the template strand along the base of the cleft, and (iii) select the transcription start site (TSS) from which to begin RNA synthesis. In bacteria, promoter recognition by RNAP requires association with a sigma (σ) factor to form a holoenzyme (Eσ) before promoter binding (Eσ structure reviewed in (*7*, *8*)). Bacteria typically have one major σ factor (σ70 in E. coli) that drives expression of 'housekeeping' genes and varying numbers of alternative σ factors (6 in E. coli, but up to 60 in Streptomyces coelicolor) that direct expression of genes in response to specific environmental conditions or stress (*2*, *9*–*11*). The σ70 subunit consists of four domains (R1-4), connected by flexible linkers, that interact with separate DNA sequence elements in the promoter (*9*). For Eukaryotic organisms, a set of general transcription factors(GTFs)—Transcription Factor(TF)IID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH—form an assembly with Pol II on core promoter DNA, the pre-initiation complex (PIC) (*12*, *13*). While RNAP holoenzyme formation is obligatory prior to promoter binding, experiments have

supported two models for PIC formation—the sequential assembly model and the holoenzyme model of PIC formation (*12–15*). The sequential model of PIC assembly involves individual GTFs sequentially binding in an ordered manner, beginning with the initial binding of TFIID, TFIIB and TFIIA on the upstream region of promoter DNA to form the upstream promoter assembly (UPA), which is responsible for promoter recognition. The TFIIF-Pol II subcomplex is then recruited by the UPA before incorporation of TFIIE and TFIIH, which are required for promoter melting (*16*), completes PIC assembly. In contrast, the holoenzyme model describes association of GTFs with Pol II in a holoenzyme prior to promoter DNA binding (*14*).

Transcription can be divided into three major stages: initiation, elongation, and termination (Figure 2A). Each stage is a multistep process that offers multiple targets for regulation. Initiation is the most regulated step in transcription; for bacterial RNAP it begins with Eσ binding to promoter DNA to form the Eσ-closed promoter complex ($RP_C$), in which the DNA remains double stranded. The $RP_C$ undergoes a series of conformational changes (reviewed in (*9*)) resulting in the unwinding of ~13 bp DNA near the transcription start site (TSS) (*17*, *18*). This forms an Eσ-open promoter complex ($RP_O$) either spontaneously (σ70) or with assistance of ATP hydrolysis (σ54) (*19*). The unwound region of DNA (open bubble) in the $RP_O$ has the template strand placed at the wall of the primary channel with the DNA base corresponding to TSS (+1) positioned at the active site, and an initiating NTP (iNTP) and a second NTP can bind to the 'i' and 'i+1' sites of the enzyme, respectively. Phosphodiester bond formation between the initial two NTPs leads to a transition from $RP_O$ to the initial transcribing complex ($RP_{ITC}$) that extends the RNA in the 5' to 3' direction. After reaching a length of 4-5 bases (*17*, *18*, *20*, *21*), the RNA 5'-end begins to sterically and electrostatically clash with the σ factor at the "σ finger" or σR3.2 domain (for σ70). This σR3.2 domain blocks access to the

RNA exit channel, resulting in either the ejection of the RNA through the secondary channel (abortive initiation) or structural re-organization of the σ factor (regions σR3.2 and σ70-R4) to open the RNA exit channel (*22*). Structural re-organization marks the entry of $RP_{ITC}$ into elongation through the process of promoter escape. The highly processive ternary RNAP-DNA-RNA elongation complex ($RP_E$) eventually terminates transcription and releases the RNA product either after reaching an intrinsic termination signal encoded in the DNA template, or by the action of termination factors such as Rho. In the case of Pol II, PIC assembly results in the closed complex (Pol II-CC) where duplex DNA is held above the cleft (*23*, *24*). For linear DNA templates *in vitro* and many promoters *in vivo* (*25*, *26*), melting the DNA to form the open complex (Pol II-OC) requires TFIIH translocase activity on DNA downstream of the TSS to increase the torsional strain on the DNA situated above the Pol II cleft (*27*). This ATP (or dATP)-dependent translocation of ~ 1 turn of the DNA helix induces a similar sized unwound region of DNA in the Pol II cleft—forming the Pol II-OC (*27*). Negative supercoiling obviates the need for TFIIH activity on certain promoters, with a 'minimal or basal' system of factors—TATA-box binding protein (TBP, a subunit of TFIID), TFIIB, and TFIIF sufficing in driving accurate transcription by Pol II (*28–30*). The evolutionarily conserved microenvironment of the Pol II active center binds and polymerizes NTPs through a mechanism that is nearly identical to that of RNAP (*1*, *6*). During initial transcription, the growing RNA chain will eventually clash with a region of TFIIB (B-reader region), which is analogous to the bacterial RNAP σR3.2 domain (*31*). This clash (RNA ≥ 6 nucleotides [31]) destabilizes the TFIIB-Pol II interaction, which weakens until TFIIB release when the RNA reaches 12-13 nucleotides in length (*32*). Disruption of numerous contacts within the PIC, collapse of the upstream bubble region and TFIIB release result in Pol II clearing the promoter to form an elongation complex (*33*), while

leaving behind a scaffolding complex responsible for binding the next Pol II for re-initiation (*34*, *35*).

Much of what we understand about transcription derives from ensemble-averaged experiments; however, single-molecule (sm) experiments performed over the past decade have also contributed significantly. The application of sm techniques was driven by their utility in providing mechanistic information that is typically concealed in the averaged data from ensemble measurements. In particular, information about rare sub-species, transient interactions between biomolecules, distinct subpopulations, or states for static or dynamic heterogeneous populations has been revealed through sm approaches (*36–39*). A variety of sm methods have been applied in the study of bacterial RNAP (*40–51*) and Pol II(*15*, *27*, *52–63*).

In fluorescence microscopy and spectroscopy, FRET (Förster Resonance Energy Transfer) is a valuable tool for distance-dependent measurements. FRET is the non-radiative transfer of energy from a donor fluorophore excited by a higher energy photon to an adjacent acceptor fluorophore. FRET is calculated by the ratio of acceptor photons to the sum of donor and acceptor photons. FRET occurs if the following conditions are met: i) the respective orientation of the donor and acceptor dipoles is not orthogonal, ii) the donor and acceptor fluorophores are within a 2-10 nm range for resonance energy transfer, and iii) there is an overlap between the donor emission and the acceptor excitation spectra (*64*). A powerful advancement for smFRET (single-molecule FRET) was the development of alternating laser excitation (ALEX), in which the excitation source is continuously alternated between donor and acceptor excitation wavelengths by acoustic optical modulation (*64–66*). Application of smFRET-ALEX enables separation of subpopulations based on the stoichiometry of donor/acceptor fluorophore labeling.

Here, we review and highlight contributions mainly from diffusion-based smFRET-ALEX and some other sm techniques that have further elucidated key steps in RNAP transcription initiation and contribution of single molecule fluorescence microscopy to human Pol II transcription initiation. These include transcription bubble opening and dynamics related to TSS selection, the mechanism of DNA scrunching, changes in RNAP clamp conformation during the transcription cycle, the characterization of a paused-backtracked state in initiation, and retention of σ70 in early elongation.

## Open Complex Formation and TSS Selection

### Promoter Search

Transcription of a particular set of E. coli genes requires the proper σ factor to position RNAP at specific promoter sequences. Determining the mechanism by which Eσ scans the genome for target promoters is a challenge. Although the precise mechanism remains obscure, sm techniques have contributed to the elucidation of molecular mechanisms that govern this process. Facilitated diffusion or random 3-dimensional (3D) diffusion mechanisms have been proposed (*67*, *68*). Potential facilitated diffusion mechanisms include: (i) a series of local Eσ-DNA binding and unbinding events, called 1-dimensional (1D) 'hopping, (ii) the directional movement of Eσ along DNA without dissociation, called 1D sliding, and (iii) the movement of Eσ from one DNA site to another that is juxtaposed through DNA looping, called intersegmental transfer (discussed in (*40*, *48*)). Facilitated diffusion of RNAP has been demonstrated using a variety of sm techniques including AFM and sm fluorescence microscopy (*48*, *69–71*). However, the random 3D diffusion of Eσ is assumed to be the predominant mechanism (*72*, *73*).

The strength of the Eσ-promoter interaction is primarily determined by the sequence properties of the promoter. Different promoter elements interact with different regions of σ or core RNAP. The six nucleotide-long (hexameric) -10 and -35 elements with "TATAAT" and "TTGACA" consensus sequences (*74*), respectively, and the length of the spacer between them, are the major determinants of promoter strength (*74*) (Figure 2B). Additionally, the UP element located upstream of the -35 element to which the α subunits of RNAP bind, is required for efficient transcription from some promoters (*74*, *75*). A summary of contacts between Eσ70 and different promoter elements is shown in Figure 2B and are extensively reviewed elsewhere (*74*, *76*).

The primary roles of Pol II core promoters, spanning a region ~50 bp upstream to ~50 bp downstream with respect to the TSS, are to serve as a platform for PIC assembly, to dictate the position of the TSS and the level of transcripts produced (*77*). In the genomic context, the basal transcriptional activity of core promoters is regulated by distal regulatory DNA elements called enhancers (*77*). Sequence analysis of native eukaryotic promoters has led to the identification of several DNA motifs within the core promoter that are bound by the PIC, including the highly conserved TATA box, the upstream and downstream TFIIB recognition elements (uBRE and dBRE respectively), the initiator (Inr) element that encompasses the TSS, the motif ten element (MTE), the downstream promoter element and some additional elements reviewed in (*78*). TFIID, a complex of TBP and ~ 13 TBP-associated factors (TAF1-13), engages in the majority of contacts with promoter DNA, with TBP binding to TATA box while several TAF subunits bind the Inr, MTE and DPE (*78*). An smFRET study by Blair et. al. (*62*) on the interaction between TBP and DNA, labeled with a FRET dye pair, containing either a consensus or mutant (lower affinity) TATA box, challenged some conclusions from previous ensemble-based studies

(reviewed in (*79*)) and gave new insight into this process. They utilized total internal reflection fluorescence (TIRF) microscopy (details on TIRF reviewed in (*80*)) to track changes in FRET over time for individual DNAs, immobilized on a glass surface, in the presence or absence of TBP alone or TBP and TFIIA (which is known to stabilize the interaction of TBP with DNA). Contrary to earlier studies, they demonstrated that the extent of DNA bending following TBP binding is similar for both consensus and mutant TATA box. In addition, as they observed individual DNA templates undergo multiple bent-unbent transitions in the presence of TBP, they identified two conformationally similar (same mean FRET) bent-DNA populations with distinct kinetic properties in returning to the unbent state that they suggest could be of functional significance (*62*). Prior to this work, similar smFRET experiments by Schluesche et al. (*57*) shed light on the mechanism through which transcription initiation is regulated by negative cofactor-2 (NC2), a complex known to bind TBP-DNA complexes and either positively or negatively impact gene expression for many genes (reviewed in (*81*)). TIRF-based smFRET-ALEX experiments, with donor dye-labeled TBP (yeast) and DNA labeled with an acceptor dye at a position upstream relative to the TATA box, were done to further probe the peculiar effect of NC2 on TBP-DNA complexes observed through ensemble-based assays [85]. DNA footprinting results showed that the TATA box region protected by TBP against DNase I digestion is reduced in the presence of NC2, despite the interaction of TBP with DNA to over an order of magnitude greater than TBP alone (electrophoretic mobility shift assays or EMSA) (*57*). The two mechanisms suggested to explain this behavior were that either the ring-structure formed by NC2 and TBP opens to provide access to DNase I or that the NC2 mobilizes TBP along DNA (*57*). Addition of NC2 to TBP-DNA resulted in the FRET signal changing from a steady pattern with well-defined states to a very dynamic pattern alternating rapidly between different states, which

is consistent with DNA footprinting results. Subsequent ensemble-based assays validated the hypothesis that NC2-TBP complexes move along the DNA (*57*). Their smFRET data also validated the EMSA results showing long residence times of NC2-TBP complexes on DNA as the NC2-induced dynamic FRET signal was observed continuously for up to four hours (*57*). A previous TIRF-based study by Zhang et. al. (*82*) demonstrated that while TFIID and TFIIA form stable associations with promoter DNA, TFIIB forms transient weak interactions (~1.5 second residence time) that are only strengthened after recruitment of the Pol II-TFIIF complex.

Another sm TIRF-based study by Horn et al. (*15*), investigated the effect of a minimal Pol II system (TBP, TFIIB, TFIIF, Pol II) or different subsets of assembled PICs, with one or more GTFs missing, on interaction with heteroduplex DNA labeled with two fluorophores (mismatch from -9 to +3 region with respect to TSS) and on transcriptional activity. To assess the effect of individual and different subsets of PIC on promoter binding, they immobilized Pol II and then either flowed in doubly-labeled DNA or flowed in different PIC factors before flowing in DNA. They observed very efficient binding by Pol II alone to heteroduplex DNA, but only slight effects with PIC subsets. Only exclusion of TFIIB significantly reduced promoter binding. This contrasted with their results demonstrating a strong dependence of transcriptional activity on GTF presence. They monitored transcriptional activity by co-localizing green spots (Cy3, labeled at upstream region of template), observed before and after NTP addition, with red spots (Cy5, labeled at downstream region of template) that emerge after NTP addition. The emergence of the Cy5 fluorescence is due to the removal of an adjacent quencher on a nicked fragment of DNA that is removed by a transcribing Pol II. Transcriptional activity was especially dependent on TBP and TFIIB.

Characterizing the RNAP clamp domain conformation during $RP_O$ formation using smFRET-ALEX

Ensemble and sm experiments have uncovered many kinetic and structural details of the $RP_C$ transition to $RP_O$. The mechanism for transcription initiation at all σ70 promoters is proposed to entail a three-step promoter opening process. Initially, the isomerization of $RP_C$ to $RP_O$ was thought to be a single step process (*83*); however, recent biophysical and biochemical data points to the existence of relatively unstable intermediate complexes (reviewed in (*9*)). During the isomerization of RPc into the first intermediate complex ($RP_{i1}$), the DNA upstream of the -35 element wraps around Eσ, while the DNA downstream of the -10 element bends towards and into the RNAP cleft (*9*). DNA unwinding in the RNAP cleft then leads to formation of a 13-base DNA bubble, converting $RP_{i1}$ into a second intermediate complex, $RP_{i2}$. Finally, additional conformational changes in the RNAP clamp that binds downstream DNA forms stable $RP_O$ from $RP_i$ (*46*, *84*). Crystal structures of RNAP-DNA complexes revealed critical conformational changes in RNAP that occur during transcription initiation (*85*, *86*). Among these, the most prominent structural transformations involved the RNAP clamp, a large mobile domain within the β' subunit that is connected to the "β' switch-2" hinge region at the base of the β' pincer (*9*).

Conformational changes in the β' switch-2 region enable the β' clamp to swing relative to the β pincer, which leads to either opening or closing of the pincers. Depending on crystallization conditions, the β' clamp adopted different conformations, varying from an open to a closed state due to clamp swinging by at least 20° (*86*). To confirm that the clamp conformations were not artifacts of crystallization, and to probe their functional relevance, the clamp conformations at different steps of transcription were examined using diffusion-based smFRET-ALEX (*86*). Changes in spatial organization of the clamp domain were monitored by

measuring shifts in FRET efficiency—a function of the distance between two FRET dyes positioned at the tips of the RNAP pincers. For example, swinging of the β' clamp away from the β pincer would result in a relative decrease in FRET efficiency, as the distance between the two pincers would increase. The smFRET-ALEX measurements by Chakraborty et al. (*86*) identified three different clamp conformations, each with a characteristic mean FRET, based on Gaussian fitting of FRET efficiency histograms. The three clamp conformations corresponded to an (i) open, (ii) closed (inward rotation by ~14°), and (iii) collapsed clamp, with the collapsed clamp reflecting a conformation more closed than any crystal structure had shown (inward rotation by ~22°). Given the short lifetime of Eσ70 $RP_C$ and its rapid isomerization to $RP_O$ at room temperature, the ATP hydrolysis-dependent Eσ54 mechanism for bubble opening was used to monitor changes in clamp conformation for $RP_C$, $RP_{i1}$, and $RP_{i2}$. After initial promoter binding, E-σ54 remains in $RP_C$ until an AAA+ ATPase (e.g. NtrCl) mediates the transition to $RP_O$. The $RP_{i1}$ and $RP_{i2}$ intermediates were trapped using ATP analogs that mimic the ground (ADP-BeFx) and transition state (ADP-AlFx) intermediates during ATP hydrolysis. The RNAP clamp exhibited an open conformation for all initial steps until formation of $RP_O$, in which a closed clamp conformation was observed. However, the flexibility in clamp conformations during the early steps of transcription initiation was implied by the minor subpopulations in the FRET histogram corresponding to the closed clamp (*86*). An open clamp conformation for Eσ up until $RP_{i2}$ is consistent with the hypothesis that opening of the RNAP pincers permits DNA loading into the RNAP cleft and for subsequent conformational changes in DNA until formation of $RP_O$. The closed clamp conformation in $RP_O$ suggests an interaction between the positively charged inner surface of the RNAP cleft and the negatively charged backbone of the single stranded DNA (ssDNA) in the bubble.

A smFRET-ALEX Assay for Measuring Open Bubble Formation and Dynamics

$RP_O$ formation is a common rate-limiting step in transcription initiation. $RP_O$ stability is controlled by various factors, including the global transcription regulators guanosine (penta- or tetra-) phosphate ((p)ppGpp) and DksA (*87*, *88*). To measure promoter opening kinetics, several ensemble and sm approaches have been used, including chemical perturbation footprinting assays (discussed in (*89*, *90*)), DNA topology assays (*89*), magnetic tweezers (*91*), and sm fluorescence techniques (discussed in (*51*)). To sense conformational change in the transcription bubble during $RP_O$ formation, a modified diffusion-based smFRET-ALEX assay (*90*) with a FRET dye pair positioned in the transcription bubble was used to conduct end-point and real-time monitoring experiments. In this FRET assay, the donor/acceptor dye pair was placed on opposite DNA strands within a region that would form an open bubble in initiation (*90*). Importantly, the dyes were in such close proximity (2 bp apart) that one dye quenched the fluorescence of the other, whereas quenching was abolished upon bubble opening (*90*). We modified this assay by generating a library of DNA probes carrying donor/acceptor dyes at various positions within the transcription bubble and monitored the FRET changes in different regions of the bubble. This library of FRET dye DNA probes can be a powerful tool for elucidating the conformational dynamics of transcription bubble opening during $E\sigma70$ $RP_O$ formation.

Several analysis methods have expanded upon the information obtained from FRET histograms to identify the dynamic behavior of single molecules in the millisecond and sub-millisecond time-scales. Burst variance analysis (BVA) is one such method (*92*). Due to photon counting statistics (shot-noise), a smFRET measurement of any molecule with a fixed distance

between donor and acceptor results in a distribution of FRET efficiencies instead of a single value. The width of the distribution increases as the number of photons decreases. The expected width (standard deviation, σE) of a distribution only broadened by shot-noise can be computed from the binomial law. A distribution broader than the shot-noise limited width indicates the superimposition of multiple heterogeneities in a FRET distribution. This heterogeneity could either be due to the presence of multiple species with distinct FRET efficiencies (static heterogeneity) or a single species dynamically interconverting between different states (dynamic heterogeneity). BVA determines if these heterogeneities are static or dynamic. BVA achieves this by equally dividing each burst into smaller sub-bursts for which a FRET efficiency is computed. The standard deviation of sub-bursts from the FRET efficiency of the burst is then computed (si). Shot-noise due to the lower number of photons per sub-burst predicts a distribution of si around σE. For static heterogeneity, the distribution of si is consistent with σE. However, if the broadening is caused by dynamic heterogeneity, then the distribution of si is deviated from σE. Application of BVA by Robb et al.(*93*) demonstrated that FRET efficiency for the $RP_O$ transcription bubble fluctuated more than the expected shot noise limit. This indicated that the millisecond time-scale dynamics of the transcription bubble could be associated with dynamics that are important for TSS selection. Both eukaryotic and bacterial RNAPs can initiate transcription from different TSSs on the same promoter. Promoter sequence characteristics play an important role in TSS selection, which in turn could affect abortive RNA synthesis, RNA stability, and translation efficiency. For bacterial RNAPs, transcription is primarily initiated by purine NTPs at sites located 4-12 bp downstream of the -10 element. The process of alternative TSS selection is thought to proceed through a scrunching and anti-scrunching mechanism (*93*–*95*). For start sites located downstream of the canonical TSS,

scrunching of the transcription bubble unwinds additional downstream DNA, incorporating it into the transcription bubble while bringing it towards the active site. However, positioning of registers upstream to the canonical TSS in the active site requires rewinding of DNA at the downstream edge of the bubble (*93*). Recently, the correlation between TSS position and scrunching was demonstrated using a high-throughput approach termed MASTER (massively systematic transcript end readout), which relied on next generation sequencing to report on the TSS and transcript yields generated from a library of DNA templates (*96*). The increase of downstream TSSs for negatively supercoiled DNA templates, which promote DNA unwinding and bubble expansion, suggested a scrunching mechanism for downstream TSS selection(*96*). Further evidence implicating DNA scrunching dynamics for TSS selection came from combining leading edge and trailing edge RNAP crosslinking to the DNA template with MASTER (*95*).

**Transcription initiation proceeds through a scrunching mechanism and can generate a paused, backtracked state**

The mechanism of initial transcription and detection of abortive initiation

RNAP can form very stable $RP_O$ complexes because of the high binding affinity between Eσ and promoter elements. However, strong Eσ-promoter interactions may impede promoter escape. A characteristic of such promoters (LacUV5, λPR, T5N25) is the increased production of short abortive RNA products (2 - 10 nucleotides for lacUV5 promoter (*97*)). Initial transcription by $RP_{ITC}$ ultimately results in promoter escape or abortive initiation, in which the abortive products are released through the secondary channel. Determinants of abortive initiation include i) the stability of the RNA-DNA hybrid (dictated by the sequence and the

length of the nascent transcript), ii) repulsive interactions between the nascent transcript 5' end and the σR3.2 finger that blocks the RNA exit channel (*17*, *21*, *98*), and iii) the interactions between Eσ and promoter DNA (*99–101*). Abortive initiation returns the complex to $RP_O$, which can continue to undergo multiple initial transcription/abortive initiation cycles until promoter escape occurs.

RNAP abortive initiation has been observed by equipping a total internal reflection (TIRF) optical microscopy setup with smFRET-ALEX and imaging surface-immobilized Eσ70 complexes labeled with a FRET dye pair (*102*). FRET changes between the leading edge of Eσ70 and downstream DNA were evaluated for $RP_O$ and different $RP_{ITC}$ states; these states were achieved through NTP starvation (i.e. by excluding one or more NTPs from the reaction) to reach a specific $RP_{ITC}$ state (e.g. ITC 2, 4, or 7). The mean FRET efficiency increased gradually in going from $RP_O$ (or earlier ITC states) to later $RP_{ITC}$ states, which suggested that longer nascent RNAs increasingly reduce the distance between the downstream DNA and the leading edge of RNAP (*102*).

Insight into the mechanism of initial transcription by $RP_{ITC}$ was offered in early DNA footprinting studies (*97, 103–105*), in which three possible models were proposed: [1] transient excursions, [2] RNAP inchworming, and [3] scrunching. The transient excursion model proposed that RNAP translocates downstream during initial transcription before returning back to the TSS following abortive initiation. The inchworming mechanism hypothesized structural flexibility in RNAP that would permit the leading edge of RNAP to translocate downstream and return, while the trailing edge of RNAP remained stationary. The scrunching mechanism hypothesized that a stationary RNAP reeled downstream DNA into the RNAP cleft during initiation, resulting in a larger 'scrunched' bubble that recoils and returns to the original bubble

size during abortive transcript release (*106*). These mechanisms were studied by smFRET-ALEX experiments in solution, which supported a scrunching mechanism (*106*).

The smFRET-ALEX experiments tested different configurations of FRET pairs on RNAP and DNA: [1] the leading edge of RNAP and downstream DNA, [2] the trailing edge (σ70-R4) of the Eσ70 and upstream DNA, [3] the trailing and leading edges of Eσ70 relative to the spacer element (position -20 relative to TSS), and [4] upstream and downstream DNA. The transient excursion model was refuted because during initial transcription, no changes in FRET were observed in the upstream DNA-RNAP trailing edge configuration for $RP_{ITC\leq7}$ compared to $RP_O$; by contrast, a shift toward lower FRET was expected according to the model. The inchworming model was ruled out based on the absence of FRET changes between the spacer region and the leading edge of the Eσ70 (for $RP_{ITC\leq7}$ compared to $RP_O$). The inchworming model, however, predicted decreased FRET between the leading edge of RNAP and the promoter spacer region. The observed FRET changes in the downstream DNA-leading edge and the upstream-downstream DNA configurations supported the scrunching model. The scrunching mechanism has also been demonstrated by Revyakin et al. using a different sm method, DNA nanomanipulation with magnetic tweezers (*91*). Subsequent studies showed viral and eukaryotic RNA polymerases also perform initial transcription through DNA scrunching (*107*, *108*). Thus, DNA scrunching may be a universal mechanism for initial transcription throughout all domains of life.

Despite significant progress in our understanding of transcription initiation mechanisms, many details remain obscure. One major open question is how extra ssDNA in the scrunched states are accommodated by RNAP. The structural data on transcription initiation complexes suggest that the internal space within the RNAP cleft is the main limiting factor (*17*) for

accommodation of ssDNA. However, analysis of the size distribution of abortive products from strong promoters shows the presence of abortive products of significant length (up to 18-25 nt) (*93*, *95*, *96*, *98*, *103*). A recent RNAP-DNA crosslinking and footprinting study demonstrated that the initiation complex may accommodate the additional scrunched bases of the nontemplate strand by extruding them out of the cleft and into the solvent (*109*). Moreover, Ploetz and Lerner et al. reached the same conclusion using sm experiments that simultaneously involved FRET and protein-induced fluorescence enhancement (PIFE) measurements (*51*). These experiments utilized the Cy3 dye, which increases its fluorescence in response to a rise in local viscosity or steric restriction. Internally labeled bases on the nontemplate strand at position +1 or +3 (relative to the TSS) acted as FRET donors to an acceptor positioned within the spacer element of the template strand. According to Winkelman et al.(*109*), the nontemplate base at position +1 should be exposed to solvent during scrunching, hence the steric restriction should be reduced. Increased FRET for the +1-Cy3 reported on open bubble formation and the intensity of Cy3 fluorescence also increased (i.e. the PIFE effect) upon $RP_O$ formation, indicating that the +1-Cy3 was in close proximity to RNAP. Upon further extension of the nascent RNA, the PIFE effect increased for +3-Cy3 (i.e. became part of the bubble owing to scrunching); however, the PIFE effect decreased for +1-Cy3 due its extrusion into solution.

Initial transcription by $RP_{ITC}$ can produce a paused and backtracked intermediate

Extension of the RNA chain by $RP_{ITC}$ involves a series of concerted motions in the holoenzyme and DNA (*6*). Nucleotide addition occurs in the active site of RNAP (see Figure 1), which is comprised of several β' and β subunit domains, including: the α-helical bridge helix (BH) that bifurcates the cleft into the primary and secondary channels, the mobile trigger loop

(TL) that controls access to the active site from the secondary channel, the F-Loop, and the catalytic loop (β'), which coordinates a Mg2+ ion (*6*, *76*) (see Figure 1B). The BH and TL are the primary actors during nucleotide addition and play a central role in RNAP translocation. The nucleotide addition cycle can be generally described as a sequence of four repeating events: (1) RNAP translocation along DNA, which moves the newly incorporated NTP from the insertion site (i+1) to the 'i' site, (2) incoming NTP binding in the 'i+1' site, (3) phosphodiester bond formation between RNA 3'-end in 'i' site and the NTP in 'i+1' site, and (4) pyrophosphate release (*6*, *110*). Completion of a nucleotide addition cycle results in an RNAP conformation termed the pre-translocated state, in which the 3'-end of the newly incorporated base occupies the insertion site. RNAP translocation relative to DNA shifts the RNA 3'-end from the 'i+1' site to the upstream 'i' site to form the post-translocated state (forward translocation). Alternative translocation states, defined by the relative position of the RNA 3'-end to the active site, are believed to alternate through rapid thermal (Brownian) fluctuations that transpire on the microsecond timescale. The driving force for forward translocation is suggested to be the stabilization energy provided by correct nucleotide binding (*6*, *110*). The backward translocation of the pre-translocated state results in formation of a 'backtracked' state in which the RNA 3'-end enters the secondary channel.

For certain promoters (e.g. lacUV5 (*111*), phage T5 N25 (*112*), malT (*113*, *114*)), the rate-limiting step in transcription initiation is promoter escape. Despite efficient RNAP binding and formation of stable open complexes, these promoters exhibit an increase in the amount and size distribution of abortive products, with a reciprocal decrease in promoter escape efficiency. Previous studies suggested the existence of a paused, backtracked state in initial transcription (*22*, *103*, *115–125*).

Lately, we confirmed the existence of a paused-backtracked intermediate in RNAP transcription initiation. A paused-backtracked intermediate was characterized using a single-round quenched kinetics assay and DNA nanomanipulation with magnetic tweezers (*126*). The single-round quenched kinetics assay utilized the diffusion-based smFRET-ALEX technique to detect transcripts over time, until terminated (quenched) at different time points. Because smFRET-ALEX is a sm technique, single-round quenched kinetics assays allowed accurate quantification of transcripts generated by RNAP-DNA complexes. To confirm the existence of the paused state in RNAP initiation, the promoter escape kinetics, starting from different $RP_{ITC}$ states, were monitored by single-round quenched kinetics assays. The assays revealed a delay (~3.5 times slower) when starting from $RP_{ITC\leq7}$ compared to $RP_{ITC=2}$. This delay was nearly abolished by inclusion of GreA, indicating a paused-backtracked intermediate during initiation. This finding was corroborated with DNA nanomanipulation assays with magnetic tweezers, which can detect changes in DNA topology (e.g. changes in the size of transcription bubble) in real time. Importantly, DNA nanomanipulation assays demonstrated RNAP transcription complexes could be paused-backtracked during initial transcription in the presence of all four NTPs at biologically relevant concentrations. A modified transcription initiation model was suggested that included the paused-backtracked Eσ. This paused-backtracked intermediate in RNAP initiation was also reported by Duchi et al. using ALEX-TIRF techniques (*127*). In their report, Duchi et al. showed that $RP_{ITC}$ could form a stable paused-backtracked intermediate and σR3.2 played a key role in controlling the properties of this intermediate.

Promoter escape; σ70 retention into elongation; Pol II activity assays

For the $RP_{ITC}$ complex, the nascent RNA eventually reaches a required length (~9-15 bases) for promoter escape. The process of promoter escape requires an extensive set of conformational changes to allow Eσ to disengage from the promoter and form the processive $RP_E$ complex. The most notable structural transitions underlying promoter escape include the following: [1] Displacement of σR3.2, which otherwise impedes the nascent RNA 5'-end; removal of σR3.2 together with σR4 allows nascent RNA to enter the RNA exit channel. [2] Weakening of RNAP interactions with σ, mainly through destabilization of the σR4-β interaction (*7, 128*). [3] Breaking σ70 interactions with the promoter (σR2.3-R2.4 with the -10 element, σR4.2 with -35 element). [4] And finally, reannealing of the upstream edge of the transcription bubble to form a 12-base pair transcription bubble, which is characteristic of elongating RNAP.

It was previously hypothesized that σ release from RNAP during promoter escape was obligatory, and a 'σ cycle' was proposed in which σ associates with RNAP for initiation and is released when RNAP enters elongation (cited in (*129*)). However, the harsh separation techniques (e.g. gel electrophoresis and chromatography) used to test for σ presence in elongation could also promote σ dissociation. Several studies conducted in the last ten years have provided sufficient evidence for σ retention during transcription elongation. Kapanidis et al. demonstrated σ retention in $RP_E$ up to position +50 from the TSS using diffusion-based smFRET-ALEX (*129*). Retention of σ70 in early elongation complexes was also observed during investigation of the mechanism of σ70-dependent promoter-proximal pausing (*130, 131*).

In a study by Revyakin et al. (*52*), the transcriptional activity of human Pol II from immobilized and fluorescently labeled DNA was monitored by co-localization of fluorescent probes that are complementary to transcripts with the labeled template. To ensure Pol II stalling after entering into elongation, the super core promoter was oriented to direct Pol II towards the

immobilized end of the DNA, which would enable co-localization of labeled probe bound to nascent transcript with template DNA. Using this approach they were able to investigate the effects of TFIID and mutations in different promoter elements on transcriptional activity. No transcriptional activity was observed in the absence of TFIID and transcription was attenuated in the promoter with a mutated Inr relative to the normal super core promoter.

**Concluding Remarks and Future Perspective**

RNAP has been and continues to be among the most extensively studied biomolecular machines, which is a testament to its complexity and its fundamental importance in biology. Over the past decade, sm assays have resolved heterogeneities in molecular distributions and have revealed fundamental mechanisms through which RNAP and Pol II functions. In this review, we highlighted some of the contributions of sm fluorescence, mainly diffusion-based smFRET, and other sm approaches that have uncovered RNAP and Pol II transcription mechanisms.

The future questions and challenges facing the sm transcription field are numerous; many questions surround the rapid transformations that occur from the promoter search mechanism to the formation of $RP_O$. Interestingly, the process of TSS selection, which was linked to bubble dynamics in the millisecond timescale (*93*), was postulated to also occur through a scrunching mechanism. Furthermore, it was proposed that DNA scrunching and anti-scrunching may underlie alternate TSS selection in the absence of NTPs (*94*, *95*, *132*). This implies that the millisecond bubble dynamics are primarily caused by DNA scrunching dynamics in $RP_O$. If correct, additional BVA experiments will be important to define different conformational modes and their potential contribution to the bubble dynamics.

The intermediate-resolution structure of a paused, backtracked bacterial RNAP elongation complex has been reported (*133*), but the structure of a paused-backtracked initiation complex (*126*, *127*) remains unavailable. It is assumed that the "ratcheted" form of RNAP is a structural feature common to all paused and backtracked RNAPs. A 'ratcheted' form of RNAP is distinct from the processive $RP_E$ in the conformations of several prominent structural elements of RNAP (*133*) (e.g. a kinked/bent BH, a partially open TL, an open clamp, and expansion of DNA:RNA hybrid binding site). Whereas heterogeneity among a population of paused, backtracked RNAP complexes will complicate any structural analysis, recent advances in cryoEM (*134*, *135*) may allow detailed structural information to be uncovered.

Despite the increasing capability of other techniques--such as single particle cryo-EM, immobilized smFRET imaging, and magnetic or optical tweezers--to resolve structural intermediates hidden in ensemble data, diffusion-based smFRET maintains several advantages. For example, diffusion-based smFRET is better at characterizing structural dynamics and the kinetics in going between different intermediates than cryo-EM. It is also higher a throughput method with greater temporal resolution (a few milliseconds) in comparison to immobilized FRET imaging (tens of milliseconds) and magnetic or optical tweezers (~ 1 second).

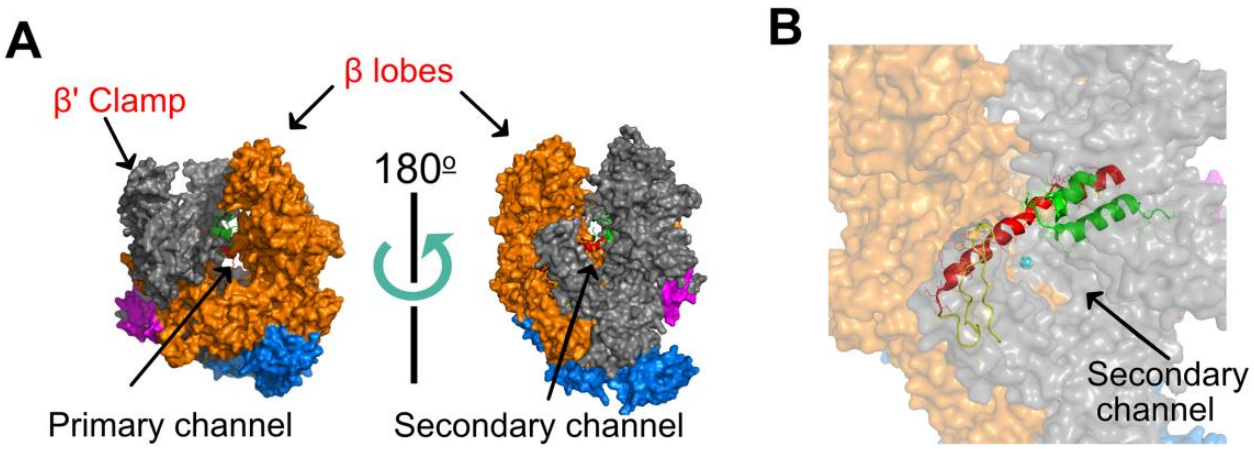



**Figure 1.** High-resolution structure of Thermus aquaticus core RNAP (PDB 1HQM). Figure generated using PyMol. (A) The five RNAP subunits are represented with different colors: The two α subunits are blue, β′ is grey, β is orange, and ω is magenta. Two orientations (related by ~180° rotation) are shown, with the LEFT showing the trailing edge of RNAP facing upstream DNA and the RIGHT showing the leading edge facing downstream DNA. The cleft between the β′ clamp and β lobes (the pincers) forms the primary channel, while the secondary channel is arranged on the opposite face of RNAP. (B) The catalytic magnesium (cyan sphere), bridge helix (red), trigger loop (green), and F-loop (yellow) of the active site can be seen from the downstream facing side of RNAP.

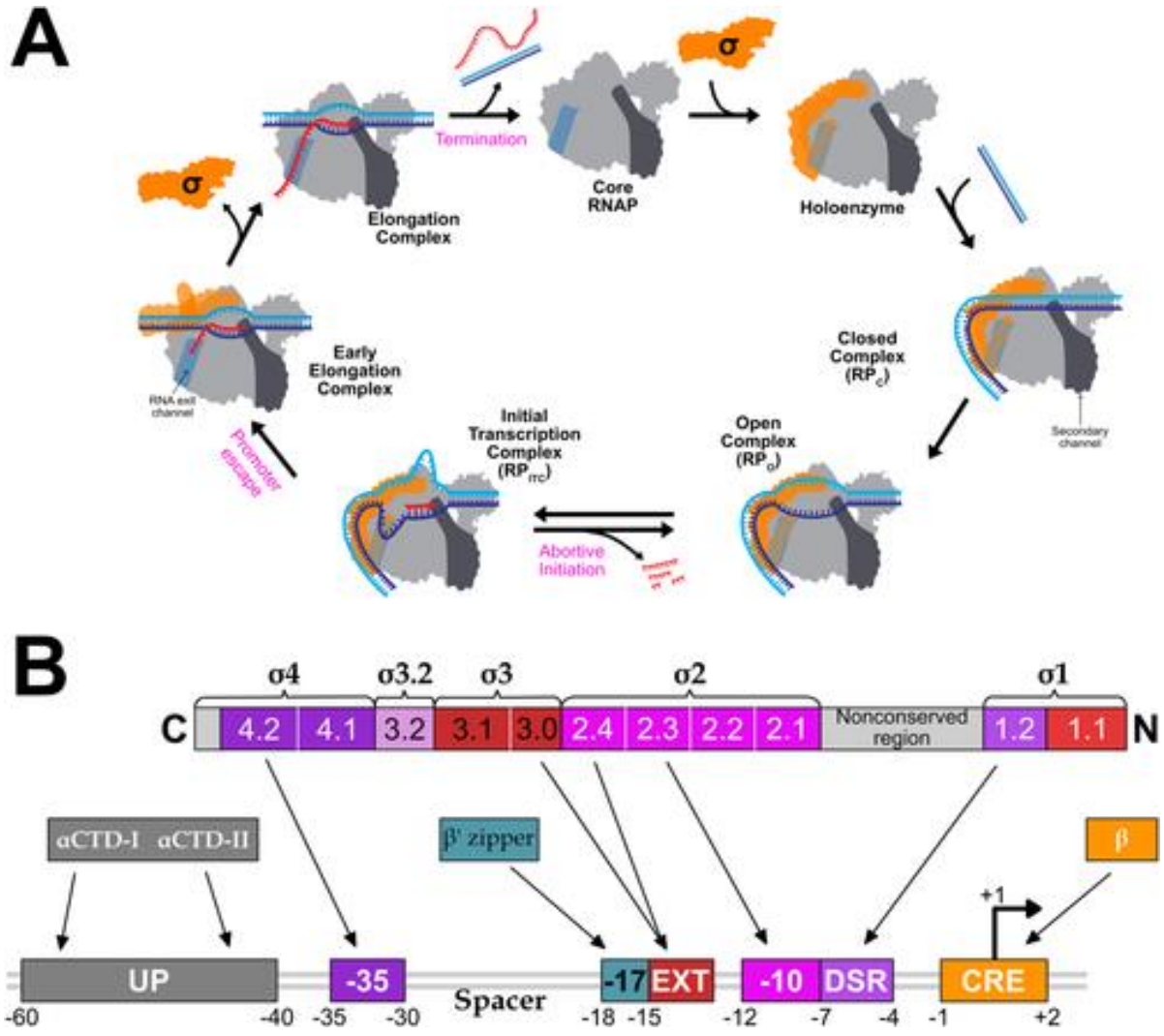



**Figure 2.** Schematic of transcription cycle and σ70 interactions with promoter DNA. (A) Association of σ (green) with core RNAP forms the RNAP holoenzyme (grey). The secondary channel (darker grey highlight) and the RNA exit channel (medium grey highlight) are represented on the right side and left side of the RNAP cartoon, respectively. σR3.2 and σR4 is shown protruding into the primary channel and occupying the RNA exit channel, respectively. HoloRNAP associates with promoter DNA (template strand in blue, and nontemplate strand in cyan) to form $RP_C$, which isomerizes to $RP_O$. Initial transcription by $RP_{ITC}$ produces short abortive RNA products. Nascent RNA can displace σR4 from the RNA exit channel by clashing with σR3.2 and can enter the RNA exit channel to form the elongation complex. Eventually,

$RP_E$ undergoes transcription termination resulting in RNAP dissociation from DNA. (B) Evolutionarily conserved structural domains and conserved regions of σ are shown as numbered and color-coded boxes. Promoter DNA is shown underneath, with arrows indicating interactions between promoter DNA sequence elements and regions of σ or RNAP.

## References


1. T. Fouqueau, F. Blombach, F. Werner, Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annu. Rev. Microbiol. Vol 71*. **71**, 331–348 (2017).
2. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*. **9**, 85–98 (2011).
3. F. Werner, Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* (2007), , doi:10.1111/j.1365-2958.2007.05876.x.
4. R. H. Ebright, RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304**, 687–98 (2000).
5. W. J. Lane, S. A. Darst, Molecular evolution of multi-subunit RNA Polymerases: Structural analysis. *J. Mol Biol.* **395**, 686 (2011).
6. E. Nudler, RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* **78**, 335–361 (2009).
7. K. S. Murakami, S. A. Darst, Bacterial RNA polymerases: The wholo story. *Curr. Opin. Struct. Biol.* (2003), , doi:10.1016/S0959-440X(02)00005-2.
8. N. Opalka *et al.*, Complete structural model of Escherichia coli RNA polymerase from a Hybrid Approach. *PLoS Biol.* (2010), doi:10.1371/journal.pbio.1000483.
9. E. Ruff, M. Record, I. Artsimovitch, Initial Events in Bacterial Transcription Initiation. *Biomolecules*. **5**, 1035–1062 (2015).
10. H. Maeda, Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* (2000), doi:10.1002/etc.2665.
11. A. Feklístov, B. D. Sharon, S. A. Darst, C. A. Gross, Bacterial Sigma Factors: A

Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* (2014), doi:10.1146/annurev-micro-092412-155737.

12. R. G. Roeder, The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* (1996), , doi:10.1016/0968-0004(96)10050-5.

13. M. C. Thomas, C.-M. Chiang, The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–78 (2006).

14. V. E. Myer, R. A. Young, RNA polymerase II holoenzymes and subcomplexes. *J. Biol. Chem.* (1998), , doi:10.1074/jbc.273.43.27757.

15. A. E. Horn, J. F. Kugel, J. A. Goodrich, Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity. *Nucleic Acids Res.* (2016), doi:10.1093/nar/gkw321.

16. U. Fiedler, H. T. Timmers, Analysis of the open region of RNA polymerase II transcription complexes in the early phase of elongation. *Nucleic Acids Res.* (2001).

17. Y. Zuo, T. A. Steitz, Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* (2015), doi:10.1016/j.molcel.2015.03.010.

18. B. Bae, A. Feklistov, A. Lass-Napiorkowska, R. Landick, S. A. Darst, Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* (2015), doi:10.7554/eLife.08504.

19. M. Buck *et al.*, The bacterial enhancer-dependent simga-54 transcription factor. *J. Bacteriol.* **54**, 4129–4136 (2000).

20. Y. Zhang *et al.*, Structural basis of transcription initiation. *Science (80-. ).* (2012), doi:10.1126/science.1227786.

21. R. S. Basu *et al.*, Structural basis of transcription initiation by bacterial RNA polymerase

holoenzyme. *J. Biol. Chem.* (2014), doi:10.1074/jbc.M114.584037.

22. L. M. Hsu, Promoter Escape by Escherichia coli RNA Polymerase. *EcoSal Plus*. **3**, 1–16 (2008).

23. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature* (2016), doi:10.1038/nature17970.

24. M. Hantsche, P. Cramer, Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.* (2017), , doi:10.1016/j.sbi.2017.03.013.

25. G. Orphanides, T. Lagrange, D. Reinberg, The general transcription factors of RNA polymerase II. *Genes Dev.* (1996), doi:10.1101/gad.10.21.2657.

26. E. Guzmán, J. T. Lis, Transcription factor TFIIH is required for promoter melting in vivo. *Mol. Cell. Biol.* (1999), doi:10.1128/MCB.19.8.5652.

27. J. Fishburn, E. Tomko, E. Galburt, S. Hahn, Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc. Natl. Acad. Sci.* (2015), doi:10.1073/pnas.1417709112.

28. J. A. Goodrich, R. Tjian, Transcription Factors IIE and IIH and ATP Hydrolysis Direct Promoter Clearance by RNA Poiymeke II. **77**, 145–156 (1994).

29. J. D. Parvin, P. A. Sharp, DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* (1993), doi:10.1016/0092-8674(93)90140-L.

30. J. D. Parvin, B. M. Shykind, R. E. Meyers, J. Kim, P. A. Sharp, Multiple sets of basal factors initiate transcription by RNA polymerase II. *J. Biol. Chem.* (1994).

31. S. Sainsbury, J. Niesser, P. Cramer, Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature*. **28**, 2–6 (2012).

32. P. Cabart, A. Ujvari, M. Pal, D. S. Luse, Transcription factor TFIIF is not required for

initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc. Natl. Acad. Sci.* (2011), doi:10.1073/pnas.1104591108.

33. D. S. Luse, Promoter clearance by RNA polymerase II. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829** (2013), pp. 63–68.

34. L. Zawel, K. P. Kumar, D. Reinberg, Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* (1995), doi:10.1101/gad.9.12.1479.

35. N. Yudkovsky, J. A. Ranish, S. Hahn, A transcription reinitiation intermediate that is stabilized by activator. *Nature* (2000), doi:10.1038/35041603.

36. S. Weiss, Fluorescence spectroscopy of single biomolecules. *Science (80-. ).* (1999), , doi:10.1126/science.283.5408.1676.

37. A. Dietrich, V. Buschmann, C. Muller, M. Sauer, Fluorescence resonance energy transfer (FRET) and competing processes in donor-acceptor substituted DNA strands: a comparative study of ensemble and single-molecule data. *Rev. Mol. Biotechnol.* **82**, 211–231 (2002).

38. L. Bai, T. J. Santangelo, M. D. Wang, SINGLE-MOLECULE ANALYSIS OF RNA POLYMERASE TRANSCRIPTION. *Annu. Rev. Biophys. Biomol. Struct.* (2006), doi:10.1146/annurev.biophys.35.010406.150153.

39. T. E. Tomov *et al.*, Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis. *Biophys. J.* (2012), doi:10.1016/j.bpj.2011.11.4025.

40. K. M. Herbert, W. J. Greenleaf, S. M. Block, Single-Molecule Studies of RNA Polymerase: Motoring Along. *Annu. Rev. Biochem.* (2008),

doi:10.1146/annurev.biochem.77.073106.100741.

41. Y. Santoso, L. C. Hwang, L. Le Reste, A. N. Kapanidis, Red light, green light: probing single molecules using alternating-laser excitation. *Biochem. Soc. Trans.* (2008), doi:10.1042/BST0360738.

42. K. Fujita, M. Iwaki, T. Yanagida, Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. *Nat. Commun.* **7**, 1–10 (2016).

43. G. Wang, J. Hauver, Z. Thomas, S. A. Darst, A. Pertsinidis, Single-Molecule Real-Time 3D Imaging of the Transcription Cycle by Modulation Interferometry. *Cell* (2016), doi:10.1016/j.cell.2016.11.032.

44. T. R. Strick, Optical investigations of the RNA polymerase molecular motor. *J. Biophotonics* (2008), , doi:10.1002/jbio.200810030.

45. D. J. Billingsley, W. A. Bonass, N. Crampton, J. Kirkham, N. H. Thomson, Single-molecule studies of DNA transcription using atomic force microscopy. *Phys. Biol.* (2012), , doi:10.1088/1478-3975/9/2/021001.

46. J. Michaelis, B. Treutlein, Single-molecule studies of RNA polymerases. *Chem. Rev.* (2013), , doi:10.1021/cr400207r.

47. B. E. Collins, L. F. Ye, D. Duzdevich, E. C. Greene, DNA curtains: Novel tools for imaging protein-nucleic acid interactions at the single-molecule level. *Methods Cell Biol.* (2014), doi:10.1016/B978-0-12-420138-5.00012-4.

48. M. Dangkulwanich, T. Ishibashi, L. Bintu, C. Bustamante, Molecular mechanisms of transcription through single-molecule experiments. *Chem. Rev.* (2014), , doi:10.1021/cr400730x.

49. L. J. Friedman, J. Gelles, Multi-wavelength single-molecule fluorescence analysis of

transcription mechanisms. *Methods* (2015), doi:10.1016/j.ymeth.2015.05.026.

50. Y. Suzuki, M. Endo, H. Sugiyama, Studying RNAP-promoter interactions using atomic force microscopy. *Methods* (2015), , doi:10.1016/j.ymeth.2015.05.018.

51. E. Ploetz *et al.*, Förster resonance energy transfer and protein-induced fluorescence enhancement as synergetic multi-scale molecular rulers. *Sci. Rep.* **6**, 1–18 (2016).

52. A. Revyakin *et al.*, Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev.* (2012), doi:10.1101/gad.194936.112.

53. M. Palangat *et al.*, Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II. *Transcription* (2012), doi:10.4161/trns.20269.

54. M. H. Larson *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci.* (2012), doi:10.1073/pnas.1200939109.

55. B. Zamft, L. Bintu, T. Ishibashi, C. Bustamante, Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc. Natl. Acad. Sci.* (2012), doi:10.1073/pnas.1205063109.

56. E. A. Galburt *et al.*, Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* (2007), doi:10.1038/nature05701.

57. P. Schluesche, G. Stelzer, E. Piaia, D. C. Lamb, M. Meisterernst, NC2 mobilizes TBP on core promoter TATA boxes. *Nat. Struct. Mol. Biol.* (2007), doi:10.1038/nsmb1328.

58. A. Muschielok *et al.*, A nano-positioning system for macromolecular structural analysis. *Nat. Methods* (2008), doi:10.1038/nmeth.1259.

59. J. Andrecka *et al.*, Single-molecule tracking of mRNA exiting from RNA polymerase II.

*Proc. Natl. Acad. Sci.* (2008), doi:10.1073/pnas.0703815105.

60. C.-Y. Chen, C.-C. Chang, C.-F. Yen, M. T-K Chiu, W.-H. Chang, Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl. Acad. Sci.* (2009), doi:10.1073/pnas.0811689106.

61. J. Andrecka *et al.*, Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase ii elongation complex. *Nucleic Acids Res.* (2009), doi:10.1093/nar/gkp601.

62. R. H. Blair, J. A. Goodrich, J. F. Kugel, Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics. *Biochemistry* (2012), doi:10.1021/bi300491j.

63. B. Treutlein *et al.*, Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol. Cell*. **46**, 136–46 (2012).

64. J. Hohlbein, T. D. Craggs, T. Cordes, Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* (2013), doi:10.1039/c3cs60233h.

65. A. N. Kapanidis *et al.*, Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38**, 523–533 (2005).

66. N. K. Lee *et al.*, Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* **88**, 2939–2953 (2005).

67. A. Feklistov, RNA polymerase: In search of promoters. *Ann. N. Y. Acad. Sci.* (2013), doi:10.1111/nyas.12197.

68. V. Svetlov, E. Nudler, Looking for a promoter in 3D. *Nat. Struct. Mol. Biol.* (2013), , doi:10.1038/nsmb.2498.

69. C. Bustamante, M. Guthold, X. Zhu, G. Yang, Facilitated target location on DNA by

individual Escherichia coli RNA polymerase molecules observed with the scanning force microscope operating in liquid. *J. Biol. Chem.* (1999), , doi:10.1074/jbc.274.24.16665.

70. M. Guthold *et al.*, Direct observation of one-dimensional diffusion and transcription by Escherichia coli RNA polymerase. *Biophys. J.* (1999), doi:10.1016/S0006-3495(99)77067-0.

71. Y. Harada *et al.*, Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys. J.* (1999), doi:10.1016/S0006-3495(99)77237-1.

72. L. J. Friedman, J. P. Mumm, J. Gelles, RNA polymerase approaches its promoter without long-range sliding along DNA. *Proc. Natl. Acad. Sci.* (2013), doi:10.1073/pnas.1300221110.

73. F. Wang *et al.*, The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nat. Struct. Mol. Biol.* (2013), doi:10.1038/nsmb.2472.

74. S. P. Haugen, W. Ross, R. L. Gourse, Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.* (2008), , doi:10.1038/nrmicro1912.

75. R. L. Gourse, W. Ross, T. Gaal, UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* (2000), , doi:10.1046/j.1365-2958.2000.01972.x.

76. J. Lee, S. Borukhov, Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action. *Front. Mol. Biosci.* (2016), doi:10.3389/fmolb.2016.00073.

77. V. Haberle, A. Stark, Eukaryotic core promoters and the functional basis of transcription

initiation. *Nat. Rev. Mol. Cell Biol.* (2018), , doi:10.1038/s41580-018-0028-8.

78. J. T. Kadonaga, Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* (2012), , doi:10.1002/wdev.21.

79. A. E. Horn, J. A. Goodrich, J. F. Kugel, Single molecule studies of RNA polymerase II transcription in vitro. *Transcription* (2014), , doi:10.4161/trns.27608.

80. S. L. Reck-Peterson, N. D. Derr, N. Stuurman, Imaging single molecules using total internal reflection fluorescence microscopy (TIRFM). *Cold Spring Harb. Protoc.* (2010), doi:10.1101/pdb.top73.

81. M. P. Klejman *et al.*, NC2alpha interacts with BTAF1 and stimulates its ATP-dependent association with TATA-binding protein. *Mol Cell Biol* (2004), doi:10.1128/MCB.24.22.10072-10082.2004.

82. Z. Zhang *et al.*, Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* (2016), doi:10.1101/gad.285395.116.

83. J. Chen, S. A. Darst, D. Thirumalai, Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc. Natl. Acad. Sci.* (2010), doi:10.1073/pnas.1003533107.

84. M. Djordjevic, R. Bundschuh, Formation of the open complex by bacterial RNA polymerase - A quantitative model. *Biophys. J.* (2008), doi:10.1529/biophysj.107.116970.

85. K. S. Murakami, Structural biology of bacterial RNA polymerase. *Biomolecules* (2015), , doi:10.3390/biom5020848.

86. A. Chakraborty *et al.*, Opening and closing of the bacterial RNA polymerase clamp. *Science*. **337**, 591–5 (2012).

87. K. Potrykus, M. Cashel, (p)ppGpp: Still Magical? *. *Annu. Rev. Microbiol.* **62**, 35–51 (2008).

88. R. M. Corrigan, L. E. Bellows, A. Wood, A. Gründling, ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. *Proc. Natl. Acad. Sci.*, 201522179 (2016).

89. V. Tchernaenko, H. R. Halvorson, M. Kashlev, L. C. Lutter, DNA bubble formation in transcription initiation. *Biochemistry*. **47**, 1871–1884 (2008).

90. T. Cordes *et al.*, Sensing DNA opening in transcription using quenchable Förster resonance energy transfer. *Biochemistry*. **49**, 9171–80 (2010).

91. A. Revyakin, R. H. Ebright, T. R. Strick, Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc. Natl. Acad. Sci.* (2004), doi:10.1073/pnas.0307241101.

92. J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, Identifying molecular dynamics in single-molecule fret experiments with burst variance analysis. *Biophys. J.* (2011), doi:10.1016/j.bpj.2011.01.066.

93. N. C. Robb *et al.*, 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**, 875–85 (2013).

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

95. J. T. Winkelman *et al.*, Multiplexed protein-DNA crosslinking:scrunching in transcription start site selection. *Science (80-. ).* **351**, 1090–1093 (2016).

96. I. O. Vvedenskaya *et al.*, Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol. Cell* (2015), doi:10.1016/j.molcel.2015.10.029.

97. A. J. Carpousis, J. D. Gralla, Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. *J. Mol. Biol.* (1985), doi:10.1016/0022-2836(85)90210-4.

98. D. Pupov, I. Kuzin, I. Bass, A. Kulbachinskiy, Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* (2014), doi:10.1093/nar/gkt1384.

99. L. M. Hsu, N. V. Vo, C. M. Kane, M. J. Chamberlin, 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* (2003), doi:10.1021/bi026954e.

100. N. V. Vo, L. M. Hsu, C. M. Kane, M. J. Chamberlin, In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* (2003), doi:10.1021/bi026962v.

101. L. M. Hsu, Monitoring abortive initiation. *Methods* (2009), , doi:10.1016/j.ymeth.2008.10.010.

102. E. Margeat *et al.*, Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys. J.* **90**, 1419–1431 (2006).

103. L. M. Hsu, Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta - Gene Struct. Expr.* (2002), , doi:10.1016/S0167-4781(02)00452-9.

104. D. C. Straney, D. M. Crothers, A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *J. Mol. Biol.* (1987), doi:10.1016/0022-2836(87)90218-X.

105. B. Krummel, M. J. Chamberlin, RNA Chain Initiation by Escherichia coli RNA Polymerase. Structural Transitions of the Enzyme in Early Ternary Complexes. *Biochemistry* (1989), doi:10.1021/bi00445a045.

106. A. N. Kapanidis *et al.*, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*. **314**, 1144–7 (2006).

107. G. Q. Tang, R. Roy, T. Ha, S. S. Patel, Transcription Initiation in a Single-Subunit RNA Polymerase Proceeds through DNA Scrunching and Rotation of the N-Terminal Subdomains. *Mol. Cell* (2008), doi:10.1016/j.molcel.2008.04.003.

108. F. M. Fazal, C. A. Meng, K. Murakami, R. D. Kornberg, S. M. Block, Real-time observation of the initiation of RNA polymerase II transcription. *Nature* (2015), doi:10.1038/nature14882.

109. J. T. Winkelman *et al.*, Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol. Cell* (2015), doi:10.1016/j.molcel.2015.06.037.

110. Y. Zuo, T. A. Steitz, A structure-based kinetic model of transcription. *Transcription* (2017), , doi:10.1080/21541264.2016.1234821.

111. J. E. Stefano, J. Gralla, Lac UV5 Transcription in Vitro. Rate Limitation Subsequent to Formation of an RNA Polymerase-DNA Complex. *Biochemistry* (1979), doi:10.1021/bi00573a020.

112. W. Kammerer, U. Deuschle, R. Gentz, H. Bujard, Functional dissection of *Escherichia*

*coli* promoters: information in the transcribed region is involved in late steps of the overall process. *EMBO J.* (1986).

113. M. Menendez, A. Kolb, H. Buc, A new target for CRP action at the malT promoter. *EMBO J.* (1987), doi:10.1073/pnas.89.15.7232.

114. P. Eichenberger, S. Déthiollaz, H. Buc, J. Geiselmann, Structural kinetics of transcription activation at the malT promoter of Escherichia coli by UV laser footprinting. *Proc. Natl. Acad. Sci. U. S. A.* (1997), doi:10.1073/pnas.94.17.9022.

115. S. Borukhov, V. Sagitov, a. Goldfarb, Transcript cleavage factors from E. coli. *Cell*. **72**, 459–466 (1993).

116. S. R. Goldman, R. H. Ebright, B. E. Nickels, Direct detection of abortive RNA transcripts in vivo. *Science (80-. ).* (2009), doi:10.1126/science.1169237.

117. J. Skancke, N. Bar, M. Kuiper, L. M. Hsu, Sequence-Dependent Promoter Escape Efficiency Is Strongly Influenced by Bias for the Pretranslocated State during Initial Transcription. *Biochemistry* (2015), doi:10.1021/acs.biochem.5b00272.

118. T. Ellinger, D. Behnke, H. Bujard, J. D. Gralla, Stalling of Escherichia coli RNA polymerase in the +6 to +12 region in vivo is associated with tight binding to consensus promoter elements. *J. Mol. Biol.* (1994), doi:10.1006/jmbi.1994.1388.

119. K. Brodolin, N. Zenkin, A. Mustaev, D. Mamaeva, H. Heumann, The σ70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nat. Struct. Mol. Biol.* (2004), doi:10.1038/nsmb768.

120. S. Samanta, C. T. Martin, Insights into the mechanism of initial transcription in escherichia coli RNA polymerase. *J. Biol. Chem.* (2013), doi:10.1074/jbc.M113.497669.

121. J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, S. M. Block, Backtracking by single

RNA polymerase molecules observed at near-base-pair resolution. *Nature* (2003), doi:10.1038/nature02191.

122. E. Stepanova *et al.*, Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *J. Bacteriol.* (2007), doi:10.1128/JB.00911-07.

123. E. Stepanova, M. Wang, K. Severinov, S. Borukhov, Early transcriptional arrest at Escherichia coli rplN and ompX promoters. *J. Biol. Chem.* **284**, 35702–35713 (2009).

124. G. Feng, D. N. Lee, D. Wang, C. L. Chan, R. Landick, GreA-induced transcript cleavage in transcription complexes containing Escherichia coli RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J. Biol. Chem.* (1994), doi:10.1016/j.bbr.2009.12.039.

125. L. M. Hsu, N. V. Vo, M. J. Chamberlin, Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc. Natl. Acad. Sci.* (1995), doi:10.1073/pnas.92.25.11588.

126. E. Lerner *et al.*, Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *Proc. Natl. Acad. Sci. United States Am.* (2016), doi:10.1073/pnas.1605038113.

127. D. Duchi *et al.*, RNA Polymerase Pausing during Initial Transcription. *Mol. Cell* (2016), doi:10.1016/j.molcel.2016.08.011.

128. S. Sengupta, R. K. Prajapati, J. Mukhopadhyay, Promoter escape with bacterial two-component σ factor suggests retention of σ region two in the elongation complex. *J. Biol. Chem.* (2015), doi:10.1074/jbc.M115.666008.

129. A. N. Kapanidis *et al.*, Retention of transcription initiation factor sigma70 in transcription

elongation: single-molecule analysis. *Mol. Cell*. **20**, 347–356 (2005).

130. S. A. Perdue, J. W. Roberts, σ70-dependent Transcription Pausing in Escherichia coli. *J. Mol. Biol.* (2011), , doi:10.1016/j.jmb.2011.02.011.

131. T. T. Harden *et al.*, Bacterial RNA polymerase can retain $\sigma^{70}$ throughout transcription. *Proc. Natl. Acad. Sci.* (2016), doi:10.1073/pnas.1513899113.

132. I. O. Vvedenskaya *et al.*, Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc. Natl. Acad. Sci.* (2016), doi:10.1073/pnas.1603271113.

133. S. I. Sekine, Y. Murayama, V. Svetlov, E. Nudler, S. Yokoyama, Ratcheting of RNA polymerase toward structural principles of RNA polymerase operations. *Transcription* (2015), doi:10.1080/21541264.2015.1059922.

134. W. Kühlbrandt, Biochemistry. The resolution revolution. *Science (80-. ).* (2014), doi:10.1126/science.1251652.

135. E. Callaway, The revolution will not be crystallized: a new method sweeps through structural biology. *Nat. News*. **525**, 172–174 (2015).

# Chapter 3:

Development of diffusion-based single-molecule FRET methodologies to study different steps in human RNA polymerase II transcription

**Introduction:**

All living organisms use a multi-subunit DNA-dependent RNA polymerase to synthesize RNA, encoding genomic information, as it translocates along DNA during the process of transcription (*1*). In humans, the ~516 kDa RNA polymerase II (Pol II) complex, consisting of 12 subunits (Rpb1-12), transcribes the genes encoding proteins and many noncoding RNAs (*2*). Despite being able to independently perform transcription from certain DNA templates, heteroduplex templates (with a mismatched region mimicking a transcription 'bubble') in the presence of an initiating RNA dinucleotide (*3*) and oligo-(dC) templates (*4*), Pol II requires the action of transcription factors to specifically bind the correct region (core promoter) on double-stranded DNA (dsDNA). A set of basal transcription factors (TF)s—TFIIA, -IIB, -IID, -IIE, IIF, and –IIH—recruit and assemble with Pol II on the proper core promoter (~100 nucleotide region from among the ~ 3 billion nucleotides (nts) of genomic DNA) to form the preinitiation complex (PIC) (*5*).

A sequential model of PIC assembly is hypothesized (*5*) to begin with TFIID, a complex of TATA-box binding protein (TBP) and ~ 13 TBP-associated factors (TAFs). TFIID nucleates PIC assembly by binding the core promoter. Subsequent TFIIA binding stabilizes the interaction of TFIID (*5*–*7*) with core promoter before TFIIB can bind to form the upstream promoter complex (UPC) (*8*). TFIID and TFIIB are the only PIC factors that have been demonstrated to bind known core promoter sequence motifs with subunits of TFIID recognizing the TATA box, initiator (Inr), motif ten element (MTE) and downstream promoter element (DPE); while TFIIB recognizes the upstream and downstream TFIIB recognition element (uBRE & dBRE, respectively) (*5*, *8*, *9*). The UPC can then recruit the Pol II-TFIIF complex to core promoter through interactions between the N-terminal zinc ribbon domain of TFIIB and Pol II (*10*) to form

a PIC that is capable of initiating transcription at the correct transcription start site (TSS) (*11*). This PIC (Pol II-CC) is analogous to the closed promoter complex of bacterial RNA polymerase, with dsDNA running along the top of the cleft, and is capable of spontaneously separating the strands of dsDNA for negatively supercoiled DNA in a region encompassing the TSS to form the open complex (Pol II-OC). However, for linear DNA *in vitro* (*12*) and many promoters *in vivo* (*13*), the inclusion of TFIIE and TFIIH is required for Pol II-OC formation. The energy-dependent unwinding of promoter DNA by TFIIH was shown to mediated by the *Xeroderma pigmentosum complementation group B* (XPB) subunit, which derives energy from hydrolysis ATP/dATP β-γ phosphate bond for its 5' to 3' translocase activity along ~10 nts of the non-template strand (*14*). The combination of TFIID/TBP holding upstream DNA while XPB twists and shuttles ~ 1 turn of downstream DNA towards the PIC induces sufficient strain to separate the DNA strands during Pol II-OC formation. TFIIB, TFIIE and TFIIF then interact with melted DNA strands to trap it in the open state.

Initial transcription begins with the formation of the 1$^{st}$ phosphodiester bond between the first two ribonucleotide triphosphates (NTPs) and frequently results in the release of unstable short RNAs (2-3 nts) during abortive initiation (*15, 16*). Eventually the growing nascent transcript clashes with the TFIIB region blocking the RNA exit channel until displacing TFIIB during promoter escape (RNA ~ 8 nts (*17*)). Numerous structural rearrangements during promoter escape transform the initial transcribing Pol II complex (Pol II-ITC) into a processive elongation complex (Pol II-EC), with the surface of Pol II previously occupied by initiation TFs being bound by the highly conserved elongation factor DSIF (short for 5,6-dichloro-1-β-d-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor) (*18*). Following promoter escape, it has been shown that some PIC TFs remain bound in the promoter region as a scaffold

that facilitates the rapid binding and promoter escape of another Pol II in a process called re-initiation (*19*, *20*). For a large number of metazoan genes, Pol II-EC associated with DSIF and negative elongation factor (NELF) stalls in early elongation (RNA between 20-80 nts). These promoter-proximal paused complexes represent an important regulatory intermediate for many genes, with establishment of the paused complex occurring prior to gene activation (further detail reviewed in (*21*, *22*). Cellular signaling culminates in positive transcription elongation factor b (P-TEFb) inducing Pol II pause escape through its kinase activity (*21*, *22*). The Pol II-EC continues processive transcription, interrupted occasionally by other regulatory pauses, until eventually reaching the gene 3'-end where a large complex of cleavage and polyadenylation factors recognize poly(A)-sites in the transcript before slowing down pol II and resulting in termination through several mechanisms (*23*).

The current mechanistic knowledge of the different steps of Pol II transcription derives largely from extensive work done using traditional biochemical and genetic studies, both *in vitro* and within the cellular context. High resolution structural snapshots of Pol II, individual PIC factors, their assemblies at different stages of initiation (Pol II-CC, OC, ITC) (*2*, *8*, *24–27*) and elongation (*18*, *28*, *29*), and recently of a promoter-proximal paused Pol-II (*30*) have been especially impactful in understanding how the numerous and complex network of molecular interactions facilitate transcription by Pol II. Despite the contributions of these methodologies, a common limitation is their reliance on interpretation of averaged data, which can overlook transient intermediates or subspecies from a heterogeneous sample that could be functionally relevant. Single-molecule (sm) analysis from fluorescence-based (*31*, *32*) or force-based (*33*, *34*) (magnetic or optical tweezers, atomic force microscopy) methodologies provide a means to characterize these subpopulations hidden in the averaged data from ensemble measurements.

Another advantage of sm methodologies is in the real time monitoring of structural dynamics or interactions between biomolecules over different time scales—from sub-seconds for magnetic and optical tweezers, and from nanoseconds to tens of milliseconds for fluorescence microscopy and spectroscopy (*35*).

Fluorescence-based sm techniques can be divided into two general categories depending on whether the labeled target is surface immobilized or freely diffusing (*36*). For the former, the use of total internal reflection (TIRF) microscopy (reviewed in (*37*)) provides a platform for tracking molecules over time to extract time-trace trajectories and enables co-localization analysis, however, surface-induced perturbations, rapid fluorophore photobleaching due to the high intensity of excitation light, and difficulty in achieving statistical significance due to its low-throughput nature are technical challenges (*31*, *36*). The latter approach avoids the limitations of immobilization-based approaches by measuring the fluorescence of a large number of molecules diffusing individually into a small observation volume (~ 1 fL) (*31*). However, a low sample concentration (~ 100 pM) is required to prevent the simultaneous diffusion of multiple molecules through the observation volume and the period for observing dynamics is limited by the residence time within the excitation volume (*31*, *36*). A common tool of fluorescence-based methods is Förster resonance energy transfer (FRET or smFRET for sm methods), which reports on the molecular distance between a pair of fluorophores (*38*). An excited fluorophore (donor) can either fluoresce (shorter wavelength photons detected in donor channel) or transfer its energy (via a non-radiative pathway, FRET) to an adjacent fluorophore (acceptor) if the donor-emission spectrum and acceptor-excitation spectrum overlap, the respective dipoles of the dyes are not orthogonal, and the distance is within the FRET permissive range (2-10 nm) (*38*). The excited acceptor eventually fluoresces (longer wavelength photons

detected in acceptor channel). The efficiency of FRET (E) is sensitive to the inter-dye distance (E is inversely proportional to sixth power of the inter-dye distance) and is computed as the ratio of photons in the acceptor channel to the sum of photons in the donor and acceptor channels. The development of alternating laser excitation (ALEX) smFRET—where the observation volume is continuously exposed to donor and then acceptor excitation wavelength (alternating on a microsecond timescale)—allowed for the sorting of the signal from dually-labeled (donor/acceptor) species from that of contaminating singly-labeled species (*39–41*).

To date, most sm studies of Pol II transcription used the yeast system (*42–60*), however, several studies have been done using human proteins (*61–64*). Here, we report on the development of new diffusion-based smFRET methods with ALEX (smFRET-ALEX) for the *in vitro* study of transcription using purified human Pol II transcription factors. We report on optimizing conditions for RNA detection by single-stranded DNA FRET probes and on the development of smFRET-ALEX assays for (1) monitoring conformational changes in promoter DNA at different stages in initiation, (2) measuring transcriptional activity, and (3) on measuring the promoter-proximal pausing of Pol II.

## Results

*Optimization of FRET probe hybridization efficiency for single-molecule assay*

All FRET probes used in our smFRET assays are short (~20 nt) antisense (complementary to RNA) single-stranded DNAs (ssDNA) that are labeled with a donor and acceptor dye pair. They adopt distinct FRET signatures in the hybridized versus unhybridized forms. We have previously employed FRET probes for the *in vitro* detection of transcripts generated by the highly active *E. coli* RNA polymerase holoenzyme (*65*). However, preliminary

results (data not shown) for transcription reactions with HeLa cell nuclear extract in combination with the basal Pol II system (TBP or TFIID, TFIIB, TFIIF, Pol-II, TFIIE, TFIIH, Pol II) and plasmid DNA gave no detectable low FRET signal (probe hybridized to transcript) unless the activator VP16 was used in the presence of Mediator (*66*), which substantially enhances *in vitro* transcription in conventional radioactive gel assays. The biophysical properties governing oligonucleotide hybridization—including the sequence, length, structure, and buffer conditions—has been extensively studied due to the broad application of oligonucleotide probes. Reducing secondary structure was previously shown to enhance binding between probe and target oligonucleotides (*63*, *67*). To evaluate the effect of oligonucleotide secondary structure on FRET probe binding, we compared the fraction of HP7 FRET probe—containing a hairpin structure—bound to complementary ssDNA containing either a stable (DNA-T1) or an unstable (mut-DNA-T1) secondary structure (Figure 1). The difference between the two ssDNA targets is in the second nucleotide from the 5'-end (G in DNA-T1, T in mut-DNA-T1), which is not included in the region bound by HP7. smFRET-ALEX measurements of HP7 with target DNA yielded FRET histograms with two well-separated distributions (Figure 1A), a population with a higher FRET efficiency (unbound HP7) and a population with a lower FRET efficiency (HP7 hybridized to ssDNA target) that emerges in the presence of ssDNA target. The fraction of hybridized probe (%LF), calculated as the population of low FRET HP7 over the total number of HP7 detected (low FRET + high FRET), was greater when incubated with unstructured mut-DNA-T1 compared to DNA-T1 (Figure 1B). We also always observed increases in the hybridized probe fraction, for different FRET probes, when samples were heated (to melt secondary structures) and then cooled (data not shown). These results demonstrate that

secondary structure in target DNA or RNA present an obstacle for probe binding, which is consistent with current knowledge in the field (*68*).

Another experimentally controllable parameter to optimize probe-target hybridization other than DNA sequence is buffer composition. A thermodynamic barrier to association of two oligonucleotides is the electrostatic repulsion between their negatively charged phosphate backbones, which can be shielded by monovalent or divalent cations to facilitate intermolecular association (*69, 70*). To optimize buffer conditions for hybridization of a FRET probe (RP1) to its ssDNA target (DNA-T2), we tested the effect of varying the concentration of different salts (NaCl, KCl, $MgCl_2$, $CaCl_2$) and denaturants (Urea, Guanidinium Chloride or GdCl) on the fraction of hybridized RP1 (Figure 2). As expected, addition of salt enhanced the low FRET fraction (Figure 2A-i, ii). The greater charge density of the divalent cations, relative to monovalent, significantly enhanced probe binding (Figure 2A, i, ii vs iii, iv). Increasing the ionic radius ($R_i$) of the cation ($R_i$-$Li^+$ < $R_i$-$Na^+$ < $R_i$-$K^+$, data not shown for LiCl) generally increased the low FRET fraction (comparison of $CaCl_2$ vs $MgCl_2$ showed slight enhancement with $CaCl_2$, especially at lower target ssDNA concentrations—data not shown) (Figure 2A). Interestingly, GdCl better enhanced RP1 hybridization to DNA-T2 in comparison to the other monovalent salts—NaCl and KCl—possibly due to its ability to disrupt secondary structure. Urea only slightly enhanced hybridization in comparison to buffer with no salt, however, its effect was significantly lower than that of the other salts and increasing the concentration of urea decreased RP1 binding (Figure 2A-v). Combining GdCl and $CaCl_2$ in the same buffer enhanced the low FRET fraction to a greater extent than for buffer containing either alone (Figure 2B). Control measurements (data not shown) testing RP1 affinity to non-complementary ssDNA

under conditions that optimally enhanced RP1 affinity to target ssDNA demonstrated that the hybridization was specific.

Analogous experiments were done using electrophoretic mobility shift assay (EMSA) to analyze the mobility of ssDNA target with complementary ssDNA probe under different conditions. The ssDNA target used (ss-HSP70-80nt) had the same sequence as the first 80 nts of the HSPA1B transcript; while a set of short anti-hsp70 probes (ah70-probes, table 1) were designed to sequentially target regions differing by 1 bp (ah70-probe 1 binds nucleotides 1-20 of ss-HSP70-80nt, while ah70-probe 2 binds nucleotides 2-21 of ss-HSP70-80nt, etc.). The migration speed for the ss-HSP70-80nt band (Figure 3, lower band) was reduced following probe binding (Figure 3, upper band), likely due to disruption of the predicted secondary structure for ss-HSP70-80nt. Similar to results with smFRET-ALEX, increasing $MgCl_2$ concentration (Figure 3, green dashed box) resulted in the complete shift of the faster migrating ss-HSP70-80nt-alone band to the slower hybridized band (for ah70-probes1-4), while two bands are apparent (with ah70-probes 2-4) for the buffer without $MgCl_2$ (Figure 3, blue dashed box). Control reactions done by heat denaturing and cooling ah70-probes with ss-HSP70-80nt showed complete binding (similar to lanes with $MgCl_2$ in buffer), which indicates the secondary structure of probe and target affecting hybridization (Figure 3).

*Monitoring transcription bubble conformation and dynamics with smFRET-ALEX*

smFRET studies were previously used to report on the conformation and dynamics of promoter DNA bound by bacterial RNA polymerase (*71–77*) and the yeast initiation machinery (*45*, *55*). However, no reports are available for the conformational dynamics of promoters bound by the human PIC. We therefore set out to develop a diffusion-based smFRET-ALEX assay to

monitor such changes. Initial strategies for labeled promoter DNA used fully complementary nontemplate and template strands bearing the adenovirus-2 major late promoter (*78*) (generously provided by Dr. Richard Ebright) that were labeled—at several different positions—within the Pol II-OC bubble. Radioactive gel and smFRET-ALEX measurements (data not shown) of transcription reactions with the minimal Pol II system (TBP, TFIIB, TFIIF, TFIIE, TFIIH, Pol II) failed to produce any RNA products (abortive or runoff) or FRET changes for these templates. To increase the chances of achieving detectable signal, heteroduplex DNA templates with different lengths of internal mismatched DNA regions (different artificial bubble sizes) were tested since they have been shown to enhance *in vitro* activity of Pol II ~100 fold (*11*). These DNA templates contained a super core (SC) promoter (*79*) identical in sequence to that in the cryo-EM structure of the human PIC (*26*). Following smFRET-ALEX experiments that screened different FRET-labeled heteroduplex DNAs—labeled at different positions in the promoter region—only one template (SCP-HD1) yielded discernable changes in FRET at different steps in initiation (Figure 4, 6). SCP-HD1 contains an internal 18-bp mismatched 'bubble' region and is labeled with donor (nontemplate strand) at the +4 register and acceptor (template strand) at the +7 register with respect to the TSS. Comparison of 1D-FRET histograms for SCP-HD1 alone (top left panel of Figure 4) and SCP-HD1 incubated with the minimal Pol II system before (middle left panel of Figure 4) and after (bottom left panel of Figure 4) NTP addition, revealed distinct changes in different regions of the FRET histogram. The 1D-histogram for SCP-HD1 alone gave a high FRET population (HF) (mean proximity ratio or E* = 0.95) with a narrow distribution ($\sigma = 0.05$) (Figure 4, A). SCP-HD1 incubated with the minimal system and an RNA primer (6 nts) resulted in the emergence of a new population with a slightly lower FRET (MF) (mean E* = 0.8) and broader distribution ($\sigma = 0.11$). The addition of NTPs to the PICs (just

before measuring) gave rise to the third FRET population with the lowest apparent FRET (LF) (E* = 0.48) and broadest distribution ($\sigma = 0.15$). The FRET histograms were fitted with a model that is the sum of three Gaussians (Figure 4), each representing a subpopulation (HF, MF, LF) (Figure 4: A, B, C respectively). These initial experiments that characterized distinct subpopulations in initiation, however, had low transcriptional activity—described here as the relative increase in the LF fraction—in the presence of NTP (compare middle and bottom 1D FRET histograms from Figure 4).

The detectable signal did however enable the eventual optimization of conditions that enhance the LF fraction. Surprisingly, it was found that incubating Pol II in buffer (pre-incubation) before combining with the remaining TFs and SCP-HD1 significantly enhanced the LF subpopulation (Figure 5). Comparing the relative change—difference in the subpopulation fractions for each condition in the presence and absence of NTP (Figure 5A)—for either not pre-incubating Pol II or pre-incubating for different times (30, 60 mins), revealed that the LF fractions increases with increasing Pol II pre-incubation time (Figure 5B). However, pre-incubation times greater than 60 mins did not yield any further enhancement (data not shown).

Use of the minimal Pol II system (TBP, TFIIB, TFIIF, and TFIIE) allows for controlling of the PIC composition and through the exclusion or addition of factors, the roles of individual factors can be dissected. Inclusion of TFIIH in the PIC enhanced the LF fraction significantly (Figure 5B, Figure 6). In contrast, addition of DSIF after PIC assembly reduced both the LF and MF fractions (Figure 6). The inclusion of a short (6 nt) RNA primer (same as in the ITC6 structure (*26*)) during PIC assembly caused a drastic increase in the LF subpopulation for different sub-assembly PICs in the presence of NTPs, relative to controls in the absence of NTPs

(Figure 7A). Interestingly, comparing the MF fraction (assembled PIC) in the absence of NTPs, showed no change with or without RNA (data not shown).

It was previously demonstrated by smFRET studies that the transcription bubble exhibits dynamics on the millisecond time scale, which is believed to be important for TSS selection and initial transcription (*73*, *76*, *77*). To evaluate if similar dynamics occur for the human PIC bubble, burst variance analysis (BVA) (*80*) was done on the smFRET-ALEX data on SCP-HD1 (Figure 7B). Contour BVA plots demonstrate that the standard deviation (Figure 7B, triangles) of all sub-bursts within a given binned region ($0 \leq E^{*} \leq 1$, 20 bins) deviates from the theoretical (dashed line in Figure 7B) standard deviation, predicted for a given FRET efficiency based on shot noise, for the LF and MF subpopulations. This deviation suggests dynamics for these subpopulations in the millisecond timescale. The HF subpopulation did not exhibit such dynamics.

*Measuring transcription activity using smFRET-ALEX*

Prior to optimizing the smFRET-ALEX conditions in the transcription bubble assay, many attempts were made to detect RNA, with different FRET probes, produced by the reconstituted Pol II system to no avail. The results of the smFRET-ALEX assay monitoring the transcription bubble indicated the conditions enhancing the NTP-dependent LF population that could be incorporated into the assay for measuring activity. Figure 8 shows a schematic for the protocol used to perform the smFRET-ALEX transcription activity assay. Different heteroduplex DNA templates were designed to be identical to SCP-HD1, except for the having a random sequence for the nontemplate strand mismatch region (instead of poly-dT) and a terminal downstream region that encoded sense sequences complementary to corresponding anti-sense

FRET probes (Figure 9). The FRET probe that succeeded in detecting transcripts contained a highly unstructured poly-dT sequence (20dT probe), which was used previously for transcription activity assays with bacterial RNA polymerase (*65*, *81*). 20dT alone gives a high FRET population that shifts to a lower FRET population after hybridizing to target (Figure 8). The fraction of bound 20dT (TE) was computed as the population of detected 20dT with low FRET over the total detected 20dT in a measurement (calculation of TE shown in Figure 8). Transcription activity was observed with three heteroduplex templates that differ in the size of the internal mismatch—13, 16, and 19 that are denoted as SCP-13rsb, SCP-16rsb, and SCP-19rsb, respectively (Figure 9).

Historically, polyacrylamide gel electrophoresis (PAGE)-based *in vitro* assays with the reconstituted system revealed the sensitivity of transcription activity to the concentration of template DNA. The effect of SCP-19rsb concentration on transcription activity was evaluated by monitoring the relative increase in the low FRET fraction (bound 20dT probe) after NTP addition to PICs (calculation for activity shown in Figure 10). The activity of the minimal Pol II system peaked at 0.75 nM of SCP-19rsb and decreased at lower or greater DNA template concentrations (Figure 10A). Also, for each template concentration, increasing the transcription reaction time (from 20 to 40 mins) resulted in more RNA production (Figure 10A). This is likely due to multi-round synthesis of RNA from the same DNA template. As in the smFRET-ALEX bubble conformation assay, pre-incubation of Pol II in buffer (Figure 8) resulted in an enhancement in activity for transcription with all 3 templates (Figure 10B). Finally, it was also shown that inclusion of TFIIH resulted in greater activity (Figure 10C), which is consistent with TFIIH's known role in enhancing promoter escape (*82*).

*Development of an in vitro smFRET-based Pol II pausing/pause release assay*

We selected the human HSPA1B gene, coding for heat shock protein 70 (HSP70), as our model for developing an smFRET-ALEX Pol II pausing/pause release assay because it is a well-characterized promoter that is regulated by Pol II promoter-proximal pausing (*83–86*). Also, our collaborators (Dylan Taatjes lab) have successfully reconstituted Pol II pausing on the HSPA1B gene *in vitro* using a gel based assay (data not shown, unpublished). The strategy for detecting *in vitro* Pol II pausing with smFRET involves targeting two regions of the RNA transcript at positions upstream and downstream of the DNA regions where Pol II pauses (between 20 to 90 nts downstream of the TSS) (Figure 11). We designed two doubly-labeled ssDNA FRET probes (Figure 11) that target sequences upstream (+4 to +23, USP probe) and downstream (+318 to +338, DSP probe) of the HSPA1B gene pause sites. The USP probe design was selected based on the relative affinity of ah70-probes to ss-HSP70-80nt (ssDNA identical in sequence to the first 80 nts of the HSPA1B transcript) using the gel shifting assay (example in Figure 3, data not shown for all ah70-probes). To avoid disrupting the native secondary structure of the transcript, which could be important for Pol II pausing, trinucleotide regions in the transcript that are likely to be unstructured were targeted (*67*). Due to the low *in vitro* activity of the HSPA1B promoter with the reconstituted Pol II system, the HSPA1B gene was engineered downstream of a strong promoter for *E. coli* RNA polymerase (RNAP) that we have previously used (*81*), the LacCONS promoter, in order to test USP and DSP binding against Hsp70 transcript. Both USP and DSP probes were demonstrated to bind to transcripts generated from reactions with 500 pM RNAP open complexes (Figure 11).

**Conclusion and Future Direction**

Over the past decade, several TIRF-based studies were done, with immobilization of DNA template to a surface, using the human reconstituted transcription system (*61–64*) (reviewed in (*87*)). Here we report on the development of diffusion-based smFRET methods for studying different steps in transcription by the reconstituted human Pol II system. Measuring the fluorescence of molecules transiting an illuminated confocal spot offers several advantages to immobilized-based approaches; it allows for a higher throughput screening of molecules since ~10 molecules (depending on concentration) pass through the confocal volume every second; it simplifies the experimental approach by avoiding the immobilization and surface passivation steps; it avoids complications that can arise due to the interaction between the surface and biomolecules, such as the inactivation of TFIID by the surface in a previous study that was alleviated by treatment of the surface with an additional polymer (63); and it also exceeds the temporal resolution for most of immobilization-based studies (~10~100 ms), reaching a few tens of microseconds resolution (88). The immobilization-based approach on the other hand has the advantage of directly tracking the trajectory of the same molecules over time (up to minutes), while the diffusion-based approach can only monitor when an individual molecule resides in the detection volume (~ms) or relies on the observation of a population of molecules. We reported here on the (1) optimization of FRET probe design and solution conditions to enhance specific targeting of FRET probes to their targets, (2) development of an smFRET-ALEX method to observe dynamic changes in promoter DNA conformation during transcription initiation by the human Pol II system, (3) development of an smFRET-ALEX method to quantitatively assess the transcriptional activity of the minimal Pol II system, and (4) proof-of-principle results on developing an smFRET-ALEX assay for studying Pol II pausing and pause release.

The affinity of two FRET probes to complementary ssDNA was enhanced by either disrupting the secondary structure of ssDNA target through a single nucleotide alteration or by including different salts in solution, with the higher charge and larger radius cations stimulating hybridization to a greater extent. We have also observed that reducing the secondary structure in the FRET probe enhances hybridization (20dT > RP1 or HP7, data not shown), which was also observed previously with other fluorescent probes (*63*). The optimized probe hybridization conditions were adapted for the Pol II smFRET-ALEX assay measuring RNA product of transcription reactions.

The conformational change and dynamics of the fluorescently-labeled transcription bubble was monitored at different steps in transcription initiation with Pol II. Three distinct populations were identified—the high FRET population for DNA alone or Pol II-CC, the medium FRET population for the open bubble with bound PIC, and the NTP-dependent low FRET population for a more open state of the bubble. Comparison of these relative fractions under different conditions revealed an enhancement in the low FRET (more open state of the bubble) population by incubating Pol II in buffer prior to mixing with other PIC factors and by inclusion of TFIIH, while DSIF seemed to have the opposite effect in lowering the low FRET population. The effects of TFIIH and DSIF are consistent with their known functions in promoter opening and in facilitating reannealing of the upstream edge of the transcription bubble during promoter escape (*88*), respectively.

Immobilized-based smFRET was previously used to detect transcripts produced by the reconstituted human Pol II system and evaluate the effect of promoter mutations and an activator (Sp1) on transcriptional activity (*63*). Here we demonstrated an smFRET-ALEX assay for measuring the transcriptional output of the reconstituted human Pol II system. The technical step

of incubating Pol II in buffer before combining with the remaining factors for PIC assembly was shown to enhance activity, which is consistent with the enhancement in the low FRET population of the transcription bubble assay. We also showed TFIIH-dependent enhancement in activity that further supports the previous findings that the XPB subunit of TFIIH enhances promoter escape (*82*).

These new smFRET methods establish a platform to dissect out the contributions of promoter DNA elements, individual PIC factors, and regulators of the PIC (Mediator, transcription factors) to different steps in transcription. The effect of TFIIH in maintaining the transcription bubble in a more open state offers interesting avenues for investigating a possible role of TFIIH and other PIC factors in transcription re-initiation. Inclusion of B2 RNA, which can bind free Pol II, will be done in the future to ensure that any transcription is restricted to already pre-formed PICs and not any re-initiation events. The increased transcription activity at 40 mins vs. 20 mins (Figure 10) provided further evidence for possible re-initiation. Current work is also being undertaken to identify the TFIIH subunit responsible for the enhancement in transcription activity. The inhibitors triptolide (*89*) and THZ1 (*90*) are currently being tested using the smFRET-ALEX Pol II activity assay to determine if the XPB helicase or CDK7 kinase is responsible, respectively. Work is also currently being done to transition to the use of fully complementary DNA templates instead of the heteroduplex templates used so far, especially given the important role of transcription bubble reannealing in promoter escape (*91*). Furthermore, experiments monitoring the conformational change of promoter DNA labeled at additional sites (Figure 12) are currently being undertaken to better characterize the 3D dynamic structure of the transcription bubble, as was recently done in bacteria (*76*).

**Methods**

FRET probe hybridization to ssDNA target.

The ssDNA FRET probes, containing a hairpin HP7 (5'-GCTAATTTTTTCTCCATTTTAGCTTCCTTAG-3') and without a hairpin RP1 (5'-CTCCATTTTAGCTTCCTTAGCT-3'), dually labeled with tetramethylrhodamine (TMR, donor) and Alexa647N (acceptor) were ordered from Lumiprobe Corporation. The unlabeled ssDNA oligonucleotides DNA-T1 (5'-AGCTAAGGAAGCTAAAATGGAGAAAAAAAT-3'), mut-DNA-T1 (5'-ATCTAAGGAAGCTAAAATGGAGAAAAAAAT-3'), and DNA-T2 (5'-AGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAAT-3') were ordered PAGE purified from Integrated DNA Technologies, Inc. (IDT).

Hybridization reactions were performed by incubating 100 pM of RP1 with different concentrations of DNA-T2 target in either TB-S0 buffer (50 mM Tris-pH 8.0, 1 mM DTT, 5% glycerol, 100 µg/mL Bovine Serum Albumin (BSA)) alone or in TB-S-X (TB with salt set to X mM, e.g. TB-NaCl-200 being TB with 200 mM NaCl) in a total volume of 20 µL for 20 minutes (min(s)) at room temperature (rt). Hybridization of 100 pM HP7 with 1 nM of DNA-T1 or mut-DNA-T1 was done using TB-KCl-100. After incubation, samples were immediately pipetted onto a glass coverslip (Fisher Scientific Company) and made into a chamber (to prevent evaporation) before loading on the smFRET-ALEX setup and measuring for 15 mins (532 nm and 638 nm lasers were set to 160 µW and 70 µW, respectively).

EMSA for probe hybridization to ssDNA target.

All ssDNA oligonucleotides, ss-HSP70-80nt (5'-GGAAAACGGCCAGCCTGAGGAGCTGCTGCGAGGGTCCGCTTCGTCTTTCGAGAGTGACTCCCGCGGTCCCAAGGCTTTCC-3') and ah70-probes (ah70-probe 1: 5'-CCTCAGGCTGGCCGTT

TTCC-3'; ah70-probe 2: 5'-TCCTCAGGCTGGCCGTTTTC-3', ah70-probe 3: 5'- CTCCTCAG GCTGGCCGTTTT-3', ah70-probe 4: 5'- GCTCCTCAGGCTGGCCGTTT-3'), were ordered from IDT. Hybridization of 67 µM of ss-HSP70-80nt with 100 µM of an ah70-probe was done in TE50 (10 mM Tris-pH 8.0, 50 mM NaCl, 1 mM EDTA) or TE50-100$MgCl_2$ (TE50, 100 mM $MgCl_2$) in a total volume of 15 µL at rt for 15 mins. For samples testing heat denaturation on hybridization, tubes were incubated in a thermocycler at 95 °C for 3 mins before cooling down to 10 °C at a rate of 0.1 °C/second. Sample were each mixed with 3 µLs of 70% glycerol before loading into a 15% TBE (45 mM Tris, 45 mM Borate, 0.5 mM EDTA) polyacrylamide gel, which was run at 90 V for 1.5-2 hours. Gels were then stained with SYBR gold (Thermo Fisher Scientific) for 20 mins and imaged using a Biorad Molecular Imager FX Pro Plus.

*In vitro* Transcription Assay with labeled promoter DNA.

The fluorescently labeled nontemplate strand of SCP-HD1 (5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGTTTTTTTTTTTTTTTTTTTTTTCGAACACT CGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3', Atto550 at underlined residue) and template strand of SCP-HD1 (5'-ACTGGGGAATTCCATGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGA CTGAGGACGAACGCGCCCCCACCCCCTTTTATAGGCGCCCTTC-3', Atto647N at underlined position) and the 6-nt RNA primer (5'-AGUCGC-3') were ordered (PAGE and reverse HPLC purified) from IBA Lifesciences. SCP-HD1 was prepared by combining equimolar ratios of each strand in TMK buffer (10 mM Tris-pH 8.0, 100 mM KCl, 2 mM $MgCl_2$) and using a thermocycler programmed with the following parameter: (1) heating to 95 °C for 3 minutes; (2) for the temperature range $T_{Tm+5°C}$ to $T_{Tm-5°C}$ ($T_{Tm+5°C}$ is 5 °C greater than Tm, $T_{Tm-}$

$_{5°C}$ is 5 °C less than Tm), a 10 minute incubation at each temperature (1 °C intervals); (3) incubation at 3 minutes at each temperature (1 °C intervals) until reaching 45 °C; and finally (4) the thermocycler was cooled to 4 °C (All cooling steps were set at a rate of 0.1 °C/second).

All components of the reconstituted human transcription system (Pol II, TBP/TFIID, TFIIB, TFIIF, TFIIE, TFIIH, NELF, DSIF) were provided by our collaborators at the Dylan Taatjes Lab (University of Colorado, Boulder). The Taatjes lab previously established the *in vitro* reconstituted human Pol II system (*92*) and provided technical information on what final amounts of Pol II and each factor to include in the transcription reaction, by titrating each newly purified factor, using radioactive PAGE gel analysis, to determine the amount needed to produce the maximum levels of RNA product. Dilution of TBP, TFIIB, TFIIF, Pol II, and the larger subunit of TFIIE (56 kDa (*92*)) were done in DB(100) buffer (20 mM Tris-pH 7.9, 100 mM KCl, 1 mM dithiothreitol (DTT), 20 µg/mL BSA, 20% (v/v) glycerol). A 2 nM stock of SCP-HD1 and 100 µM stock of 6-nt RNA were prepared in DB(100) buffer. Transcription reactions without any Pol II pre-incubation step were initiated by adding 1 µL of 2 nM SCP-HD1, 100 µM 6 nt RNA, and HDM buffer (160 mM Hepes-pH 7.6, 64 mM $MgCl_2$, 1 mM DTT) to 13 µLs of factor mix (DB(100)+components of the minimal system) (16 µLs reaction volume) and immediately incubated for 20 minutes at 30 °C to assemble PICs. Reactions that included Pol II pre-incubation, were initiated by incubating Pol II and DB(100) (total 5 µL) at 30 °C for a given time before adding 8 µLs of factor mix (DB(100)+remaining components of the minimal system) and then 1 µL of 2 nM SCP-HD1, 100 µM 6 nt RNA, and HDM buffer before incubating for 20 minutes at 30 °C to assemble PICs. For reactions including TFIIH, DSIF, and NELF; these factors were added, along with DB(100) if needed, after the incubation step for PIC assembly to bring the reaction volume to 19 µLs before mixing. Then either 1 µL of DB(100) buffer (no

NTP control) or 1 µL of 4 mM NTP (in DB(100) buffer) was added before mixing and immediately transferring to a glass cover slip (Fisher Scientific, Premium Cover Glass), assembling a chamber and quickly starting the smFRET-ALEX measurement. Measurements were done for 20 mins (532 nm and 638 nm lasers were set to 170 µW and 70 µW, respectively).

*In vitro* Transcription Activity Assay with antisense FRET probe.

The SCP-13rsb, SCP-16rsb and SCP-19rsb heteroduplex DNA templates were prepared by hybridizing the same template strand (5'- CCGCTTTTTTTTTTTTTTTTTTTTGCACGTCT GCTCGGCTCGAGTGTTCGATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTAT AGGCGCCCTTC-3') with the SCP-13rsb nontemplate (5'- GAAGGGCGCCTATAAAAGG GGGTGGGGGCGTCAAGCAGGCTTCTCGCGATCGAACACTCGAGCCGAGCAGACGTG CAAAAAAAAAAAAAAAAAAAGCGG-3'), SCP-16rsb nontemplate (5'- GAAGGGCG CCTATAAAAGGGGGTGGGGGCGTCAAGCAGGCTTCTGCCGATCGAACACTCGAGCC GAGCAGACGTGCAAAAAAAAAAAAAAAAAAAGCGG-3'), and SCP-19rsb nontemplate (5'- GAAGGGCGCCTATAAAAGGGGGTGGGGGCGTCAAGCAGGCTTCTGAGATTCGA ACACTCGAGCCGAGCAGACGTGCAAAAAAAAAAAAAAAAAAAGCGG-3'). The hybridization method was the same as that for SCP-HD1 (above section).

Transcription reactions were performed identically to that described for the with the SCP-HD1 template (above section) until the step of either NTP or DB(100) addition. However, an additional incubation, following NTP or buffer addition, was done at 30 °C for 20 mins. Transcription reactions were then quenched by adding 3 µLs of Stop solution (4 M GdCl, 0.67 M $MgCl_2$). 1 µL of 20dT probe to a final concentration of 100 pM before incubating at rt for 20 mins.. Solutions were then either immediately measured by smFRET-ALEX, as mentioned

above for SCP-HD1, or thawed in liquid nitrogen for measurement at a later time. Freezing samples in liquid nitrogen did not affect FRET histogram (data not shown).

*In vitro* Pol II pausing/pause release assay

The DNA template containing the LacCONS promoter and HSPA1B gene sequence to position +360 (downstream of TSS) was ordered from SGI-DNA, Inc. Preparation of RNAP open complexes with E. coli RNAP holoenzyme (NEB, M0551S) was done as previously described (*81*). Transcription reactions were carried out by addition of either transcription buffer (*81*) or NTPs to a final concentration of 100 µM to 0.5 nM of open complexes (calculated based on template DNA concentration) in a final volume of 40 µL before incubating at 37 °C for 20 mins. The reaction was then quenched by adding 4 µLs of 6M GdCl, followed by addition of either USP or DSP to 100 pM final concentration in a total of 48 µLs and incubating for 20 mins at rt. smFRET-ALEX were done for 10 mins (532 nm and 638 nm lasers were set to 170 µW and 70 µW, respectively).

smFRET setup and analysis

Details on the ALEX-FAMS setup used for all measurements are described in (*81*). The FRETbursts software package was used for all analysis (*93*). Analysis of dual-colored time stamps obtained from the ALEX-FAMS was analyzed using the same burst search algorithm and parameters described before (*81*). The dual channel burst search to isolate the FRET only subpopulation was also performed as described before (*81*), however, the threshold for each channel was set to 20 instead of 25:

i) $(n_{DA} - lk \cdot n_{DD} - dir \cdot n_{DD}) + \gamma \cdot n_{DD} \geq 20$

ii) $n_{AA} \geq 20$

All FRET histograms for the transcription activity measurements are globally fitted to a sum of two Gaussians. For FRET histograms of the transcription bubble conformation, the measurements are globally fitted to a sum of three Gaussians. In the global model, the means and widths of the subpopulations are constrained to be constant as a function of time (i.e., the same for all datasets), whereas the amplitudes are left free to vary for each time point.

Illustrations

All illustrations were done using Inkscape software.

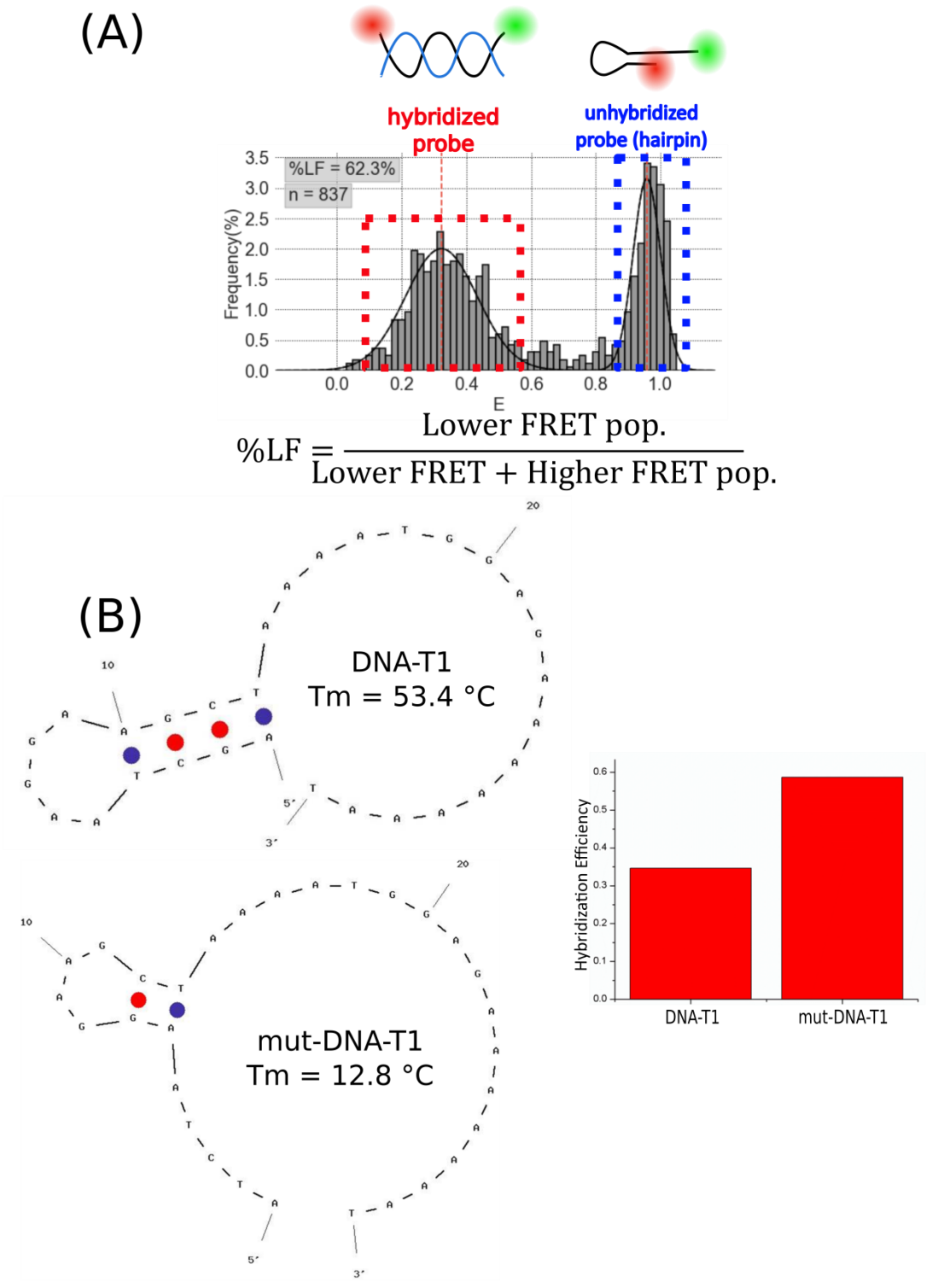



**Figure 1.** Effect of ssDNA structure on FRET probe hybridization. **(A)** Extraction of hybridization efficiency from 1D-FRET histogram. Histogram data were fitted to the sum of two Gaussians (black solid line), a lower FRET (red dotted box) and a higher FRET (blue dotted box)

populations are for hybridized and unhybridized probes, respectively. The %LF or relative fraction of hybridized HP7 probe was computed as the ratio of the area of the low FRET population to the total fitted area. **(B)** Left: Predicted secondary structure and $T_m$ (UNAfold software(*94*, *95*) on IDT website) for ssDNA targets DNA-T1 and mut-DNA-T1. Right: Bar graph showing the hybridization efficiency (%LF) of HP7 probe binding to either DNA-T1 or mut-DNA-T.

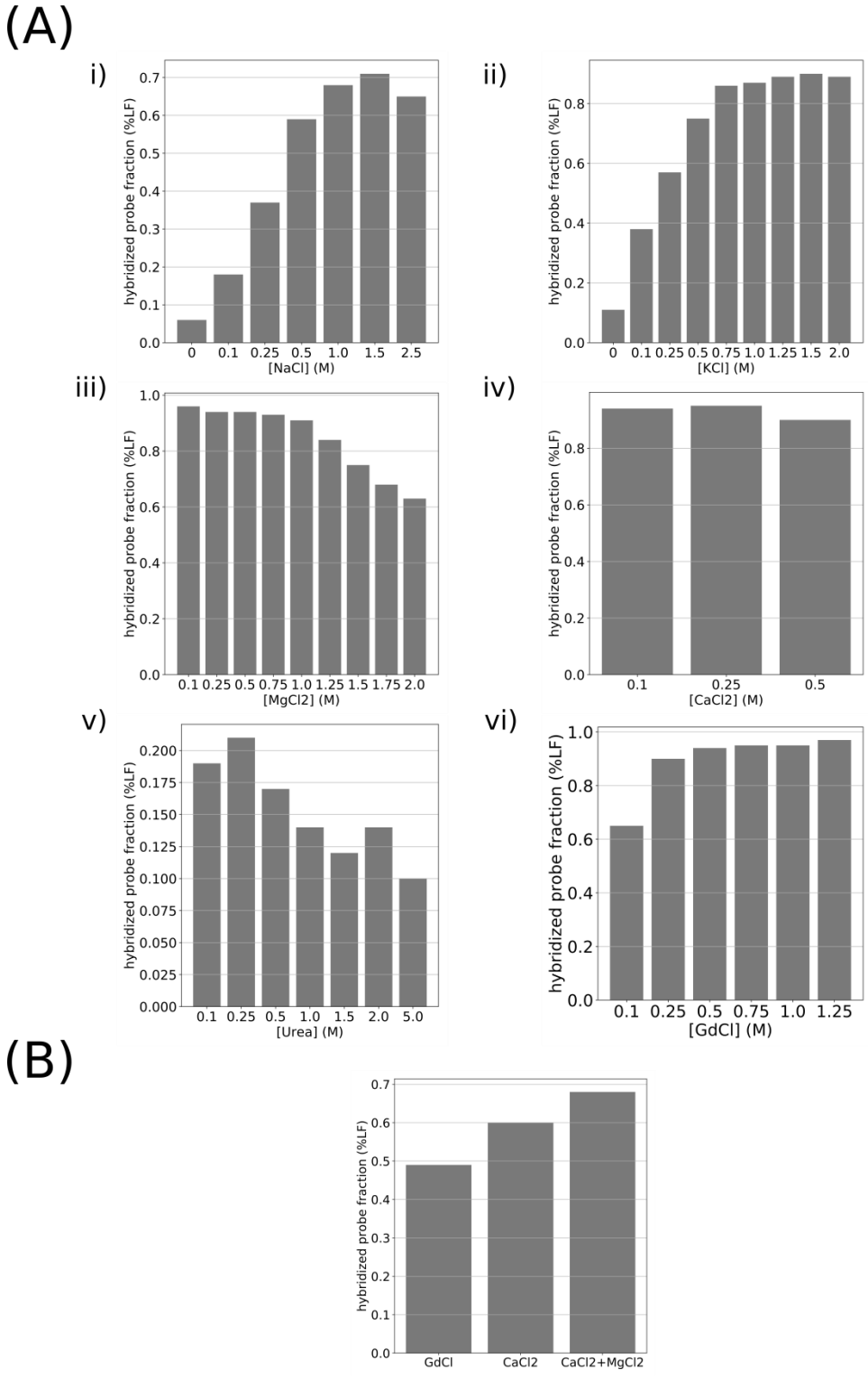



**Figure 2.** Effect of different salts on RP1 probe hybridization to ssDNA. **(A)** Bar graphs showing the hybridized probe fraction (%LF) of RP1 binding to 500 pM DNA-T1 at different

salt concentrations for (i) NaCl, (ii) KCl, (iii) MgCl2, (iv) CaCl2, (v) Guandinium Chloride (GdCl), and (vi) urea. **(B)** The hybridized probe fraction of RP1 binding to 100 pM DNA-T1 for GdCl, $CaCl_2$, and GdCl+$CaCl_2$.

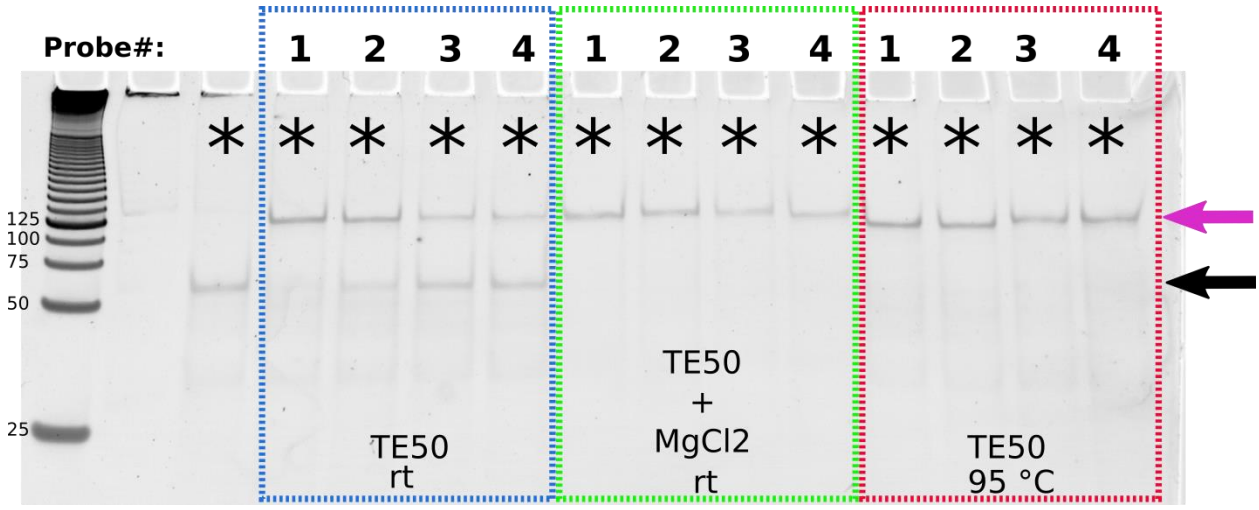



**Figure 3.** Effect of $MgCl_2$/heat denaturation on probe-target hybridization analyzed using EMSA. Different ah70-probes (top: ah70-probe #) were tested against ss-HSP70-80nt target (* indicates lanes with ss-HSP70-80nt) under different conditions: in TE50 buffer at rt (blue box), TE50+100 mM MgCl2 buffer at rt (green box), and TE50 buffer at 95 °C before cooling to 10 °C (red box). Oligonucleotides were separated through a 15% TBE PAGE gel and stained using SYBR Gold. Arrows indicated band for ss-HSP70-80nt ssDNA (black arrow) and for ss-HSP70-80nt hybridized to probe (magenta arrow). Band size was determined in reference to 25-bp Ladder from Invitrogen (DNA size in bp shown).

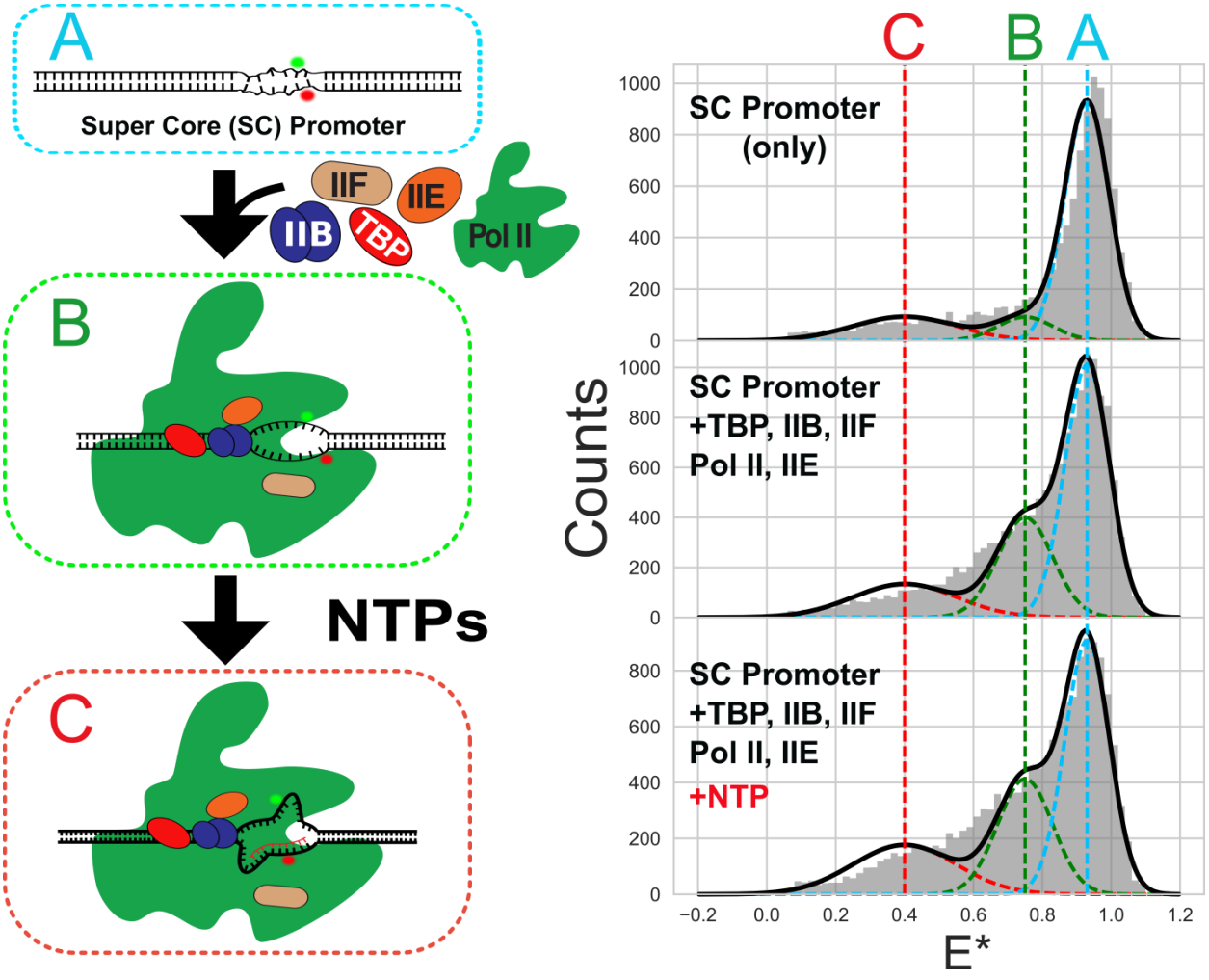
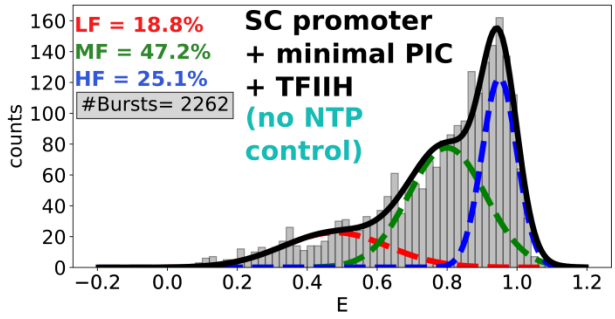
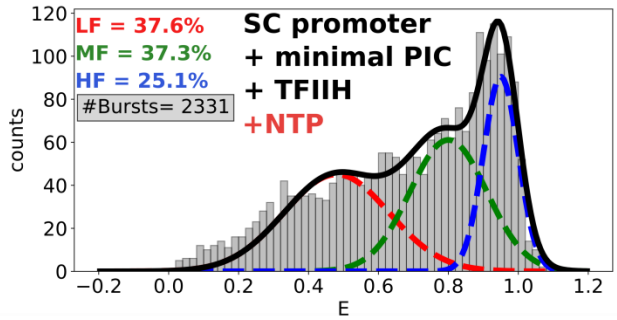



**Figure 4.** Three proposed states in initiation for labeled promoter DNA. To assess promoter structure during transcription initiation, DNA (SCP-HD1) containing super core (SC) promoter was labeled on the non-template (+4 relative to TSS, donor) and template (+7 relative to TSS, acceptor). **Left:** Schematic for the 3 proposed states of SCP-HD1 DNA: **(A)** relatively closed transcription bubble SCP-HD1 alone, **(B)** slightly open transcription bubble after PIC assembly, and **(C)** expanded transcription bubble following NTP addition. **Right:** 1D-FRET histograms from smFRET-ALEX measurements of different stages of initiation**;** histograms were fit as the sum of 3 Gaussians with each subpopulation (LF: lower FRET; MF: middle FRET; HF: Higher

FRET subpopulations) having a mean FRET efficiency **(vertical dotted lines)**. **E*** is the proximity ratio.

A




Relative Change (%) = s.p.(+NTP) - s.p.(no NTP)

B

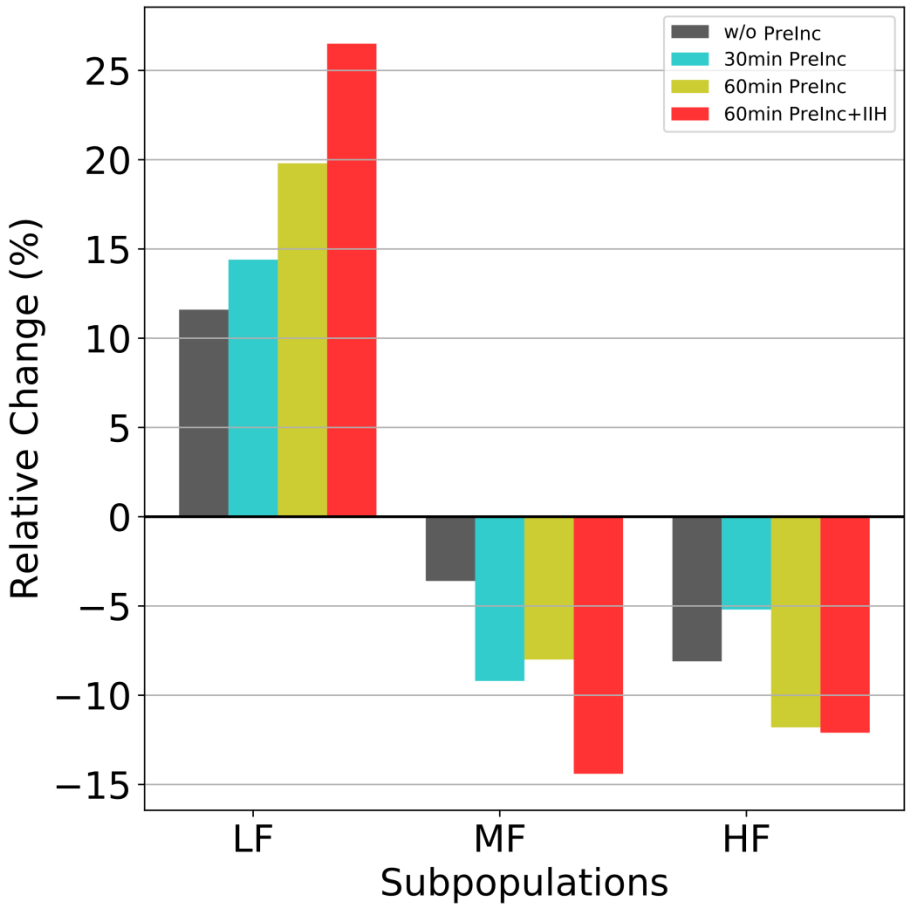


**Figure 5.** The NTP-dependent LF (lower FRET) state (more open state of transcription bubble) is enhanced by TFIIH and by pre-incubating Pol II (incubating Pol II in buffer before addition of DNA template and other TFs). **(A)** To compare different transcription reactions conditions, the change in each subpopulation **(LF, MF, HF)** after NTP addition was determined with respect to corresponding controls lacking NTP. The relative change for each subpopulation is calculated as the difference in the amplitude (or area) of each subpopulation (s.p.) in the presence or absence of NTP. The total number of bursts and the amplitude for each subpopulation is shown in the upper left of each histogram. E is the proximity ratio. **(B)** The relative change for each subpopulation for transcription reactions performed under different conditions. Pol II was either mixed simultaneously with TFs and DNA to assemble the PIC (w/o PreInc) or it was incubated in buffer at 30 °C for different times (PreInc) before adding the TFs and DNA. Inclusion of TFIIH to the minimal PIC resulted in the greatest shift of the histogram to the LF subpopulation.

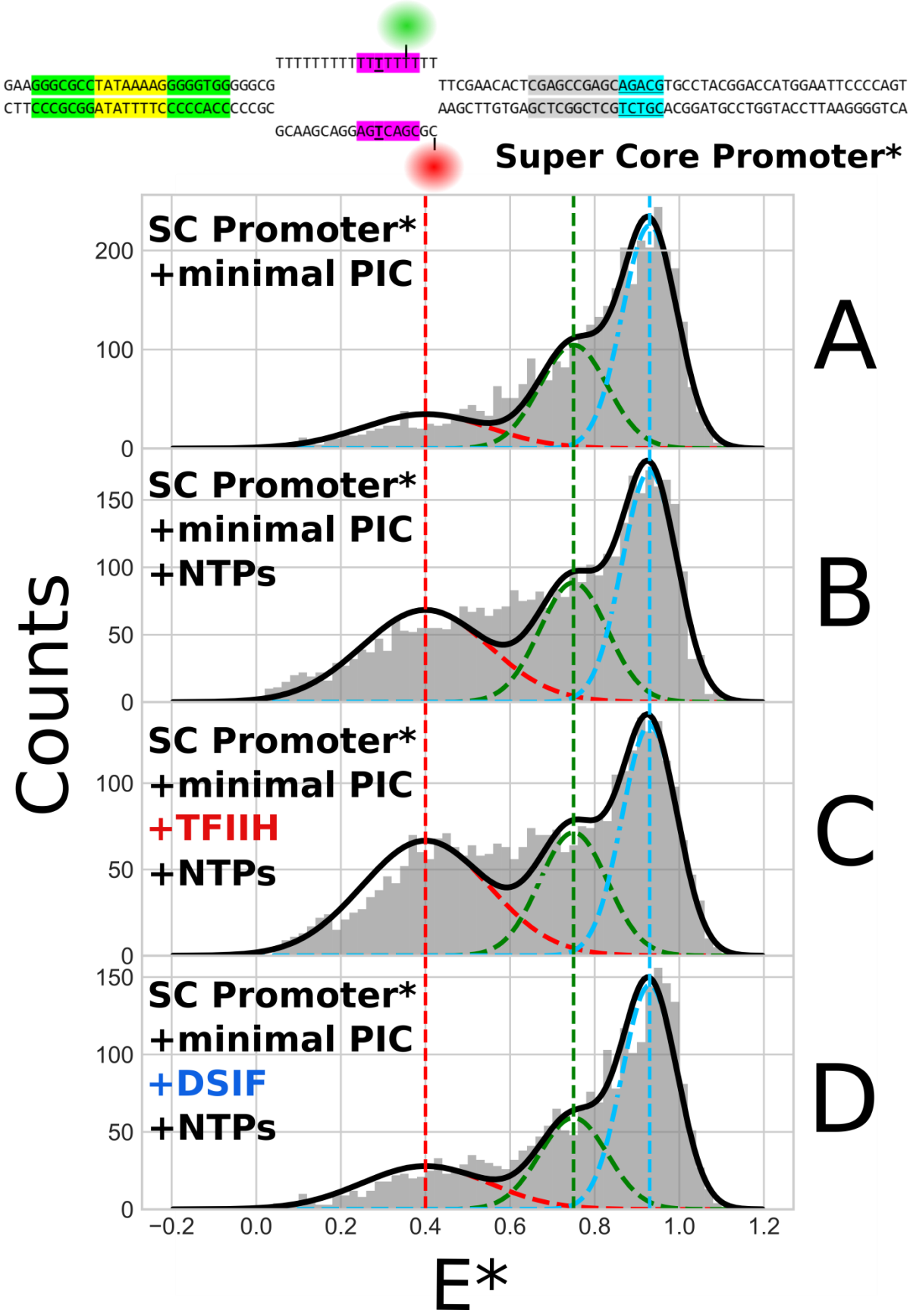
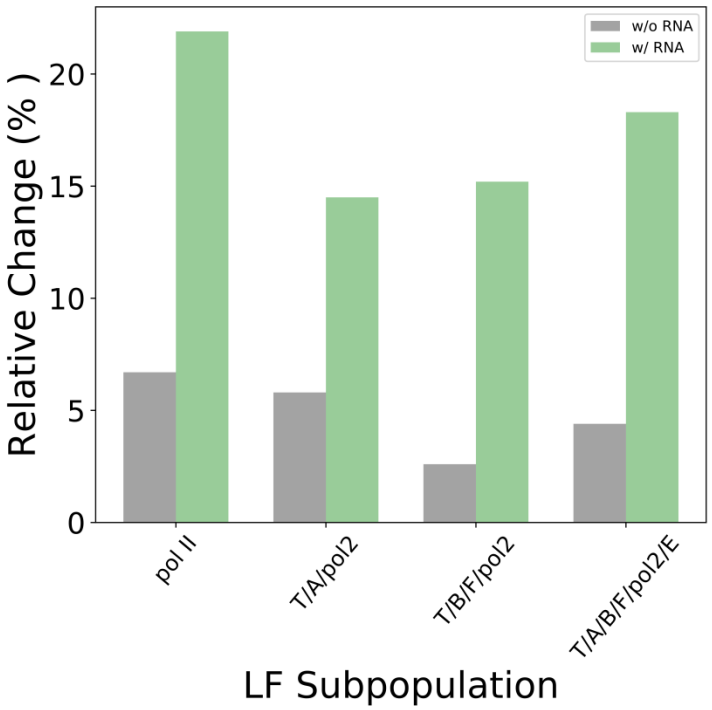

TTTTTTTTTTTTTTTTT
GAAGGGCGCCTATAAAAGGGGGTGGGGGCG
CTTCCCGCGGATATTTTCCCCCACCCCCGC
TTCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT
AAGCTTGTGAGCTCGGCTCGTCTGCACGGATGCCTGGTACCTTAAGGGGTCA
GCAAGCAGGAGTCAGCGC
Super Core Promoter*
SC Promoter*
+minimal PIC
A
SC Promoter*
+minimal PIC
+NTPs
B
SC Promoter*
+minimal PIC
+TFIIH
+NTPs
C
SC Promoter*
+minimal PIC
+DSIF
+NTPs
D
Counts
E*
200
100
0
150
100
50
0
100
50
0
150
100
50
0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

**Figure 6.** TFIIH appears to maintain a more open promoter; DSIF appears to promote re-annealing. **Top:** labeled SCP-HD1 heteroduplex DNA template (containing mismatch region mimicking initiation bubble) with SC promoter used. A-D) smALEX-FRET assays. Low-FRET population (red line) proposed to represent distinct open state. Addition of NTPs (B) increases low-FRET state, and this is increased further with TFIIH (C). By contrast, DSIF markedly reduces low-FRET state (D).

A


w/o RNA
w/ RNA
Relative Change (% )
0
5
10
15
20
pol II
T/A/pol2
T/B/F/pol2
T/A/B/F/pol2/E
LF Subpopulation

B

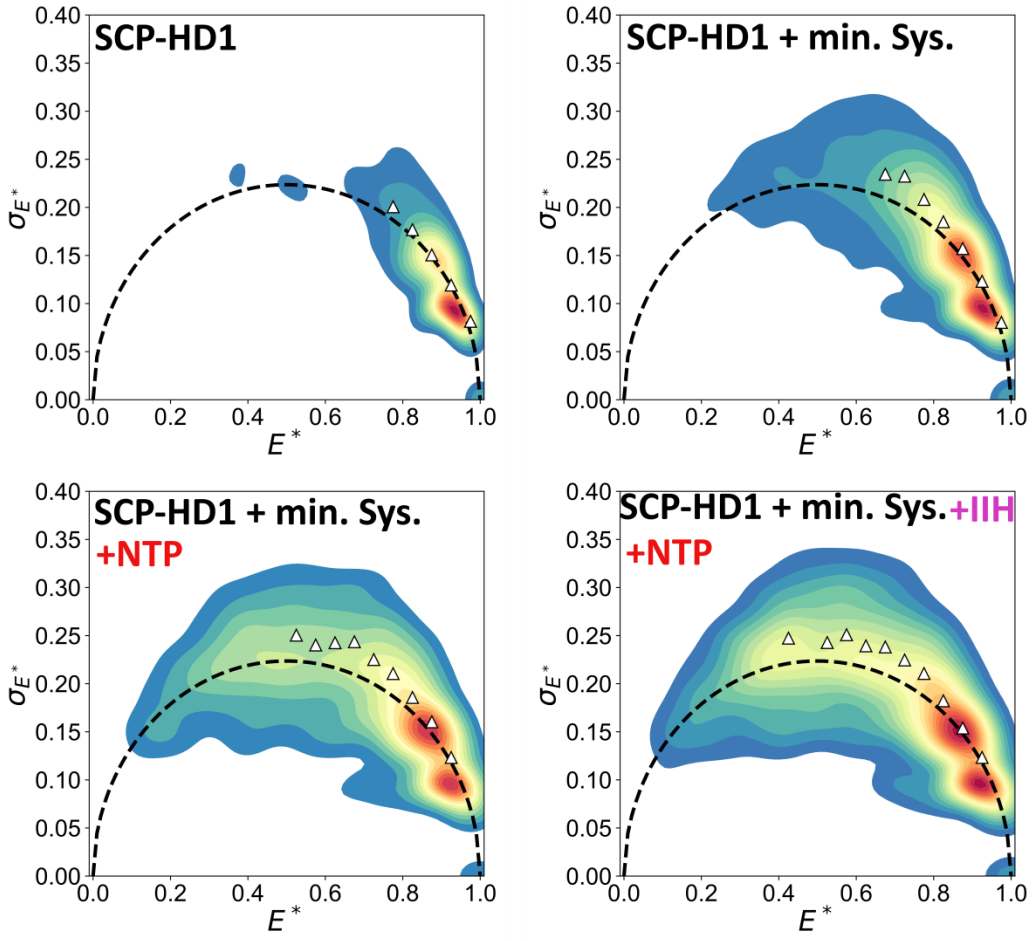
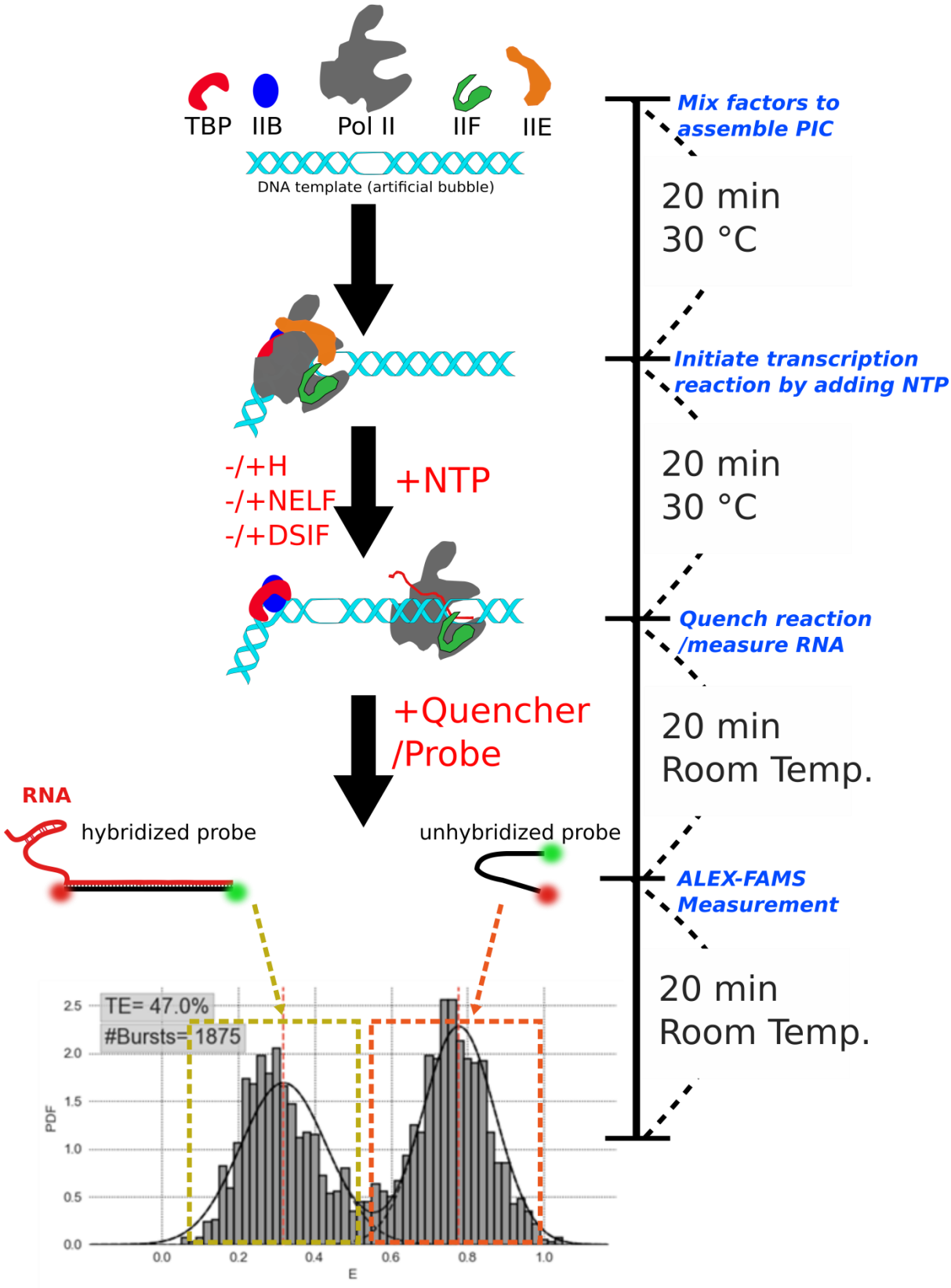

SCP-HD1
SCP-HD1 + min. Sys.
SCP-HD1 + min. Sys.
+NTP
SCP-HD1 + min. Sys.+IIH
+NTP
$\sigma_{E^*}$
$E^*$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.0
0.2
0.4
0.6
0.8
1.0

**Figure 7.** 6-mer RNA primer enhances transcriptional activity; MF and LF subpopulations exhibit millisecond dynamics. **(A)** Relative change in LF subpopulation for different PICs (Pol II alone vs Pol II+TFIIA vs Pol II+TFIIB+TFIIF vs Pol II+TFIIA+TFIIB+TFIIF+TFIIE) is enhanced by inclusion of an RNA primer. **(B)** Contour plot for BVA analysis of smFRET-ALEX measurements of SCP-HD1 promoter DNA alone, SCP-HD1 with the Pol II minimal system with and without NTP, and SCP-HD1 with a minimal system that includes TFIIH in the presence of NTP. Triangles in the BVA plot represent calculated standard deviations of all sub-bursts (containing 5 photons) in the binned regions ($0 \leq E^{*} \leq 1$, 20 bins), from their mean E* in order to increase the statistical power of BVA (*80*). In addition, to focus on fluctuations from the sample dynamics (rather than photo-physicals of fluorophores), only binned regions, which contain bursts more than or equal to thresholds (5% of total number of bursts) are considered for the calculated standard deviations.

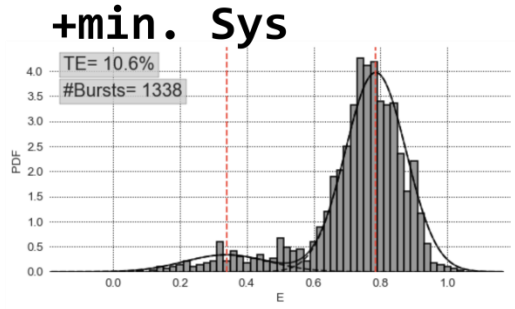


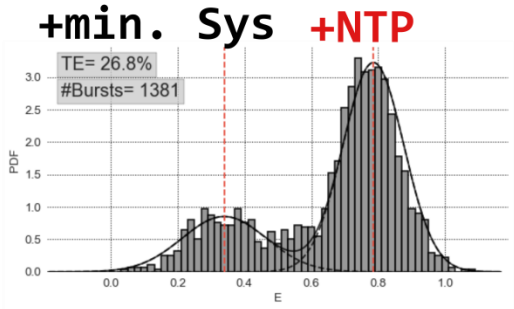

$$\text{TE}\,(\%) = \frac{\text{Low FRET pop. (hybridized probe)}}{\text{Low FRET pop. (hybridized probe)} + \text{High FRET pop. (unhybridized probe)}} \times 100$$

**Figure 8.** *In vitro* smFRET-ALEX activity assay schematic. Components of the Pol II minimal system are mixed with heteroduplex template to form PIC. Additional factors (TFIIH, NELF, DSIF) are added just before NTP addition. Transcriptional efficiency (TE%) is calculated according to equation.


# SCP-13rsb

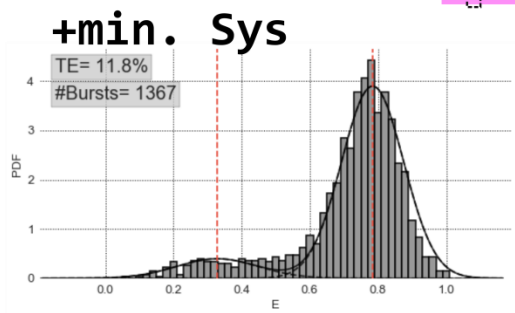
TCAAGCAGGCTTC
GAAGGGCGCCTATAAAAGGGGGTGGGGGCG TCGCGATCGAACACTCGAGCCGAGCAGACGTGCAAAAAAAAAAAAAAAAAAAAAGCGG
CTTCCCGCGGATATTTTCCCCCACCCCCGC AGCGCTAGCTTGTGAGCTCGGCTCGTCTGCACGTTTTTTTTTTTTTTTTTTTTCGCC
GCAAGCAGGAGTC

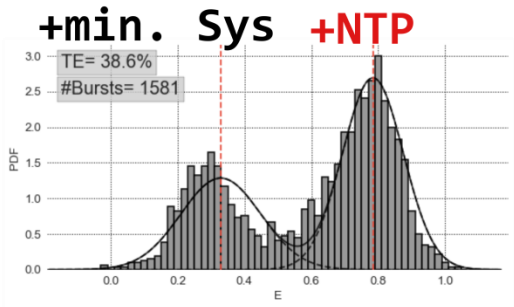



# SCP-16rsb

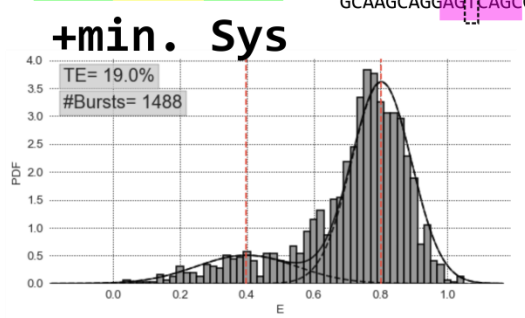
TCAAGCAGGCTTCTGC
GAAGGGCGCCTATAAAAGGGGGTGGGGGCG CGATCGAACACTCGAGCCGAGCAGACGTGCAAAAAAAAAAAAAAAAAAAAAGCGG
CTTCCCGCGGATATTTTCCCCCACCCCCGC GCTAGCTTGTGAGCTCGGCTCGTCTGCACGTTTTTTTTTTTTTTTTTTTTCGCC
GCAAGCAGGAGTCAGC




# SCP-19rsb

TCAAGCAGGCTTCTGAGAT
GAAGGGCGCCTATAAAAGGGGGTGGGGGCG TCGAACACTCGAGCCGAGCAGACGTGCAAAAAAAAAAAAAAAAAAAAAGCGG
CTTCCCGCGGATATTTTCCCCCACCCCCGC AGCTTGTGAGCTCGGCTCGTCTGCACGTTTTTTTTTTTTTTTTTTTTCGCC
GCAAGCAGGAGTCAGCGCT



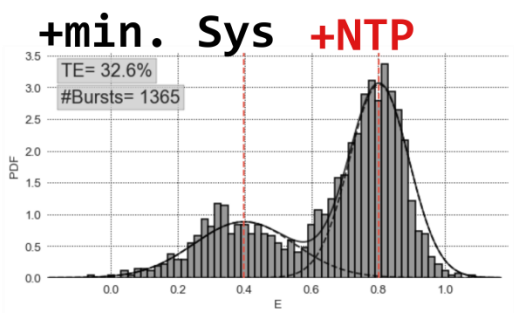
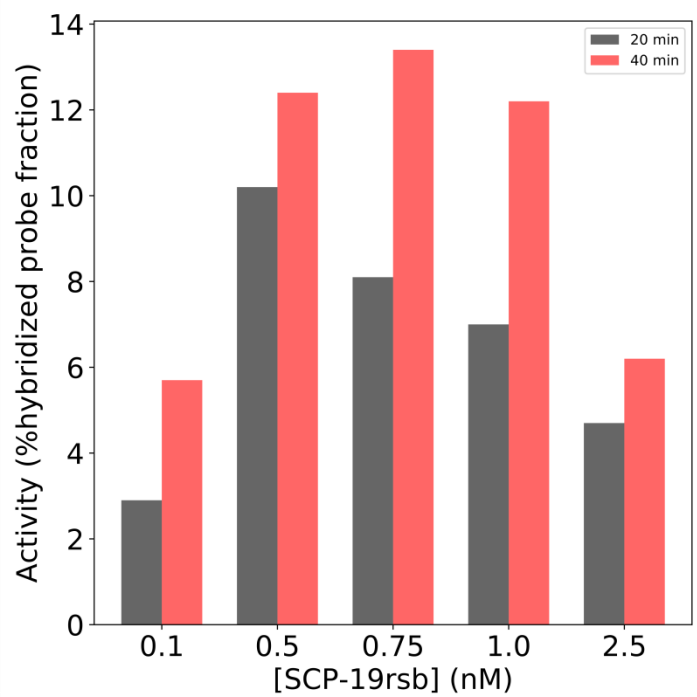



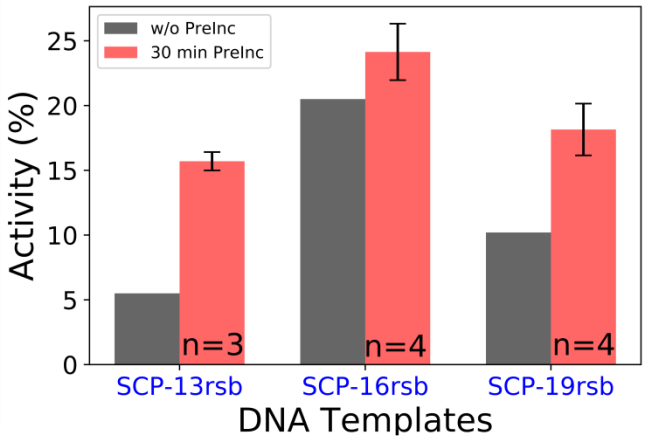
**Figure 9.** Heteroduplex templates with super core promoter used for smFRET-ALEX activity assay. The SC promoter contained the following consensus elements: TATA box, the upstream and downstream TFIIB Recognition Element (BREu and BREd respectively), Initiator Element (Inr), Motif Ten Element (MTE), and Downstream Promoter Element (DPE). The 3 heteroduplex templates (SCP-13rsb, SCP-16rsb, SCP-19rsb) differed in the size of the internal sequence mismatch (13, 16, and 19 nts). The TSS is indicated by the dashed box. Transcription reactions with each template done with the minimal Pol II system gave activity (Increase in lower FRET peak).

**Activity (%)=TE%(+NTP) - TE%(no NTP control)**

A




B




C

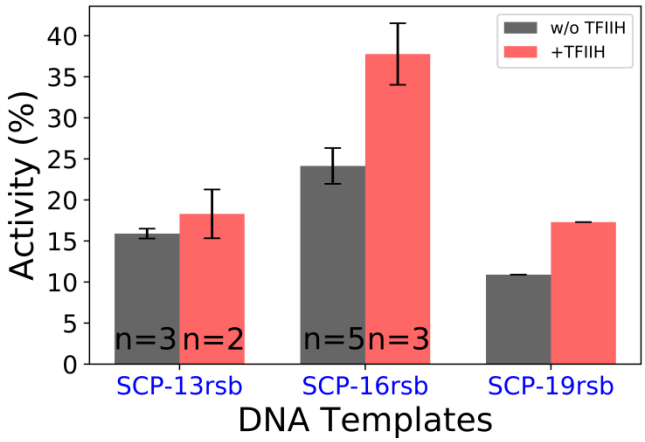


**Figure 10.** Optimizing the activity of the minimal Pol II system; TFIIH enhances promoter escape. The transcription activity is the relative increase in the low FRET population in the presence of NTPs (Top: equation for calculating relative activity). **(A)** Different DNA template (SCP-19rsb) concentrations were used to optimize assay conditions. **(B)** 30 min incubation of Pol II in buffer before the addition of the remaining factors to assemble minimal system PICs (DNA, TBP, IIB, IIF, IIE) enhances the transcription activity for the 3 heteroduplex templates (all at 0.5 nM final concentration) in comparison to standard protocol of simultaneous mixing of factors. **(C)** TFIIH enhanced transcription activity for all 3 heteroduplex DNAs. All reactions included a 30 min Pol II pre-incubation. Error bars are shown for measurements done more than once (n is number of replicates).

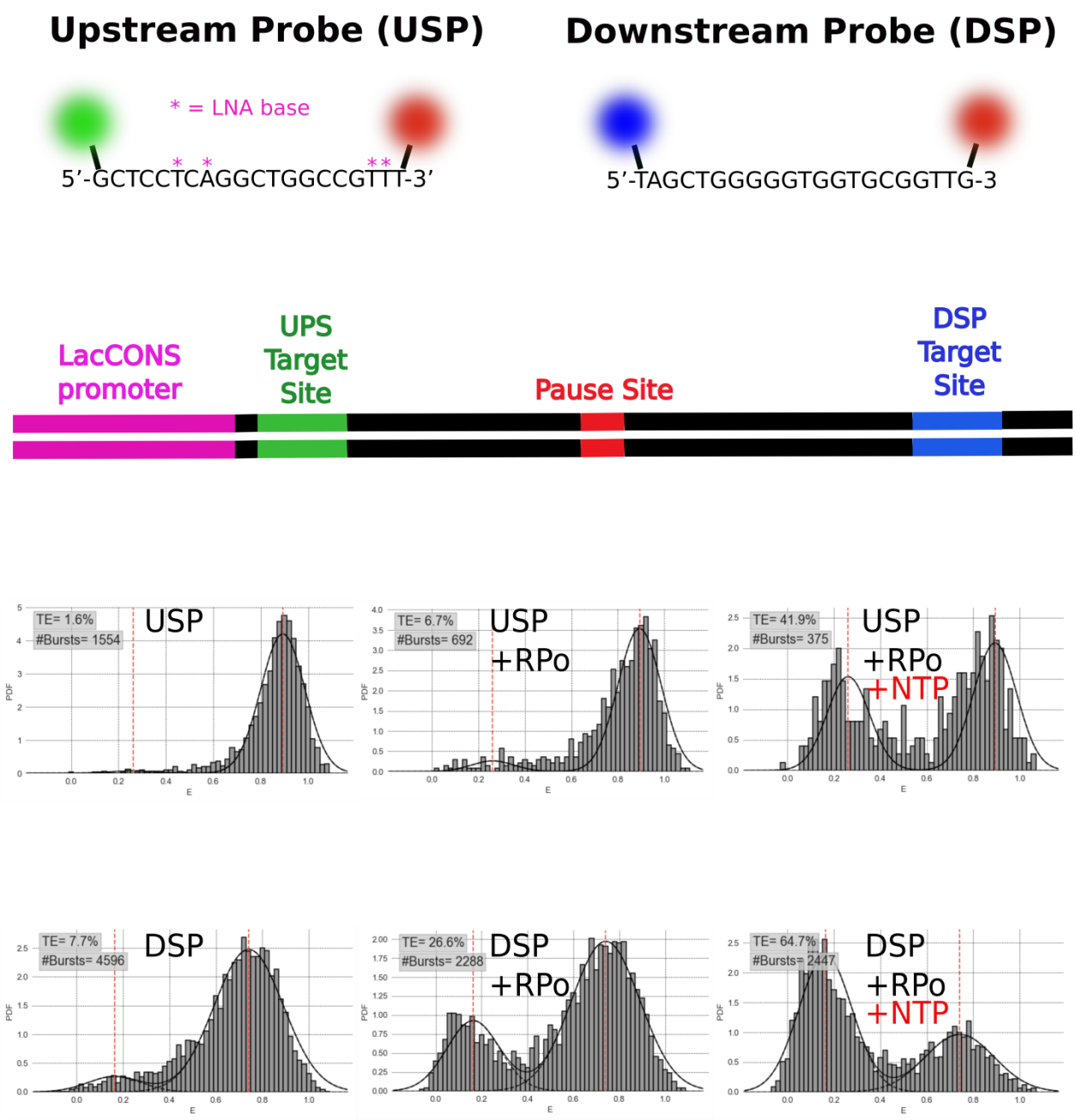



**Figure 11.** Development of smFRET-based Pol II pausing/pause release assay. **Top:** Upstream probe sequence containing 4 locked nucleic acids (LNA) labeled with Atto550 donor and Atto647N acceptor; downstream probe sequence labeled with Atto488 donor and Atto647N acceptor. **Middle:** Schematic for DNA template used in transcription reactions with *E. coli* RNAP to generate RNA for proof-of-principle tests of pausing assay probes. **Bottom:** smFRET-

ALEX data for probe alone (left panels), probe in the presence of RNAP open complex before (middle panels) and after NTP addition (right panels).

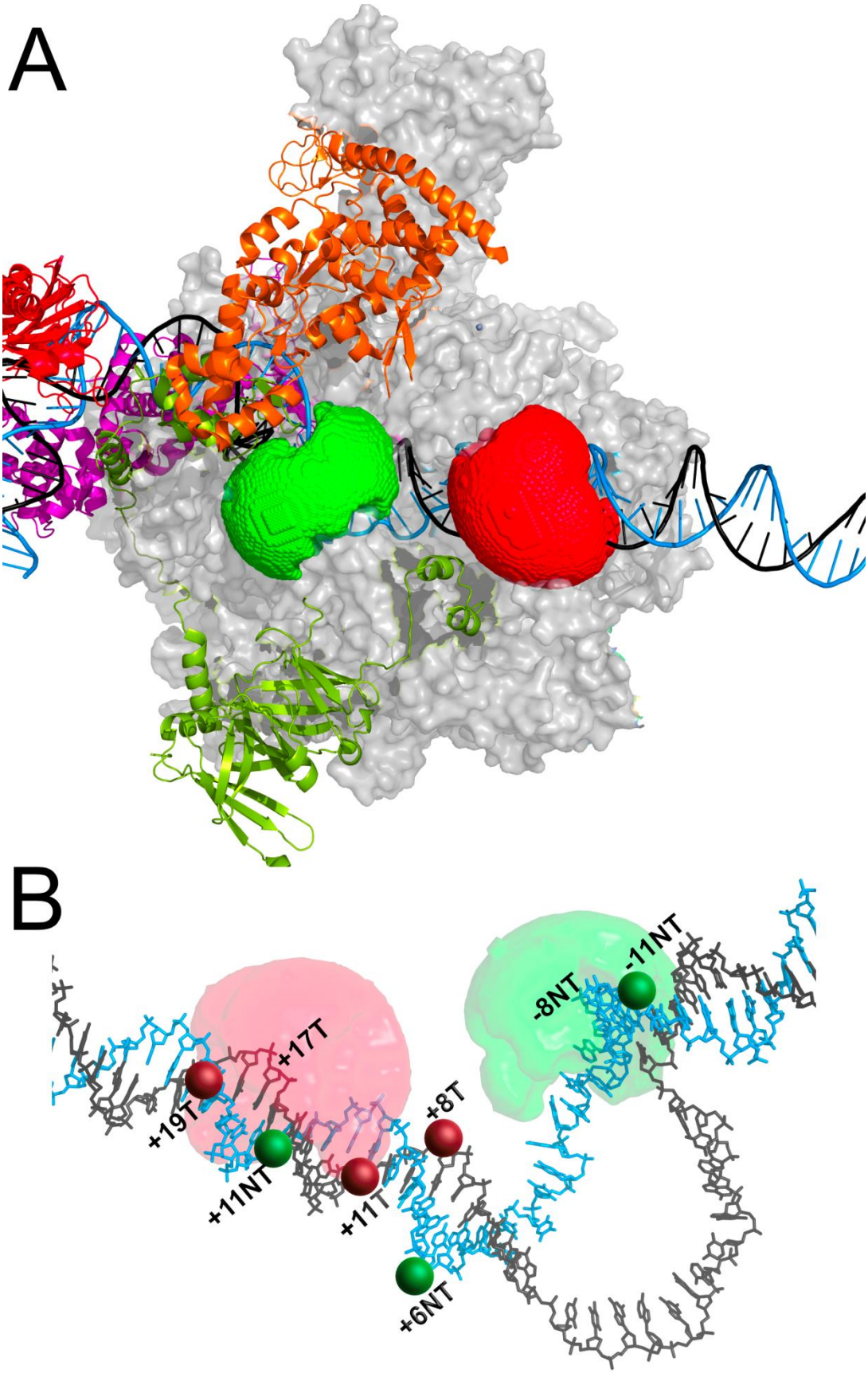

A
B
+17T
+19T
+11NT
+11T
+8T
+6NT
-8NT
-11NT

**Figure 12.** Studying the structural dynamics of the human Pol II transcription bubble in initiation. (A) Human PIC structure showing the accessible volumes (AVs) of the donor (Cy3B, green, position -8 on NT) and acceptor (Atto647N, red, position +17 on T) dyes on the NT (marine blue) and T (black) strands of the promoter DNA. PIC components: Pol-II (grey), TBP (red), TFIIB (magenta), TFIIF (light green), and TFIIE (orange). (B) Dye AVs for donor (-8 NT) and acceptor (+17 T) shown as in (A), while some additional candidate labeling positions for either the donor (green spheres) or acceptor (red spheres) on the NT and T, respectively. AVs were generated for donor and acceptor dyes at each position on the T and NT strands, both in the context of PIC and for free DNA, using FRET-restrained positioning and screening software analysis (*96*). Candidate positions for labeling promoter DNA with FRET pairs of dyes, within the PIC footprint, were selected based on minimal spatial restriction of the PIC structure on the dye AVs. Figure generated using PyMOL ((*26*), PDB 5IY7).

## References


1. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*. **9**, 85–98 (2011).
2. M. Hantsche, P. Cramer, Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.* (2017), , doi:10.1016/j.sbi.2017.03.013.
3. R. G. Keene, D. S. Luse, Initially transcribed sequences strongly affect the extent of abortive initiation by RNA polymerase II. *J. Biol. Chem.* (1999), doi:10.1074/jbc.274.17.11526.
4. T. R. Kadesch, M. J. Chamberlin, Studies of in vitro transcription by calf thymus RNA polymerase II using a novel duplex DNA template. *J. Biol. Chem.* (1982), doi:10.1002/mabi.201100029.
5. M. C. Thomas, C.-M. Chiang, The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–78 (2006).
6. S. Bagby *et al.*, TFIIA-TAF regulatory interplay: NMR evidence for overlapping binding sites on TBP. *FEBS Lett.* (2000), doi:10.1016/S0014-5793(00)01213-8.
7. J. H. Geiger, S. Hahn, S. Lee, P. B. Sigler, Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science (80-. ).* (1996), doi:10.1126/science.272.5263.830.
8. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129 (2015).
9. T. Juven-Gershon, J. T. Kadonaga, Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* (2010), , doi:10.1016/j.ydbio.2009.08.009.
10. Y. Wang, S. G. E. Roberts, New insights into the role of TFIIB in transcription initiation.

*Transcription* (2010), doi:10.4161/trns.1.3.12900.

11. G. Pan, J. Greenblatt, Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J. Biol. Chem.* **269**, 30101–30104 (1994).
12. H. T. Timmers, Transcription initiation by RNA polymerase II does not require hydrolysis of the beta-gamma phosphoanhydride bond of ATP. *EMBO J.* (1994).
13. E. Guzmán, J. T. Lis, Transcription factor TFIIH is required for promoter melting in vivo. *Mol. Cell. Biol.* (1999), doi:10.1128/MCB.19.8.5652.
14. J. Fishburn, E. Tomko, E. Galburt, S. Hahn, Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc. Natl. Acad. Sci.* (2015), doi:10.1073/pnas.1417709112.
15. D. S. Luse, G. A. Jacob, Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *J. Biol. Chem.* (1987).
16. J. A. Goodrich, R. Tjian, Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell*. **77**, 145–156 (1994).
17. A. R. Hieb, S. Baran, J. a Goodrich, J. F. Kugel, An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation. *EMBO J.* **25**, 3100–9 (2006).
18. C. Bernecky, J. M. Plitzko, P. Cramer, Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* (2017), doi:10.1038/nsmb.3465.
19. N. Yudkovsky, J. A. Ranish, S. Hahn, A transcription reinitiation intermediate that is stabilized by activator. *Nature* (2000), doi:10.1038/35041603.
20. S. Hahn, Structure and mechanism of the RNA polymerase II transcription machinery.

*Nat. Struct. Mol. Biol.* (2004), , doi:10.1038/nsmb763.

21. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–31 (2012).

22. H. Kwak, J. T. Lis, Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508 (2013).

23. N. J. Proudfoot, Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (80-. ).* (2016), , doi:10.1126/science.aad9926.

24. S. Grünberg, S. Hahn, Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* (2013), , doi:10.1016/j.tibs.2013.09.002.

25. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature*. **495**, 481–6 (2013).

26. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature* (2016), doi:10.1038/nature17970.

27. E. Nogales, R. K. Louder, Y. He, Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu. Rev. Biophys.* (2017), doi:10.1146/annurev-biophys-070816-033751.

28. X. Liu, L. Farnung, C. Wigge, P. Cramer, Cryo-EM structure of a mammalian RNA polymerase II elongation complex inhibited by α-amanitin. *J. Biol. Chem.* (2018), doi:10.1074/jbc.RA118.002545.

29. S. M. Vos *et al.*, Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*. **560**, 607–612 (2018).

30. S. M. Vos, L. Farnung, H. Urlaub, P. Cramer, Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* (2018), doi:10.1038/s41586-018-0442-2.

31. S. Weiss, Fluorescence spectroscopy of single biomolecules. *Science (80-. ).* (1999), , doi:10.1126/science.283.5408.1676.

32. S. Shashkova, M. C. Leake, Single-molecule fluorescence microscopy review: shedding new light on old problems. *Biosci. Rep.* (2017), doi:10.1042/BSR20170031.

33. T. Lionnet *et al.*, Single-molecule studies using magnetic traps. *Cold Spring Harb. Protoc.* (2012), doi:10.1101/pdb.top067488.

34. M. T. Woodside, S. M. Block, Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annu. Rev. Biophys.* (2014), doi:10.1146/annurev-biophys-051013-022754.

35. A. A. Deniz, S. Mukhopadhyay, E. A. Lemke, Single-molecule biophysics: At the interface of biology, physics and chemistry. *J. R. Soc. Interface* (2008), , doi:10.1098/rsif.2007.1021.

36. C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, T. Ha, Advances in Single-Molecule Fluorescence Methods for Molecular Biology. *Annu. Rev. Biochem.* (2008), doi:10.1146/annurev.biochem.77.070606.101543.

37. S. L. Reck-Peterson, N. D. Derr, N. Stuurman, Imaging single molecules using total internal reflection fluorescence microscopy (TIRFM). *Cold Spring Harb. Protoc.* (2010), doi:10.1101/pdb.top73.

38. J. Hohlbein, T. D. Craggs, T. Cordes, Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* (2013), doi:10.1039/c3cs60233h.

39. A. N. Kapanidis *et al.*, Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8936–8941 (2004).

40. A. Kapanidis, T. Laurence, Alternating-laser excitation of single molecules. *Accounts Chem. …*. **38**, 523–533 (2005).

41. N. K. Lee *et al.*, Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* **88**, 2939–2953 (2005).

42. E. A. Galburt *et al.*, Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* (2007), doi:10.1038/nature05701.

43. A. Muschielok *et al.*, A nano-positioning system for macromolecular structural analysis. *Nat. Methods* (2008), doi:10.1038/nmeth.1259.

44. M. Palangat *et al.*, Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II. *Transcription* (2012), doi:10.4161/trns.20269.

45. B. Treutlein *et al.*, Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol. Cell*. **46**, 136–46 (2012).

46. A. Gietl *et al.*, Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res.* (2014), doi:10.1093/nar/gku273.

47. V. Schweikhard *et al.*, Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6642–7 (2014).

48. T. Ishibashi *et al.*, Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics. *Proc. Natl. Acad. Sci.* (2014), doi:10.1073/pnas.1401611111.

49. F. M. Fazal, C. A. Meng, K. Murakami, R. D. Kornberg, S. M. Block, Real-time observation of the initiation of RNA polymerase II transcription. *Nature* (2015),

doi:10.1038/nature14882.

50. V. Fitz *et al.*, Nucleosomal arrangement affects single-molecule transcription dynamics. *Proc. Natl. Acad. Sci.* (2016), doi:10.1073/pnas.1602764113.

51. A. Lisica *et al.*, Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc. Natl. Acad. Sci.* (2016), doi:10.1073/pnas.1517011113.

52. J. B. Crickard, J. Lee, T. H. Lee, J. C. Reese, The elongation factor Spt4/5 regulates RNA polymerase II transcription through the nucleosome. *Nucleic Acids Res.* (2017), doi:10.1093/nar/gkx220.

53. J. Andrecka *et al.*, Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci.* (2008), doi:10.1073/pnas.0703815105.

54. C.-Y. Chen, C.-C. Chang, C.-F. Yen, M. T-K Chiu, W.-H. Chang, Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl. Acad. Sci.* (2009), doi:10.1073/pnas.0811689106.

55. J. Andrecka *et al.*, Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase ii elongation complex. *Nucleic Acids Res.* (2009), doi:10.1093/nar/gkp601.

56. C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, C. Bustamante, Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (80-. ).* (2009), doi:10.1126/science.1172926.

57. L. Bintu *et al.*, The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat. Struct. Mol. Biol.* (2011), doi:10.1038/nsmb.2164.

58. B. Zamft, L. Bintu, T. Ishibashi, C. Bustamante, Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc. Natl. Acad. Sci.* (2012),

doi:10.1073/pnas.1205063109.

59. L. Bintu *et al.*, Nucleosomal elements that control the topography of the barrier to transcription. *Cell* (2012), doi:10.1016/j.cell.2012.10.009.

60. M. H. Larson *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci.* (2012), doi:10.1073/pnas.1200939109.

61. P. Schluesche, G. Stelzer, E. Piaia, D. C. Lamb, M. Meisterernst, NC2 mobilizes TBP on core promoter TATA boxes. *Nat. Struct. Mol. Biol.* (2007), doi:10.1038/nsmb1328.

62. R. H. Blair, J. A. Goodrich, J. F. Kugel, Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics. *Biochemistry* (2012), doi:10.1021/bi300491j.

63. A. Revyakin *et al.*, Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev.* (2012), doi:10.1101/gad.194936.112.

64. A. E. Horn, J. F. Kugel, J. A. Goodrich, Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity. *Nucleic Acids Res.* (2016), doi:10.1093/nar/gkw321.

65. S. Kim *et al.*, High-throughput single-molecule optofluidic analysis. *Nat. Methods*. **8**, 242–245 (2011).

66. Z. C. Poss, C. C. Ebmeier, D. J. Taatjes, The Mediator complex and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* (2013), , doi:10.3109/10409238.2013.840259.

67. Z. Zhang, a. Revyakin, J. B. Grimm, L. D. Lavis, R. Tjian, Single-molecule tracking of the transcription cycle by sub-second RNA detection. *Elife*. **3**, e01775–e01775 (2014).

68. Y. Gao, L. K. Wolf, R. M. Georgiadis, Secondary structure effects on DNA hybridization

kinetics: A solution versus surface comparison. *Nucleic Acids Res.* (2006), doi:10.1093/nar/gkl422.

69. T. Špringer, H. Šípová, H. Vaisocherová, J. Štepánek, J. Homola, Shielding effect of monovalent and divalent cations on solid-phase DNA hybridization: Surface plasmon resonance biosensor study. *Nucleic Acids Res.* (2010), doi:10.1093/nar/gkq577.

70. R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, J. A. Wälder, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* (2008), doi:10.1021/bi702363u.

71. A. N. Kapanidis *et al.*, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*. **314**, 1144–7 (2006).

72. T. Cordes *et al.*, Sensing DNA opening in transcription using quenchable Förster resonance energy transfer. *Biochemistry*. **49**, 9171–80 (2010).

73. N. C. Robb *et al.*, 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**, 875–85 (2013).

74. D. Duchi *et al.*, RNA Polymerase Pausing during Initial Transcription. *Mol. Cell* (2016), doi:10.1016/j.molcel.2016.08.011.

75. A. Ingargiola *et al.*, Multispot single-molecule FRET: Highthroughput analysis of freely diffusing molecules. *PLoS One* (2017), doi:10.1371/journal.pone.0175766.

76. E. Lerner, A. Ingargiola, S. Weiss, Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example. *J. Chem. Phys.* (2018), doi:10.1063/1.5004606.

77. D. Duchi *et al.*, Conformational heterogeneity and bubble dynamics in single bacterial

transcription initiation complexes. *Nucleic Acids Res.* (2018), doi:10.1093/nar/gkx1146.

78. J. D. Dignani, R. M. Lebovitz, R. G. Roeder, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* (1983), doi:10.1093/nar/11.5.1475.

79. T. Juven-Gershon, S. Cheng, J. T. Kadonaga, Rational design of a super core promoter that enhances gene expression. *Nat. Methods*. **3**, 917–922 (2006).

80. J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, Identifying molecular dynamics in single-molecule fret experiments with burst variance analysis. *Biophys. J.* (2011), doi:10.1016/j.bpj.2011.01.066.

81. E. Lerner *et al.*, Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *Proc. Natl. Acad. Sci. United States Am.* (2016), doi:10.1073/pnas.1605038113.

82. R. J. Moreland *et al.*, A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274**, 22127–22130 (1999).

83. T. Henriques *et al.*, Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* (2013), doi:10.1016/j.molcel.2013.10.001.

84. K. L. Zobeck, M. S. Buckley, W. R. Zipfel, J. T. Lis, Recruitment Timing and Dynamics of Transcription Factors at the Hsp70 Loci in Living Cells. *Mol. Cell* (2010), doi:10.1016/j.molcel.2010.11.022.

85. C. H. Wu *et al.*, NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* (2003), doi:10.1101/gad.1091403.

86. K. Adelman *et al.*, Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell*. **17**, 103–112 (2005).

87. A. E. Horn, J. A. Goodrich, J. F. Kugel, Single molecule studies of RNA polymerase II transcription in vitro. *Transcription* (2014), , doi:10.4161/trns.27608.

88. J. B. Crickard, J. Fu, J. C. Reese, Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *J. Biol. Chem.* (2016), doi:10.1074/jbc.M116.716001.

89. D. V. Titov *et al.*, XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* (2011), doi:10.1038/nchembio.522.

90. N. Kwiatkowski *et al.*, Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* (2014), doi:10.1038/nature13393.

91. M. Pal, A. S. Ponticelli, D. S. Luse, The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* (2005), doi:10.1016/j.molcel.2005.05.024.

92. M. T. Knuesel, K. D. Meyer, C. Bernecky, D. J. Taatjes, The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439–51 (2009).

93. A. Ingargiola, E. Lerner, S. Y. Chung, S. Weiss, X. Michalet, FRETBursts: An open source toolkit for analysis of freely-diffusing Single-molecule FRET. *PLoS One* (2016), doi:10.1371/journal.pone.0160716.

94. N. R. Markham, M. Zuker, DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* (2005), doi:10.1093/nar/gki591.

95. N. R. Markham, M. Zuker, UNAFold: Software for nucleic acid folding and hybridization. *Methods Mol. Biol.* (2008), doi:10.1007/978-1-60327-429-6_1.

96. S. Kalinin *et al.*, A toolkit and benchmark study for FRET-restrained high-precision

structural modeling. *Nat. Methods* (2012), doi:10.1038/nmeth.2222.

# Chapter 4:

Investigating the effect of different 5'-RNA moieties on *E. coli* RNA polymerase transcription initiation using single-molecule FRET

## Introduction

Transcription of the bacterial genome is implemented by a multi-subunit RNA polymerase (RNAP) enzyme. RNA synthesis by RNAP is achieved via the nucleotide addition cycle, in which each iteration incorporates an incoming ribonucleoside triphosphate (NTP) (*1*). Incoming NTPs are selected based upon sequence complementarity to the DNA template. The 370 kDa E. *coli* RNAP core enzyme is composed of five subunits ($\alpha_I$, $\alpha_{II}$, β', β, ω). The core RNAP is catalytically competent for transcription elongation; however, sequence-specific recognition of promoter DNA requires association with a sigma (σ) factor to form a holoenzyme, holoRNAP (RPσ) (*2*). Bacteria typically have one major σ factor (σ70 in *E. coli*) that drives expression of 'housekeeping' genes and varying numbers of alternative σ factors that direct expression of genes in response to specific environmental conditions or stress (*2*).

Transcription can be divided into three major stages: initiation, elongation, and termination. Initiation is the most regulated step in transcription; it begins with RPσ binding to promoter DNA to form the RPσ -closed promoter complex ($RP_C$), in which the DNA situated above the cleft remains double stranded. The $RP_C$ undergoes a series of conformational changes that unwinds ~13 bp DNA around the transcription start site, which transforms into a stable RPσ-open promoter complex ($RP_O$) following additional structural adjustments—mainly in downstream DNA(*3*). Two initiating NTPs can then bind to complementary bases in the DNA template strand in the $RP_O$ active site. Phosphodiester bond formation between the initial two NTPs leads to a transition from $RP_O$ to the initial transcribing complex ($RP_{ITC}$). When the RNA reaches a length of ~5 nucleotides (nts), the 5'-end of the RNA clashes with the σ3.2 loop, which blocks the entrance to the RNA exit channel (*4*). The strain induced by the clash increases as more NTPs are incorporated into the RNA until either the RNAP backtracks to generate a paused

initiation complex (*5*, *6*) or the σ70 regions obstructing the RNA exit channel are removed—allowing the RNA to enter the channel and the process of promoter escape to begin. One factor governing the outcome of the interplay between the σ3.2 loop and RNA is the moiety at the 5'-end of the RNA. Canonical 5'-moietyies of the ribose carbon five include the unphosphorylated hydroxyl or a mono-, di-, or triphosphate group (*7*). Some non-canonical moieties were shown to be incorporated during transcription initiation (*8*). We previously demonstrated that whether using dinucleotide primers containing a hydroxyl or triphosphate at the 5'-end to initiation transcription, RNAP pausing occurs in initiation (*5*). It was even shown that a triphosphate 5'-end enhances escape from the paused state in initiation (*9*). Eventually RNAP escapes the promoter to form the highly processive ternary RNAP-DNA-RNA elongation complex ($RP_E$).

In bacteria, cyclic dinucleotides (CDN) are a ubiquitous class of signaling molecules that regulate many biological processes (*10*). Bis-(3'-5')-cyclic diguanylic acid (c-di-GMP) and 3',5'-cyclic di-adenosine monophosphate (c-di-AMP) are the most well characterized CDNs in bacteria and are synthesized by cyclases from GTP and ATP, respectively (*10*). Hydrolysis of c-di-GMP and c-di-AMP into their nucleoside monophosphates—GMP and AMP, respectively—is achieved by several different phosphodiesterases (PDE) (reviewed in (*10*)) in bacteria with some PDEs only partially hydrolyzing CDNs to form linear dinucleotides with a 5' monophosphate (pNpN), which are 5′-phosphoadenylyl-(3′-5′)-guanosine (pGpG) and 5′-phosphoadenylyl-(3′-5′)-adenosine (pApA), respectively (*10*, *11*). Increased levels of pGpG was shown to stimulate biofilm formation in bacteria (*10*), while pApA has been demonstrated to inhibit the activity of the PDE hydrolyzing c-di-AMP (*12*).

In fluorescence microscopy and spectroscopy, FRET (Förster Resonance Energy Transfer) is a valuable tool for distance-dependent measurements (*13*, *14*). FRET is the non-

radiative transfer of energy from a fluorophore excited by a higher energy photon (donor) to an adjacent fluorophore (acceptor). A major advance for smFRET (single-molecule FRET) technology was the development of alternating laser excitation technology (ALEX), in which the excitation source is continuously alternated between donor and acceptor excitation wavelengths using an acoustic optical modulator (*15*). The smFRET-ALEX technology enables the virtual separation of subpopulations based on the stoichiometry of donor/acceptor fluorophore labeling (donor-only labeled, acceptor-only labeled, donor- and acceptor-labeled) (*16*).

Using smFRET-ALEX, we investigated the effect of the RNA 5'-end on transcription initiation. We demonstrate an apparent destabilization of $RP_O$ induced by pApA and pGpG binding, which bind to complementary bases in the +1/+2 and -1/-2 sites with respect to the canonical TSS. We also demonstrate an enhancement in the entrance to the paused state for initiation RNAs containing a 5'-triphosphate.

## Results and Discussion

### *Entrance kinetics into the paused-backtracked state in initiation is enhanced by RNA 5'-triphosphate*

We previously characterized the kinetics of RNAP that paused and backtracked in initiation (*5*). We also demonstrated that the kinetics of promoter escape starting from a later initiation stage—($RP_{ITC7}$) generated by adding a 5'-triphosphate (pppApA) or a 5'-hydroxyl (ApA) dinucleotide primer along with a subset of NTPs—was delayed in comparison to starting from $RP_{ITC2}$ (*5*). To determine what effect the 5'-triphosphate has on the rate of entering into the paused-backtracked state, the entrance kinetics (*5*) was determined for $RP_{ITC2}$ complexes made by incubating either pppApA or ApA with $RP_O$ (RPσ + LacCONS_20dA DNA template, Figure

1: top panel) with. The entrance kinetics was performed by incubating $RP_{ITC2}$ (pppApA) or $RP_{ITC2}$ (ApA) with a subset of NTPs—that permits a maximum RNA length of 7 nts (described in (*5*))—for different lengths of time before adding the missing NTPs. Reactions were quenched after 15 minutes and the 20dT FRET probe was added to measure RNA production (*5*). The entrance kinetics for $RP_{ITC2}$ containing pppApA was significantly faster than similar $RP_{ITC2}$ containing ApA (Figure 2).

This result provides interesting insight into the effect of the negatively charged phosphates at the 5'-end of the RNA when clashing with the negatively charged σ3.2 loop. Although the steric clash between σ3.2 loop and the 5-hydroxyl of RNAs initiated from ApA is sufficient to induce RNA backtracking and RNAP pausing; the additional electrostatic repulsion between the negative phosphates and the conserved aspartic residues of the σ3.2 loop (*17*) significantly enhances RNAP pausing.

***Reduction in fraction of open complexes in the presence of pApA and pGpG***

While performing experiments on the effect of the 5'-end moiety on the propensity for entering the backtracked state in initiation, $RP_{ITC2}$ complexes with pApA were exhibiting a gradual loss in activity over time (data not shown), which was not observed for $RP_{ITC2}$ complexes with pppApA or ApA. To investigate the possible cause, we employed a previously used smFRET method reporting on the conformation of DNA in the region where the DNA strands are separated in $RP_O$ (*5*, *18, 19*). Two template DNAs with alternative labeling configurations were used (Figure 1, middle and bottom panel), for the -5NTD-8TA_DNA template the nontemplate(NT) -5 and template (T) -8 position relative to TSS are labeled donor and acceptor, respectively; while for the -8NTD-3TA_DNA template the NT -8 and T -3

positions relative to TSS were labeled(Figure 1). These templates were previously used to report on the transcription bubble opening (*18*, *19*). In the double stranded (ds) form of the labeled DNA, the relatively close proximity of the FRET dyes results in a population with a higher FRET efficiency (reported here as an uncorrected FRET efficiency or proximity ratio (E*)) from the donor to acceptor fluorophores (Figure 1, left 1D-FRET histogram for middle and bottom panel). Upon formation of $RP_O$ and separation of the DNA in the promoter region results in a population with a lower FRET efficiency (Figure 1, right 1D-FRET histogram for middle and bottom panel).

Comparing the smFRET-ALEX results of $RP_O$ with -8NTD-3TA_DNA template after incubating for 7.5 minutes with either 2.5 mM GpG (5'-OH) or 3 mM pGpG reveals a relative decrease in the lower FRET ($RP_O$) population, which is more pronounced at 37 °C than room temperature, for pGpG (Figure 3, compare C vs D and E vs F). No apparent effect was observed for GpG (Figure 3, C and D vs B).

Similar experiments using $RP_O$ with -5NTD-8TA_DNA revealed the same effect for 3 mM pGpG and 2 mM pApA in decreasing the lower FRET $RP_O$ population to a greater extent 37 °C versus room temperature (Figure 4; compare B to E/ F and G/H). To ensure the FRET dyes in the transcription bubble are not interfering with RNAP activity, the effect of NTP addition was evaluated for $RP_O$ with -5NTD-8TA_DNA (Figure 4D) and -8NTD-3TA_DNA (not shown). For both templates, addition of NTP resulted in a decrease in the lower FRET fraction. To determine if the apparent effect is allosteric or through active site binding, we tested the effect of pApG on $RP_O$ stability. Unlike, pApA and pGpG, pApG is not complementary to any sequence in the transcription bubble. No significant decrease in the $RP_O$ population was observed for pApG (Figure 4C).

The effect of pApA on the fraction of $RP_O$ was also monitored over time at room temperature and 37 °C (Figure 5A). The fraction of $RP_O$ at each time following pApA addition was normalized to the $RP_O$ fraction without pApA. The fraction of $RP_O$ decrease by ~30% over 30 minutes following pApA addition (Figure 5A). Fluorescence correlation spectroscopy (FCS), which provides a measure of the diffusion time for fluorescent molecules diffusing through the laser excitation volume, of $RP_O$ incubated pApA indicates dissociation of $RP_O$ after incubation with pApA (Figure 5B). Addition of pApA caused the FRET labeled DNA to diffuse as quickly as DNA alone control, instead of the diffusion profile of $RP_O$.

**Methods**

Entrance kinetics assay

The LacCONS_20dA T and NT strands were ordered from IDT and prepared as described in (*5*). The entrance kinetics was also performed as in (*5*).

smFRET-ALEX assay measuring fraction of $RP_O$ with labeled DNA

The NT and T strands for -5NTD-8TA_DNA and -8NTD-3TA were ordered (IBA life sciences) with PAGE and reverse HPLC purification. -5NTD-8TA_DNA and -8NTD-3TA were prepared by combining equimolar ratios of each strand in TMK buffer (10 mM Tris-pH 8.0, 100 mM KCl, 2 mM $MgCl_2$) and using a thermocycler programmed with the following parameter: (1) heating to 95 °C for 3 minutes; (2) for the temperature range $T_{Tm+5°C}$ to $T_{Tm-5°C}$ ($T_{Tm+5°C}$ is 5 °C greater than Tm, $T_{Tm-5°C}$ is 5 °C less than Tm), a 10 minute incubation at each temperature (1

°C intervals); (3) incubation at 3 minutes at each temperature (1 °C intervals) until reaching 45 °C; and finally (4) the thermocycler was cooled to 4 °C (All cooling steps were set at a rate of 0.1 °C/second).

Preparation of $RP_O$ is as described before in (*5*). Before testing the effect of different dinucleotides on $RP_O$ stability, a working stock for all reagents—1 mM NTPs (GE Healthcare, high purity 100 mM original stock), 10 mM for pApA/pGpG/GpG/pApG—was prepared in 1X transcription buffer [40 mM Hepes KOH, 50 mM KCl, 10 mM MgCl2, 1 mM dithiotreitol (DTT), 1 mM 2-mercaptoethylamine-HCl (MEA), 100 μg/mL BSA, pH 7]. In a total volume of 20 μLs, 100 pM of $RP_O$ (calculated based on concentration of DNA template) was incubated with different concentrations of pNpNs or NpNs in 1X transcription buffer for a given time at room temperature or 37 °C before measuring with smFRET-ALEX. For transcription reactions, NTPs were added to a final concentration of 100 μM in 20 μL total volume containing 100 pM of $RP_O$ before immediately incubating the reactions at 37 °C for 15 minutes. Reactions were then immediately measured with smFRET-ALEX. All smFRET-ALEX measurements and analysis were done as previously described (*5*), with the only difference being the threshold for performing the dual channel burst search was identical to what is described in the third chapter of this dissertation.

FCS measurements

All FCS measurements were performed at 25 °C using a home-built confocal microscope based on an Olympus IX71 with CW laser at 532nm (Melles Griot Inc., Carlsbad, CA). The emitted photon stream from the labeled complexes was split by a dichroic mirror (635LP,

Chroma, VT) into two different color channels (one for the fluorescence from Atto550, and the other one for the fluorescence from Alexa647) equipped with SPADs (Perkin Elmer Inc., Waltham, MA, USA). Detected fluorescence signals were sent to an ALV-6000 MultiCorr digital real-time correlator and then cross-correlated to eliminate the unwanted signal from singly labeled species and detector after-pulsing.

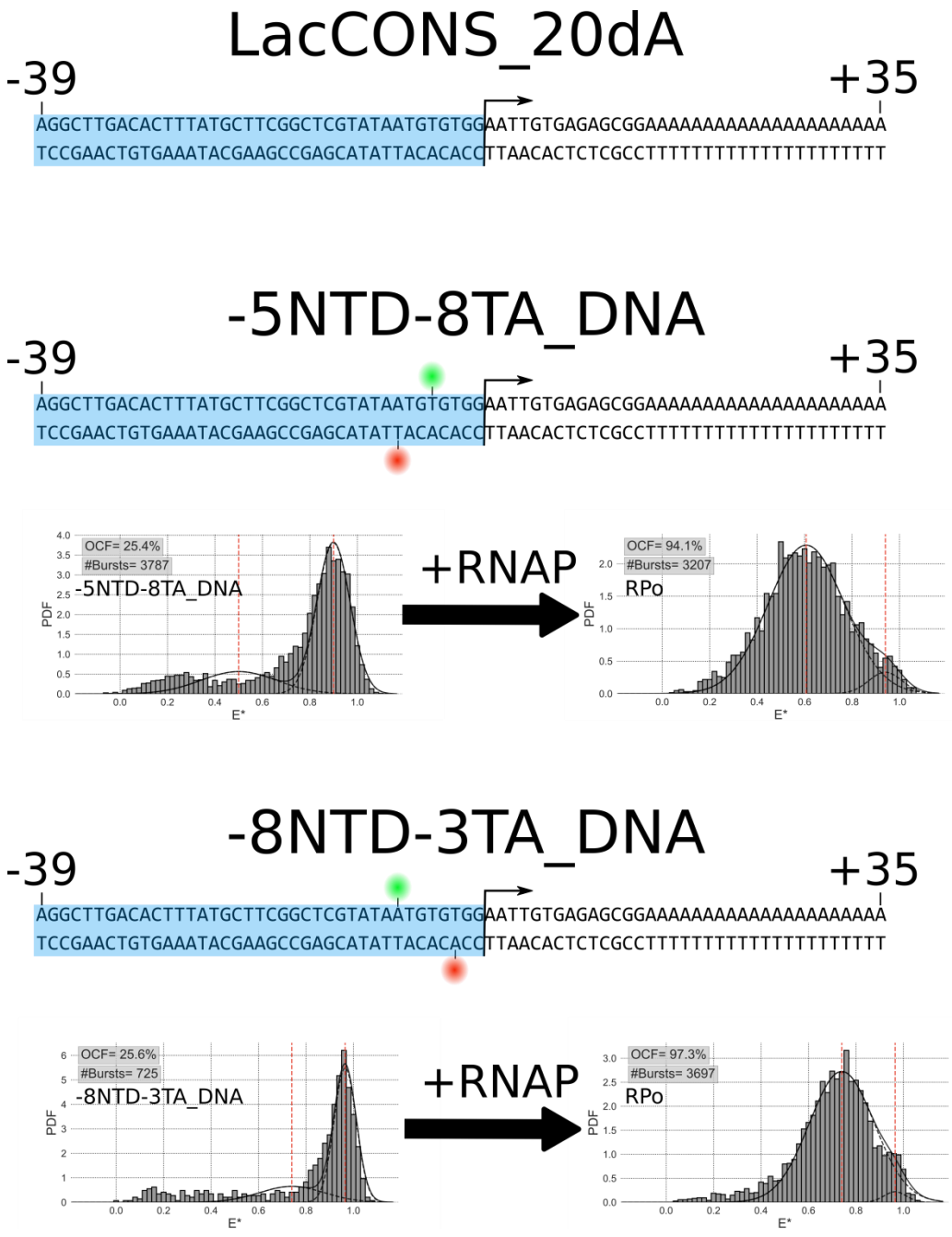



**Figure 1.** Unlabeled and Labeled DNA templates used for smFRET-ALEX assays. LacCONS promoter highlighted in blue. **Top:** linear DNA template used for investigating effect of 5'-RNA dinucleotide moiety on entrance kinetics into backtracked ITC7 state. **Middle:** -5NTD-8TA DNA template labeled at positions -5 (donor, Cy3B, NT) and -8 (acceptor, Atto647N, T)

relative to TSS. 1D FRET-histogram of DNA alone and after incubating with RNAP to form $RP_O$. **Bottom:** -8NTD-3TA DNA template labeled at positions -8(donor, Cy3B, NT) and -3 (acceptor, Atto647N, T) relative to TSS. 1D FRET-histogram of DNA alone and after incubating with RNAP to form $RP_O$.

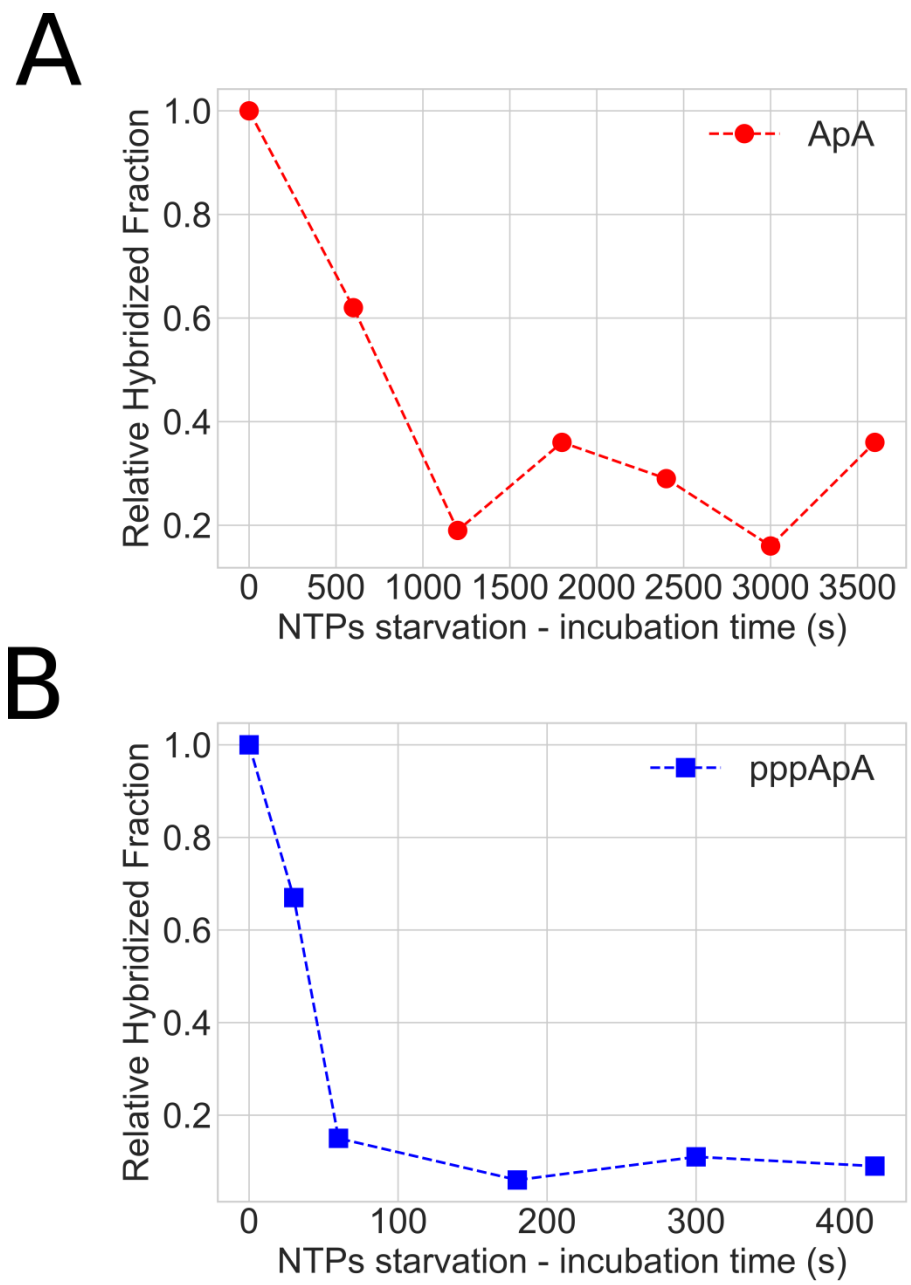



**Figure 2.** Entrance kinetics into the RNAP-paused backtracked state is enhanced for 5'-triphosphate RNAs relative to 5'-OH RNAs. $RP_{ITC2}$ (pppApA or ApA) was incubated with a subset of NTPs for different times before adding the missing NTP and measuring RNA after running transcription reactions for 15 minutes. The fraction of hybridized 20dT at each incubation time was normalized to the fraction of hybridized 20dT without any NTP starvation

step (5). **(A)** Entrance kinetics for $RP_{ITC2}$ with ApA. **(B)** Entrance kinetics for $RP_{ITC2}$ with pppApA.

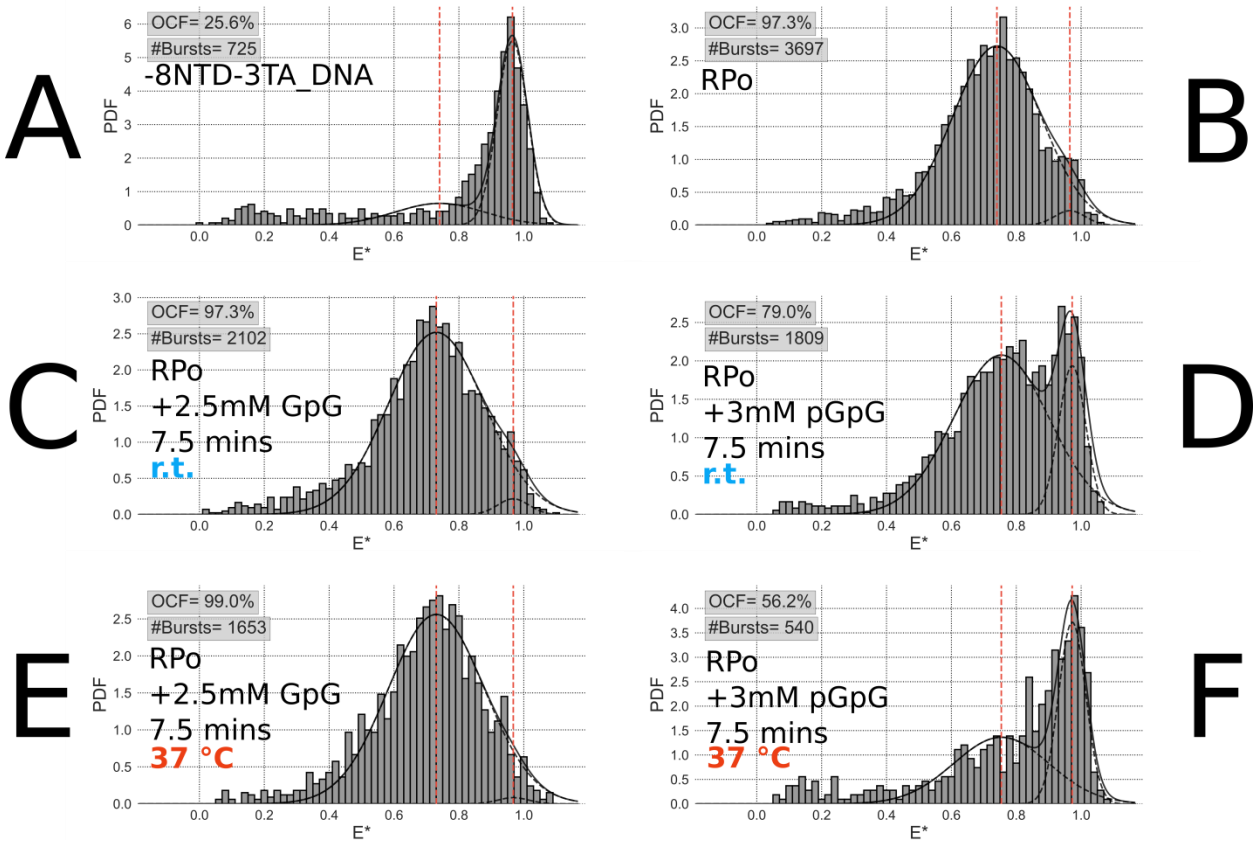



**Figure 3.** 1D-FRET histograms showing the effect of 5'-monophosphate vs 5'-hydroxyl linear dinucleotides on the population of open complexes (lower FRET peak) at different temperatures. All measurements include the -8NTD-3TA_DNA template labeled with Cy3B (NT, -8 relative to TSS) and Atto647N (T, -3 relative to TSS). **(A)** DNA alone, **(B)** $RP_O$. **(C-F)** smFRET-ALEX measurements done after incubating $RP_O$ for 7.5 minutes with GpG (5'-OH) or pGpG (5'-monophosphate) at room temp. (rt) or 37 °C.

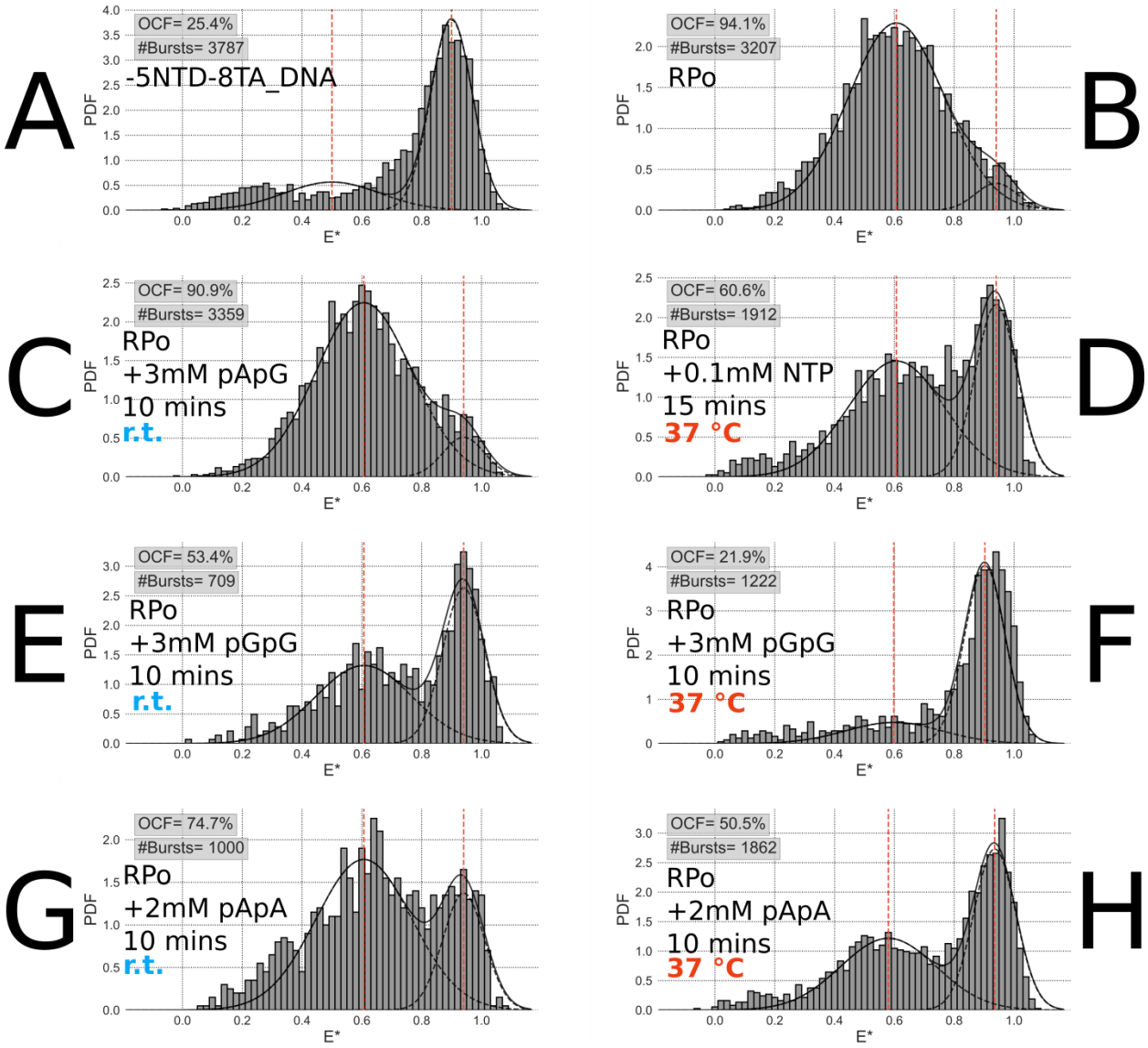



**Figure 4.** 1D-FRET histograms showing the effect of 5'-monophosphate linear dinucleotides on the population of open complexes (lower FRET peak) at different temperatures. All measurements include the -5NTD-8TA_DNA template labeled with Cy3B (NT, -5 relative to TSS) and Atto647N (T, -8 relative to TSS). **(A)** DNA alone, (**B)** $RP_O$. **(C-H)** smFRET-ALEX measurements done after incubating $RP_O$ with the non-complementary pNpN (pApG), NTP, and complementary pNpNs (pApA, pGpG). Incubation time and temperature indicated.

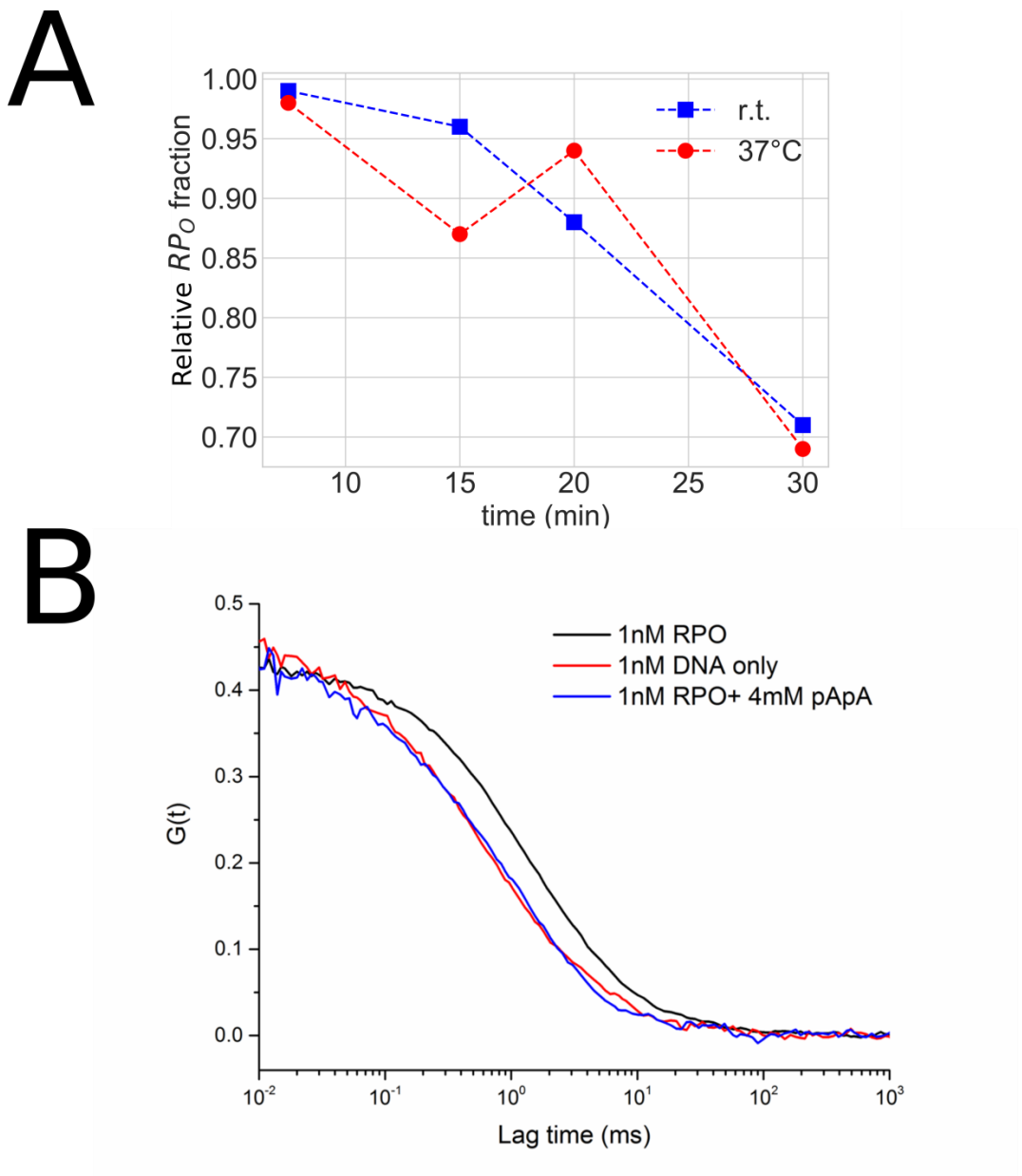



**Figure 5.** Fraction of open complexes (lower FRET pop. in Fig. 2,3) decreases over time in the presence of 4 mM pApA at different temperatures; pApA results in dissociation of the open complex. **(A)** Fraction of $RP_O$ (lower FRET) incubated at different times with 4 mM pApA before smFRET-ALEX measurement relative to $RP_O$ without any pApA. Incubations were done

at room temp. (r.t., blue dashed line, filled in squares) and 37 °C (red dashed line, filled in circles). Measurements done with the -8NTD-3TA_DNA template. **(B)** FCS curves of -5NTD-8TA_DNA (DNA only, red) only, $RP_O$ in the presence (blue) and absence (black) of 4mM pApA. FCS curves show that after 25 minutes incubation with 4 mM pApA, diffusion profile of $RP_O$ becomes similar to that of DNA only.

## References


1. E. Nudler, RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* **78**, 335–361 (2009).

2. A. Feklístov, B. D. Sharon, S. A. Darst, C. A. Gross, Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* (2014), doi:10.1146/annurev-micro-092412-155737.

3. E. Ruff, M. Record, I. Artsimovitch, Initial Events in Bacterial Transcription Initiation. *Biomolecules*. **5**, 1035–1062 (2015).

4. Y. Zuo, T. A. Steitz, Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* (2015), doi:10.1016/j.molcel.2015.03.010.

5. E. Lerner *et al.*, Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *Proc. Natl. Acad. Sci. United States Am.* (2016), doi:10.1073/pnas.1605038113.

6. D. Duchi *et al.*, RNA Polymerase Pausing during Initial Transcription. *Mol. Cell* (2016), doi:10.1016/j.molcel.2016.08.011.

7. B. E. Nickels, S. L. Dove, NanoRNAs: A class of small RNAs that can prime transcription initiation in bacteria. *J. Mol. Biol.* (2011), , doi:10.1016/j.jmb.2011.06.015.

8. J. G. Bird *et al.*, The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. *Nature* (2016), doi:10.1038/nature18622.

9. D. Dulin *et al.*, Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nat. Commun.* (2018), doi:10.1038/s41467-018-03902-9.

10. T. A. N. Huynh, J. J. Woodward, Too much of a good thing: Regulated depletion of c-di-AMP in the bacterial cytoplasm. *Curr. Opin. Microbiol.* (2016), , doi:10.1016/j.mib.2015.12.007.

11. U. Jenal, A. Reinders, C. Lori, Cyclic di-GMP: Second messenger extraordinaire. *Nat. Rev. Microbiol.* (2017), , doi:10.1038/nrmicro.2016.190.

12. L. Bowman, M. S. Zeden, C. F. Schuster, V. Kaever, A. Gründling, New insights into the cyclic di-adenosine monophosphate (c-di-AMP) degradation pathway and the requirement of the cyclic dinucleotide for acid stress resistance in Staphylococcus aureu. *J. Biol. Chem.* (2016), doi:10.1074/jbc.M116.747709.

13. J. Hohlbein, T. D. Craggs, T. Cordes, Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* (2013), doi:10.1039/c3cs60233h.

14. E. Lerner *et al.*, Toward dynamic structural biology: Two decades of single-molecule förster resonance energy transfer. *Science (80-. ).* (2018), , doi:10.1126/science.aan1133.

15. A. N. Kapanidis *et al.*, Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38**, 523–533 (2005).

16. A. N. Kapanidis *et al.*, Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8936–8941 (2004).

17. R. S. Basu *et al.*, Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* (2014), doi:10.1074/jbc.M114.584037.

18. A. Ingargiola *et al.*, Multispot single-molecule FRET: Highthroughput analysis of freely diffusing molecules. *PLoS One* (2017), doi:10.1371/journal.pone.0175766.

19. E. Lerner, A. Ingargiola, S. Weiss, Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example. *J. Chem. Phys.* (2018), doi:10.1063/1.5004606.

# Chapter 5:

Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase

# Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase

Eitan Lerner[a,1], SangYoon Chung[a,1], Benjamin L. Allen[b,1], Shuang Wang[c,1], Jookyung Lee[d], Shijia W. Lu[a], Logan W. Grimaud[a], Antonino Ingargiola[a], Xavier Michalet[a], Yazan Alhadid[a], Sergei Borukhov[d], Terence R. Strick[c,e,f,2], Dylan J. Taatjes[b,2], and Shimon Weiss[a,g,h,2]

[a]Department of Chemistry & Biochemistry, University of California, Los Angeles, CA 90095; [b]Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309; [c]Institut Jacques Monod, Centre National de la Recherche Scientifique (CNRS), UMR7592, University Paris Diderot, Sorbonne Paris Cité, F-75205 Paris, France; [d]Rowan University School of Osteopathic Medicine, Stratford, NJ 08084; [e]Institut de Biologie de l'Ecole Normale Supérieure, Institut de Biologie de l'Ecole Normale Superieure (IBENS), CNRS, Inserm, Ecole Normale Supérieure, Paris Sciences et Lettres (PSL) Research University, F-75005 Paris, France; [f]Programme Equipe Labellisées, Ligue Contre le Cancer, 75013 Paris, France; [g]Molecular Biology Institute, University of California, Los Angeles, CA 90095; and [h]Department of Physiology, University of California, Los Angeles, CA 90095

Edited by Steven M. Block, Stanford University, Stanford, CA, and approved September 13, 2016 (received for review March 30, 2016)

**Initiation is a highly regulated, rate-limiting step in transcription. We used a series of approaches to examine the kinetics of RNA polymerase (RNAP) transcription initiation in greater detail. Quenched kinetics assays, in combination with gel-based assays, showed that RNAP exit kinetics from complexes stalled at later stages of initiation (e.g., from a 7-base transcript) were markedly slower than from earlier stages (e.g., from a 2- or 4-base transcript). In addition, the RNAP–GreA endonuclease accelerated transcription kinetics from otherwise delayed initiation states. Further examination with magnetic tweezers transcription experiments showed that RNAP adopted a long-lived backtracked state during initiation and that the paused–backtracked initiation intermediate was populated abundantly at physiologically relevant nucleoside triphosphate (NTP) concentrations. The paused intermediate population was further increased when the NTP concentration was decreased and/or when an imbalance in NTP concentration was introduced (situations that mimic stress). Our results confirm the existence of a previously hypothesized paused and backtracked RNAP initiation intermediate and suggest it is biologically relevant; furthermore, such intermediates could be exploited for therapeutic purposes and may reflect a conserved state among paused, initiating eukaryotic RNA polymerase II enzymes.**

transcription | pausing | backtracking | RNAP | RNA polymerase

Transcription in *Escherichia coli* comprises three stages: initiation, elongation and termination. Initiation, which is typically the rate-limiting and the most regulated stage of transcription, is by itself a complex, multistep process consisting of the following successive steps (1, 2): (*i*) association of RNA polymerase (RNAP) core enzyme (subunit composition $\alpha 2\beta\beta'\omega$) with the promoter specificity factor σ (such as σ70 for transcription of housekeeping genes) to form RNAP holoenzyme; (*ii*) binding of holoenzyme to the −10 and −35 DNA elements in the promoter recognition sequence (PRS) upstream to the transcription start site (TSS) to form closed promoter complex (RPc); (*iii*) isomerization of RPc through multiple intermediates into an open promoter complex (RPo), in which a ∼12-bp DNA stretch (bases at registers −10 to +2) is melted to form a transcription bubble, the template DNA strand is inserted into RNAP major cleft positioning the base at register +1 of TSS at the active site, the nontemplate strand is tightly bound by σ70 and the downstream DNA duplex (bases +3 up to +20) is loaded into RNAP β′ DNA-binding clamp; (*iv*) an abortive initiation step (AI), where binding of nucleoside triphosphate (NTPs) and the start of RNA synthesis leads to formation of an initial transcribing complex ($RP_{ITC}$) followed by RNAP cycling through multiple polymerization trials via a DNA scrunching mechanism (3, 4), release of short "abortive transcripts" and repositioning itself in $RP_O$ for a new synthesis trial (5–7); and finally, (*v*) RNAP promoter escape, when enough strain is built in the enzyme, the σ70 undergoes structural transition to relieve blockage of the RNA exit channel in RNAP and loses its grip on the PRS, nascent RNA enters the RNA exit channel and transcription enters the elongation stage.

In AI, the interactions between σ70 and the PRS limit the lengths of abortive transcripts. The stronger these interactions are, the longer the time RNAP will spend cycling in AI (8) and the longer the lengths of abortive transcripts will be (9). For these reasons, after establishing tight promoter interactions, transcription initiation is rather slow. Transcription elongation, on the other hand, is a very efficient and fast process (10–13). However, in elongation, RNAP may encounter specific sequences that cause transcriptional pausing, sometimes accompanied by backtracking of the RNA chain pushing its 3′ end into the RNAP secondary channel (14), the site of NTPs entry (15). Pausing and pause release in elongation constitute additional steps (discussed below).

Whereas the basic biochemical steps of RNA polymerization (including NTP entrance and insertion, phosphodiester bond formation, pyrophosphate removal and RNAP translocation) are the same both in initiation and in elongation, their overall rates are far slower in initiation due to the structural differences

## Significance

**Transcription initiation by RNA polymerase (RNAP) is a highly regulated rate-limiting step in many genes and involves numerous intermediate states that remain incompletely understood. Here, we report the characterization of a previously hypothesized slow initiation pathway involving RNAP backtracking and pausing. This backtracked and paused state is observed when all nucleoside triphosphates (NTPs) are present at physiologically relevant concentrations, but becomes more prevalent with unbalanced NTP levels, which may occur in vivo under conditions of metabolic stress. Pausing and backtracking in initiation may play an important role in regulating RNAP transcription. Moreover, similar RNA backtracked states may contribute to promoter-proximal pausing among eukaryotic RNA polymerase II enzymes.**

Author contributions: E.L., S.C., B.L.A., S. Wang, and S. Weiss designed research; E.L., S.C., B.L.A., S. Wang, S.W.L., L.W.G., Y.A., T.R.S., and D.J.T. performed research; S.C., J.L., A.I., X.M., and S.B. contributed new reagents/analytic tools; E.L., S.C., B.L.A., S. Wang, S.B., T.R.S., D.J.T., and S. Weiss analyzed data; E.L., S.C., B.L.A., S. Wang, S.B., T.R.S., D.J.T., and S. Weiss wrote the paper; E.L., S.C., T.R.S., D.J.T., and S. Weiss designed this study; E.L. and S.C. developed and designed the quenched-kinetics experiments; E.L., S.C., S.W.L., L.W.G., and Y.A. performed the quenched-kinetics measurements, in consultation with A.I. and X.M; B.L.A. designed and performed gel-based experiments; S. Wang and T.R.S. designed, performed, and analyzed magnetic tweezers experiments; and J.L. and S.B. prepared GreA protein.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]E.L., S.C., B.L.A., and S. Wang contributed equally to this work.

[2]To whom correspondence may be addressed. Email: sweiss@chem.ucla.edu, strick@biologie.ens.fr, or taatjes@colorado.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605038113/-/DCSupplemental.

between these states (16–18). In particular, the strong σ70–PRS interactions (absent in elongation) (19–21) and the interactions (steric and electrostatic) of the acidic tip of the σ70 region 3.2 (σR3.2, which blocks the RNA exit channel) and the 5′ end of the nascent transcript (22–25) contribute most to the slow rate of initiation. Nascent RNAs of up to 9 nt in length are stabilized by hybridization to the melted template DNA in the RNA:DNA hybrid (18, 26–28). Although RNAP can accommodate an RNA of up to 14 nt long (i.e., 9 nt in RNA:DNA hybrid + 5 nt in RNA exit channel) (29), the transition into elongation, whereby the RNA exit channel blockage by σR3.2 is relieved, begins when RNA reaches the length of 11 nt (16, 30–32). After removal of the exit channel blockage, RNAP undergoes a global conformational change, σ70 loses its grip on the PRS and RNAP processively and rapidly elongates until it reaches the termination signal (23).

In what ways can RNAP overcome the barrier between initiation and elongation? Previous experimental data suggested that in AI, RNAP scrunches downstream DNA into its active site (3, 4). Because σ70 holds the −10 and −35 DNA elements tightly, the dsDNA upstream to the transcription bubble stays as a duplex, whereas the 10 DNA bases in the noncoding part of the transcription bubble (bases −10 to −1 in both strands) become compressed. This results in a strain buildup into what is known as the "stressed intermediate" complex (5, 33). Additionally, the longer the initially transcribed nascent RNA is, the closer its negatively charged 5′ end will be to the acidic tip of σR3.2 (19, 24). Simultaneously, a longer nascent transcript makes more stable RNA:DNA hybrid, which, in turn, helps counteract strain from the compressed DNA bubble and electrostatic repulsion from the acidic tip of σR3.2. In AI, transcripts are held in RNAP until reaching a certain length, above which the strain is relieved either through RNA backtranslocation and its abortive release through the secondary channel (14, 34–36) or by pushing σR3.2 and unblocking the RNA exit channel (22–25).

The electrostatic repulsion between the acidic tip of σR3.2 and the negatively charged 5′ end of the nascent transcript should decrease with each backtranslocation step (i.e., the distance between the negatively-charged groups increases). Nascent transcripts >4 nt in length could be stabilized in a backtracked state resulting in initiation pausing without immediate abortive release. Similar situations in elongation stabilize the nascent transcript in the pretranslocated state (37–41), eventually causing the nascent RNA 3′ end to disengage from the active site, backtrack into the secondary channel and pause transcription. The backtracked RNA can be cleaved either by the intrinsic endonucleolytic activity of RNAP (very inefficient at physiological pH) or by the action of extrinsic transcript cleavage factors GreA and GreB (TFIIS in eukaryotic RNA polymerase II system (42)). This cleavage leads to realignment of the newly formed RNA 3′-terminus with the active site and reactivation of transcription (pause release) (43–45). Due to these parallels, we wanted to examine if backtracking and pausing occur in transcription initiation by RNAP as well.

Previously it has been shown that stalling of transcription initiation occurs in strong promoters (8, 46). In addition, it has been shown that GreA and GreB induce RNA cleavages in AI, reduce abortive cycling and stimulate RNAP transition to elongation in vitro and in vivo at certain promoters (5, 25, 47–55). Linking these

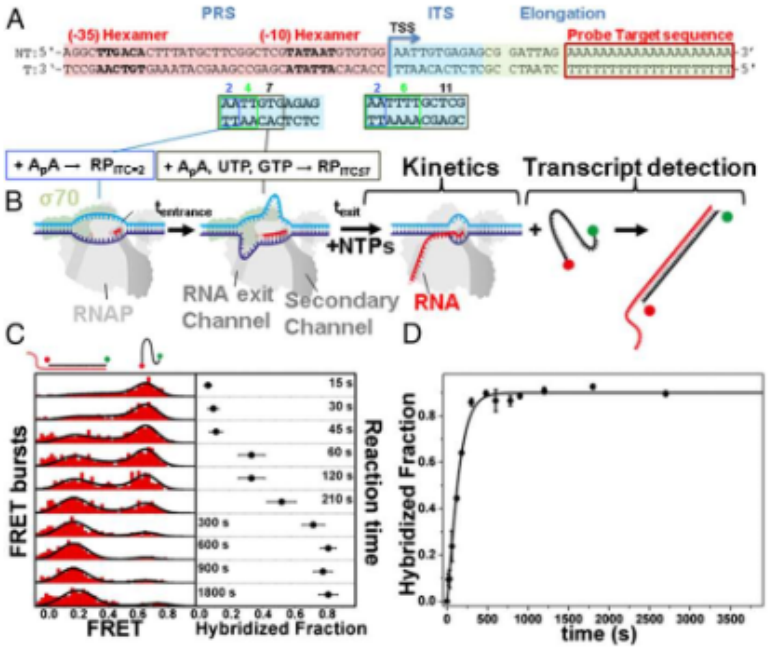



**Fig. 1.** Single-round transcription quenched kinetics assay. (*A*) Representative promoter sequence used here (lacCONS promoter) to show how by changing the initially transcribed sequence (ITS; cyan), different NTP-starved states can be generated ($RP_{ITC=2}$, $RP_{ITC\leq4,6,7}$, $RD_{E=11}$). Other regions of the promoter include the promoter recognition sequence (PRS; pink) and the elongation sequence (yellow), including a probe target complementary sequence (red). All promoters measured are described in Fig. S1. (*B*) Schematic of RNAP runoff transcription starting from a particular NTP-starved state (incubation with a partial set of NTPs for $t_{entrance}$). Upon supplementing all NTPs, transcription kinetics start and transcripts are quantified via hybridization to a ssDNA FRET probe for different incubation times ($t_{exit}$). (*C*) Example of quenched kinetics data generated from quantification of runoff transcripts. The example follows one repetition of the kinetics exiting from $RP_{ITC=2}$. (*D*) As an example for kinetic curve extraction, average runoff kinetics from various $RP_{ITC=2}$ are shown. The data points are averages of three repeats and the error bars are the SDs about these averages. The data are represented as points and the solid line represents the best-fit result to the model described in *Methods*. The best-fit values of the model parameters are shown in Table S1.

biochemical studies to the aforementioned endonucleolytic activity that the GreA–RNAP complex has in elongation, it was hypothesized that RNAP backtracking may occur in initiation and that GreA factors, as in elongation, act upon the nascent RNA bases that enter into the secondary channel of RNAP through backtracking. To directly address this question and to study the mechanism of transcription initiation in greater detail, we developed a solution-based, single-round quenched kinetics transcription assay that measures the kinetics of runoff transcript production. This assay was initially used to assess the kinetics of exit out of NTP-starved $RP_{ITC}$ states, using an *E. coli* transcription system reconstituted from native RNAP core enzyme and σ70. Such NTP-starvation experiments provided us with reliable means to interrogate RNAP transcription from specific initiation states and offered mechanistic insights that were further validated in the follow-up experiments with minimal or no starvation. In addition to quenched kinetics assays of runoff RNA, we performed gel-based in vitro transcription assays that focused on kinetics of abortive products. Finally, we carried out single-molecule magnetic tweezers experiments to monitor changes in transcription bubble size during transcription initiation by RNAP. Altogether, we report here the detection and direct observation of RNAP in a backtracked, paused state during early stages of transcription initiation. Although previously hypothesized, our results establish paused, backtracked RNAP as a bona fide initiation intermediate. This observation could have potential applications for molecular therapeutics and mechanistic implications for mammalian RNA polymerase II enzymes.

## Results and Discussion

**Single-Round Transcription Quenched Kinetics Assay.** We developed a single-round quenched kinetics assay (Fig. 1) to probe the kinetics of transcription by directly counting the number of transcripts produced over time. Using this assay, we initially examined the kinetics of *E. coli* RNAP transcription from distinct $RP_{ITC}$ states generated via NTP starvation (Fig. 1 *A* and *B*). The assay was based on quantification of single runoff transcripts by hybridization with a doubly labeled ssDNA probe (Fig. 1*B*). The number of hybridized probes (and hence the number of transcripts) was accurately determined using microsecond alternating-laser excitation (μsALEX)-based fluorescence-aided molecule sorting (ALEX-FAMS) (56, 57) (Fig. 1*C*). ALEX-FAMS is a method based on single-molecule Förster resonance energy transfer (smFRET) (58). A significant advantage of smFRET and ALEX-FAMS is the ability to identify distinct populations in a model-free manner, simply by counting single-molecule events of one sort (with a given FRET efficiency population) and comparing their number to single-molecule events of another sort (representing a distinct FRET efficiency population) (56, 57, 59, 60) (Fig. 1*C*). Hence, FAMS is a suitable method for the quantification of runoff transcripts at picomolar probe concentrations.

Our experimental design circumvented potential drawbacks of the quenched kinetics assay. Because we used low concentrations of RNAP and promoter DNA, the probability for RNAP reassembly/reinitiation was low, whereas single-round transcription reaction conditions were achieved kinetically and thermodynamically (61, 62) (details in *Supporting Information*, *Rationale of the Single-Round Character of the Quenched Kinetics Transcription Assay*). Furthermore, we designed the DNA template so that the probe hybridization sequence would anneal at the end of the fully elongated RNA (Fig. 1*A*); in this way, probe hybridization would not interfere with transcription initiation but instead report on successful promoter escape. Also, because transcription initiation in many genes is much slower than elongation (i.e., initiation is typically rate limiting) (63), the synthesis of relatively short, yet full-length RNA products (39-base and 41-base transcripts for the promoters studied here) reflected the rate of transcription initiation.

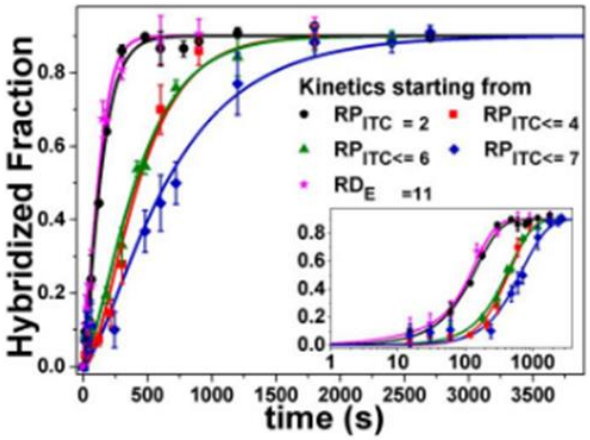



**Fig. 2.** Quenched kinetics transcription results identify an initiation-related stalled state. Shown are runoff kinetics from various NTP-starved states. Kinetics starting from late initiation states (e.g., $RP_{ITC\leq7}$, blue) are slower than from an earlier initiation state (e.g., $RP_{ITC=2}$, black). The data are represented as points and solid lines represent best-fit results to the model described in *Methods*. The best-fit values of the model parameters are shown in Table S1.

The ssDNA probe was doubly labeled with a FRET pair. When free in solution, the probe yields a single FRET population with a peak FRET efficiency of E ~ 0.75 (Fig. 1*C*). Hybridization of the ssDNA FRET probe with the elongation part of the runoff transcript yields a FRET efficiency population with a lower peak value of E ~ 0.3, due to the probe being stretched via hybridization to the 20A target sequence segment of the runoff transcript (Fig. 1 *A* and *C*). Correct assessment of transcription initiation rates by the quenched kinetics assay requires (*i*) formation of a stable initial state, (*ii*) addition of NTPs at t = 0, (*iii*) efficient and rapid quenching of the reaction at selected times, (*iv*) full hybridization of the ssDNA FRET probe to transcripts and (*v*) prevention of RNA degradation. As outlined below, our experiments were conducted under conditions that satisfied each of these requirements.

*i*) The formation of a stable initial transcribing state was achieved by using partial sets of NTPs that limited transcript polymerization to a given maximal length, which was dependent on the initially transcribed sequence (ITS) of the promoter DNA. We used the ITS of either the lacCONS (64, 65) (Fig. 1*A* and Fig. S1) or T5N25 (4) (Fig. S1) promoters, which allowed production of abortive transcripts of varying maximal lengths upon addition of different partial NTP sets (Fig. S1). Stable open complexes were formed by adding an initiating dinucleotide to achieve $RP_{ITC=2}$ (control measurements in Figs. S2 and S3; detailed discussion in *Supporting Information*, *Initiating Dinucleotide With/Without a 5′-Triphosphate Group*). To stabilize an $RP_{ITC}$ state ($RP_{ITC\leq i}$, i ϵ [4,6,7]) or an RNAP–DNA elongation complex state ($RD_{E=11}$), $RP_{ITC=2}$ was incubated for a given time, $t_{entrance}$, in the presence of a partial set of NTPs (Fig. 1*B*).

*ii*) In the absence of specific NTPs, NTP-starved $RP_{ITC}$ or $RD_E$ states should not produce runoff transcripts. The stability of NTP-starved $RP_{ITC}$ or $RD_E$ states was verified by negative control experiments, in which ssDNA FRET probe ALEX-FAMS measurements were performed with a limited set of NTPs, whereby hybridization of the FRET probe was not observed, and hence no runoff transcript was produced. Only after all four NTPs were added were runoff transcripts detected. Therefore, for the purposes of the assays described herein, the time at which all NTPs were added was set as t = 0.

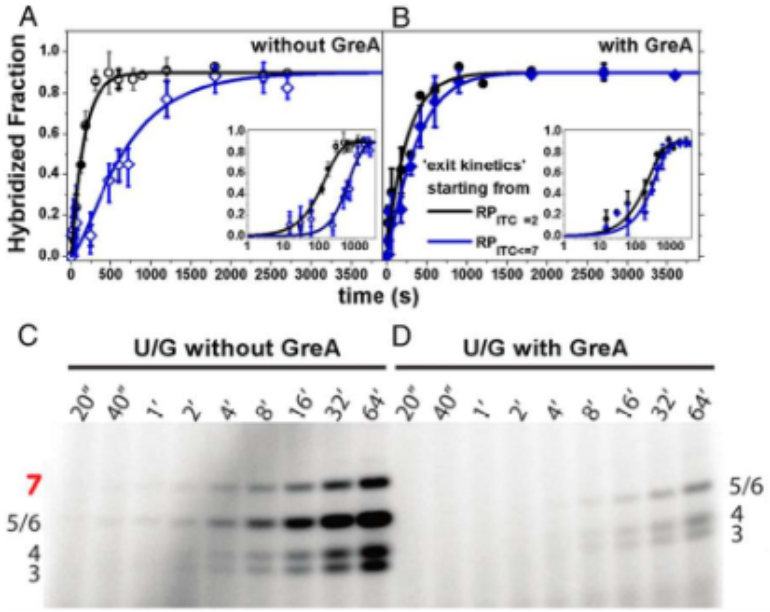



**Fig. 3.** GreA suppresses the kinetic delay in transcription initiation. (*A*) Runoff transcription kinetics are slower when starting from $RP_{ITC\leq7}$ (blue) than from $RP_{ITC=2}$ (black) (Fig. 2). (*B*) With 1 μM GreA, the delay in transcription initiation is reduced. The data are represented as points and solid lines represent best-fit results to the model described in *Methods*. The best-fit values of the model parameters are shown in Table S1. (*C* and *D*) Gel-based abortive initiation kinetics: Without GreA, NTP-starved $RP_{ITC\leq7}$ produced abortive transcripts up to 7 bases long, whereas the 7-base product was not produced with 1 μM GreA, suggesting 2 bases of 3′-backtracked RNA were cleaved by RNAP in a GreA-catalyzed reaction during initiation. Band assignment is provided in Fig. S8*A* and the accompanying legend.

*iii*) The kinetics of runoff transcript production, starting from an NTP-starved state, were measured using a constant $t_{entrance}$ = 40 min, followed by a varying incubation time with all four NTPs ($t_{exit}$), followed by efficient quenching of the transcription reaction. $t_{entrance}$ = 40 min was selected to be in the plateau of the measured ~7-min entrance kinetics (Fig. S4). Guanidinium hydrochloride (GndHCl; 0.5 M) served both as a reaction quencher (Fig. S5) and as an enhancer of hybridization of the ssDNA FRET probe to the runoff transcript (Fig. S6).

*iv*) After quenching the transcription reaction, the 20-base ssDNA FRET probe was added. The hybridization of the ssDNA FRET probe to the elongation segment of the runoff transcript was probed by ALEX-FAMS (Fig. 1*C*). The number of runoff transcripts per time point was calculated from the results of a global fit to the ALEX-FAMS–derived FRET histograms to a two-Gaussian model (Fig. 1*C* and *Supporting Information*, *μsALEX-smFRET Analysis for the Quantification of Transcription Kinetics*). Then the fraction of events in the low-FRET subpopulation, "hybridized fraction" (the number of runoff transcripts produced, relative to known amount of ssDNA FRET probe) is plotted as a function of $t_{exit}$ (Fig. 1*C*). Repeats of the same measurement are performed to produce reliable averaged kinetic curves (as in Fig. S7) that can be later used for curve fitting (Fig. 1*D*).

*v*) Throughout the experiments, an RNase inhibitor was added to the reaction mixture to prevent degradation of RNA products.

**Slower Transcription Initiation Kinetics from Select NTP-Starved States.** Because promoter escape is the rate-limiting step in initiation, we anticipated that RNAP transcription kinetics starting from different $RP_{ITC}$ states ("exit kinetics") would be similar. However, we found that exit kinetics from the $RP_{ITC\leq4}$, $RP_{ITC\leq6}$, or $RP_{ITC\leq7}$ states were slower than from $RP_{ITC=2}$ state (Fig. 2). In fact, whereas exit kinetics from $RP_{ITC=2}$ were similar to those of RNAP already in the elongation state ($RD_{E=11}$), the exit kinetics from $RP_{ITC\leq7}$ were at least 3.5 times slower (Table S1). These results suggested the existence of a previously hypothesized paused state in RNAP transcription initiation ($RP_{ITC\leq i}$*; see Fig. 5). Importantly, this state was transient and overall RNAP activity remained unchanged, given that all "delayed" RNAP complexes ($RP_{ITC\leq4}$, $RP_{ITC\leq6}$, or $RP_{ITC\leq7}$) eventually transitioned to elongation (Fig. 2).

**Delayed Initiation Kinetics Are Associated with Backtracking.** It is well established that elongating RNAP can backtrack and pause (44, 45). In such circumstances, the nascent RNA 3′ end backtranslocates into the secondary channel, where it can undergo endonucleolytic cleavage by the GreA–RNAP complex (25, 44, 48, 66). To test whether delayed exit kinetics for $RP_{ITC\leq7}$ were associated with backtracking during initiation, we assessed the effect of GreA, using our single-round quenched kinetics assay. As shown in Fig. 3 *A* and *B*, the addition of GreA at physiologically relevant concentration of 1 μM accelerated the exit kinetics from $RP_{ITC\leq7}$ relative to the exit kinetics from $RP_{ITC=2}$ (~50% recovery from the $RP_{ITC\leq7}$* stalled state; *Methods* and Table S1). These results were consistent with GreA-dependent release of RNAP from a backtracked and paused state in elongation (44, 45, 47).

To further test the effect of GreA during RNAP transcription initiation, we performed in vitro transcription assays in which $^{32}$P-labeled abortive transcripts were quantified following polyacrylamide gel electrophoresis (PAGE). This enabled identification of various abortive transcripts (band assignment in Fig. S8*A* and accompanying legend) and thus provided a means to determine whether GreA catalyzed the cleavage of short transcripts during transcription initiation. As shown in Fig. 3 *C* and *D*, the

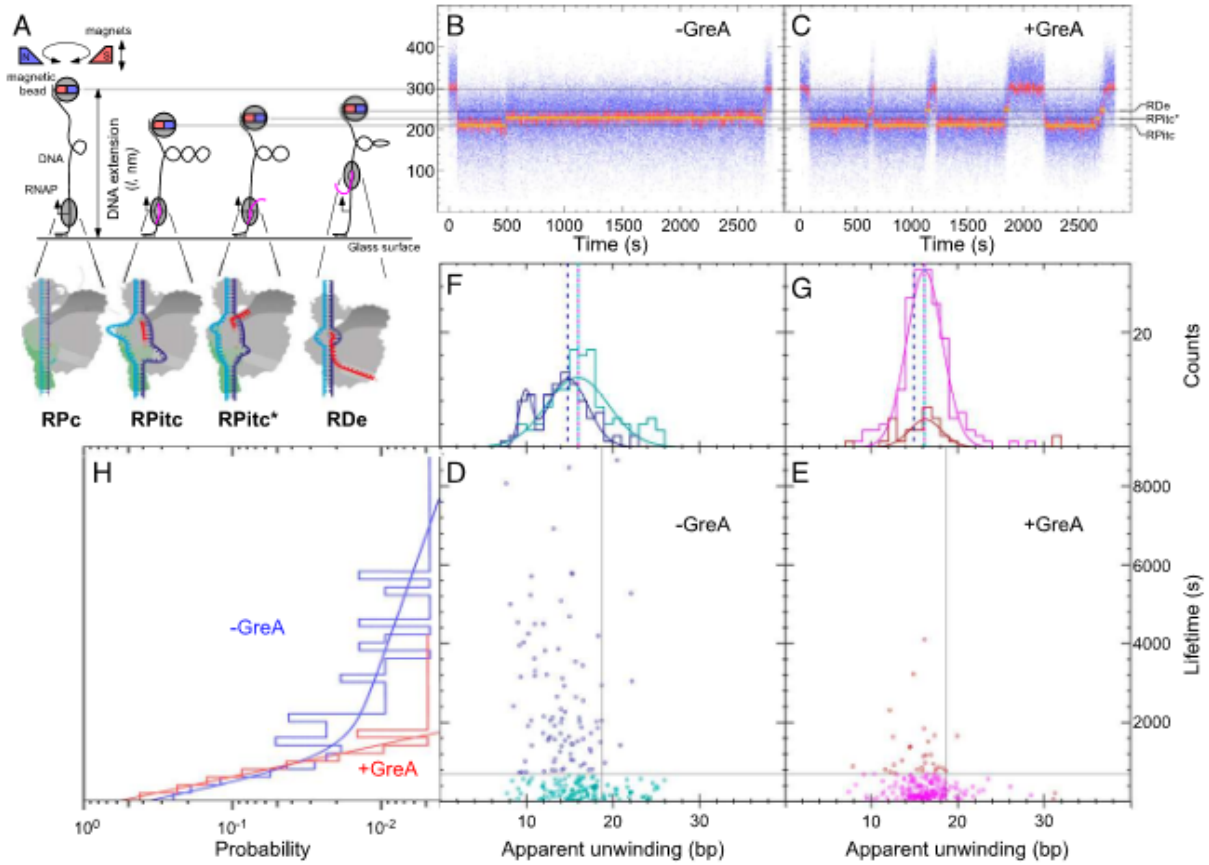



**Fig. 4.** Backtracking in initiation correlates with RNAP pausing in the presence of equimolar NTPs. (*A*) Schematics of the magnetic tweezer transcription assay (*Methods*). (*B* and *C*) Representative bead extension trajectories shown for single-molecule transcription experiments without (*B*) or with (*C*) 1 μM GreA. Unwinding levels (gray lines) are shown, indicating different bubble sizes imposed by different RNAP states (below). Yellow lines highlight typical lifetimes in each state. (*D–H*) Unwinding levels and $RP_{ITC}$ and $RP_{ITC}$* lifetimes of individual initiation events (i.e., averaging over all $RP_{ITC}$ and $RP_{ITC}$* states seen from initiation to promoter escape) are summarized into unwinding–lifetime scatter plots without (*D*) or with (*E*) 1 μM GreA; their 1D projections are shown in *F–H*. Quadrant structure is built as discussed in *Methods*, *Illustrations*. Lifetime data in the absence or presence of GreA were first fitted to single or double exponentials based on goodness-of-fit. Then, the 2D data were temporally separated into events shorter than (cyan, absence of GreA; magenta, presence of GreA) or longer than (dark blue, absence of GreA; dark red, presence of GreA) the fast timescale for promoter escape (~300 s) plus 1 SD. Similarly, the 2D data were spatially separated into events with apparent unwinding amplitude smaller or larger than the mean unwinding observed during short-duration escape events plus 1 SD. Apparent unwinding data associated with short- or long-escape timescales were then fitted to single- or double-Gaussian distributions based on goodness-of-fit and according to the color code described above. Dotted vertical lines are visual guides to the maximum of the respective Gaussian distributions. Twenty to fifty DNA templates, carrying the lacCONS promoter sequence, were used for each condition, with 5–10 transcription pulses per template.

production of the 7-base abortive RNA transcript was significantly suppressed in the presence of GreA. Because GreA stimulates cleavage only of "backtracked" RNA (i.e., when the RNA 3′ end is inserted into the secondary channel), these data, combined with our single-round kinetics data, confirmed that RNAP backtracks during transcription initiation. Because GreA predominantly induces cleavage of 2–3 bases from the RNA 3′-terminus (44, 45, 50–52), the lack of the 7-base abortive transcript is most likely due to GreA-induced shortening (compare Fig. 3 *C* and *D*) to a 5-mer, suggesting that RNAP in NTP-starved $RP_{ITC\leq7}$ backtracks by 1 base.

**RNAP Backtracking and Pausing in the Presence of All NTPs.** The data summarized in Figs. 2 and 3 studied pausing and backtracking in RNAP transcription initiation under conditions in which one or more NTPs were absent. Whereas the complete absence of select NTPs is highly unlikely in vivo, changes in pools of NTP levels may occur in bacteria in several conditions (e.g., metabolic stress, stages in cell growth; discussions in refs. 67–69). Accordingly, we examined RNAP transcription with the quenched kinetics assay with all NTPs present but under NTP concentration imbalance (UTP and GTP >> ATP and CTP at the lacCONS promoter; UTP and ATP >> CTP and GTP at the T5N25 promoter). According to the DNA sequences of these promoters (Fig. S1), these conditions were expected to temporarily populate delayed $RP_{ITC}$ states ($RP_{ITC\leq7}$ in lacCONS and $RP_{ITC\leq8}$ in T5N25, respectively). Consistent with the results described above, we observed a delay in exit kinetics from the $RP_{ITC=2}$ state under conditions of NTP imbalance (Fig. S9, orange) compared with equimolar conditions (Fig. S9, black) at each of the two different tested promoter templates. In addition, the observed delay was smaller in magnitude than the delay achieved upon exit from an NTP-starved $RP_{ITC}$ state (Fig. S9, blue). This result implies the kinetic delay inversely correlates with the abundance of the NTPs required for promoter clearance.

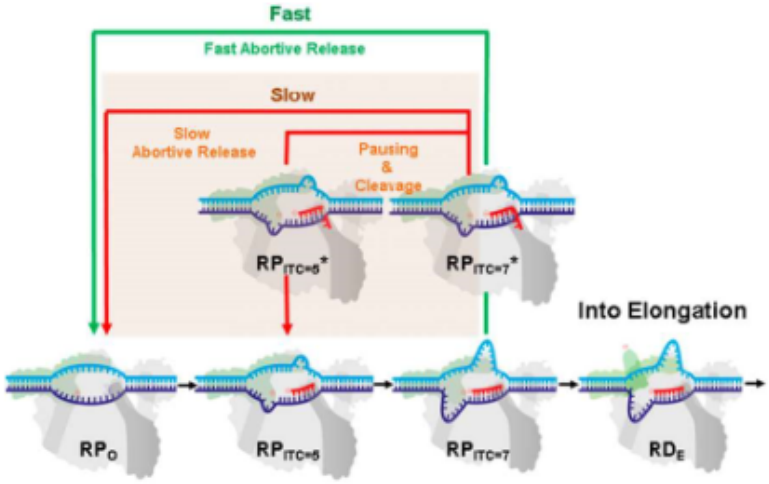



**Fig. 5.** A modified transcription initiation model. RNAP transcription initiation branches to (*i*) promoter clearance and transitions into elongation (black arrows) or into (*ii*) release of abortive transcripts (green and red arrows). After initial backtracking steps (e.g., from $RP_{ITC=7}$ to $RP_{ITC=7*}$), the complex can continue with either (*iii*) fast abortive transcript release (classic model, green arrow) or (*iv*) transition into a paused and backtracked state. Exit from the paused–backtracked state can occur either by successive slow backtracking steps (red arrow) or through intrinsic cleavage of RNA bases that are in the secondary channel, which prepares RNAP in, e.g., the $RP_{ITC=5}$ state. Upon cleavage, the complex can release the abortive transcript or reestablish RNA polymerization (e.g., from the $RP_{ITC=5}$ state).

The single-round quenched transcription kinetics assay and transcription assays with $^{32}$P-labeled UTP are ensemble experiments and hence cannot reliably detect infrequent or transient events. Under conditions of NTP starvation, abortive transcripts are identifiable by the gel-based assay (Fig. 3*B*). However, in the presence of all NTPs we did not obtain radiolabeled bands with sufficient intensity for the quantification of abortive transcripts, due to a limited number of abortive transcripts produced under single-round transcription conditions (Fig. S8*B*). The single-round transcription kinetics starting from $RP_{ITC=2}$ may already be influenced by pausing and backtracking; however, in the quenched kinetics assay $RP_{ITC=2}$ serves as a reference point.

To detect the full initiation behavior on the lacCONS promoter in the presence of equimolar NTPs (100 μM each) with high sensitivity, we implemented magnetic tweezers experiments with positively supercoiled promoter templates (Fig. 4*A*). This assay allowed us to track individual RNAP complexes over time and simultaneously detect and identify distinct $RP_{ITC}$ states, based upon well-established changes in DNA extension (4). In the absence of GreA, we observed short- and long-lived $RP_{ITC}$ states (Fig. 4 *B*, *D*, *F*, and *H*). The lifetimes of $RP_{ITC}$ states (n = 216) are summarized in a histogram fitted with a double exponential, in which ~60% of events were short-lived ($\tau = 340 \pm 60$ s, SEM) and ~40% were long-lived ($\tau = 4{,}600 \pm 2{,}700$ s, SEM; Fig. 4*H*, blue; fractions based on quadrant analysis as discussed in Fig. 3 legend). Correlating these lifetimes with DNA bubble sizes (representing distinct $RP_{ITC}$ states; Fig. 4 *A*, *D*, and *F* and accompanying legend) revealed that short-lived $RP_{ITC}$ states are characterized by a Gaussian distribution of transcription bubble sizes with a mean apparent unwinding of 15.9 ± 0.4 bp (SEM) and a SD of 3.6 ± 0.4 bp, whereas long-lived events display two distinct states of apparent unwinding: a 20% fraction with mean apparent unwinding of only 9.9 ± 0.3 bp (SEM) and a SD of 0.7 ± 0.3 bp (SEM) and an 80% fraction with mean apparent unwinding of 14.8 ± 0.4 bp (SEM) and a SD of 2.3 ± 0.4 bp (SEM).

These data suggested that, in addition to the well-characterized $RP_{ITC}$ state (Fig. 4*A*), a large fraction of RNAP complexes entered a distinct, long-lived state characterized by a smaller transcription bubble (denoted $RP_{ITC}$* in Fig. 4 *A*–*C*). We hypothesized that this long-lived initiation intermediate represented a backtracked RNAP that was characterized also in our quenched kinetics and gel-based transcription assays. If correct, the addition of GreA would be expected to markedly reduce the number of these long-lived events. In agreement, $RP_{ITC}$ states (n = 209), observed in the presence of GreA, became uniformly short-lived ($\tau = 350 \pm 30$ s, SEM) and displayed a bubble size distribution similar to that observed for the short-lived $RP_{ITC}$ states in the absence of GreA [16.2 ± 0.2 bp (SEM); SD 2 ± 0.2 bp (SEM); Fig. 4 *C*, *E*, *G*, and *H*]. The results shown in Fig. 4 followed transcription rounds of single RNAP molecules on immobilized DNA molecules carrying the lacCONS promoter sequence. These experiments were performed in the presence of the initiating dinucleotide used in the quenched kinetics assay (Figs. 1*D*, 2, and 3 and Figs. S3, S4, S8, and S9), $A_pA$. Additional magnetic tweezer transcription experiments using the lacCONS promoter with all NTPs in the absence of $A_pA$ yielded similar trends to those in the presence of it (example of a trajectory in Fig. S10). Experiments conducted on a different promoter template (T5N25) at 100 μM each NTP also showed the same trends, which were furthermore independent of the presence (Fig. S11) or absence (Fig. S12) of the initiating dinucleotide of T5N25 promoter, $A_pU$. Specifically, we observed both short-lived $RP_{ITC}$ states lasting tens of seconds and long-lived $RP_{ITC}$* states lasting thousands of seconds, for which the $RP_{ITC}$* state systematically displays a smaller mean unwinding amplitude with narrower distribution than the $RP_{ITC}$ state. As observed for the lacCONS promoter, addition of GreA once again abrogates the minimally unwound, long-lived $RP_{ITC}$* state in favor of short-lived $RP_{ITC}$ states with extensive unwinding and leading to highly efficient promoter escape.

Finally, we performed magnetic tweezer transcription experiments on the lacCONS promoter and at higher NTP concentrations (1 mM each NTP; no initiating dinucleotide), which resemble the concentrations of NTPs under normal physiological conditions. The results obtained in the absence of GreA (Fig. S13) continue to show long-lived $RP_{ITC}$* events that resemble the ones measured under lower NTP concentrations. The transcription initiation complexes formed at the lacCONS promoter in these conditions continue to display double-exponential lifetimes before bona fide escape and regular processive elongation: 84%

with a fast phase of only 91 ± 15 s (SEM) and 16% with a slow phase remaining in the thousands of seconds (Fig. S13).

In summary, the presence of GreA abrogates $RP_{ITC}$* states characterized by long lifetimes and smaller transcription bubbles in favor of $RP_{ITC}$ states with short lifetimes and larger transcription bubbles, leading to highly efficient promoter escape (Figs. S11 and S12 *D* and *E*, quadrants, and S13*C*, quadrants). Hence, we conclude that GreA catalyzes the cleavage of nascent RNA transcript in backtracked complexes. The cleavage allows scrunching to resume so that the transcription bubble may reach its maximal size, efficiently driving the transition to a productive elongation. It also implies that the long-lived paused state is either caused by or stabilized through backtracking. These data further support the existence of a long-lived initiation state, in which the transcription bubble is smaller than its size in the presence of GreA, most likely due to backtracking. Altogether, the results provide evidence for the existence of the previously hypothesized paused and backtracked initiation state.

## Conclusions

Our results support the existence of a previously hypothesized, yet uncharacterized, state in which RNAP backtracks and pauses during transcription initiation. We identified this paused–backtracked initiation intermediate under in vitro conditions with NTP concentrations that resemble physiological conditions (~100 μM to 1 mM). Moreover, GreA and NTP availability appear to play key roles in regulating the flux in or out of this state. Based upon these results, we propose a modified transcription initiation model (Fig. 5). The model proposes that transcription initiation is rate limiting not only due to multiple abortive cycles that can occur before transitioning into elongation (5, 31, 63, 70), but also due to the existence of a backtracked, paused state. Thus, we hypothesize the presence of two initiation pathways. If nascent RNA backtracks into the secondary channel, either RNAP can swiftly release it as an abortive product, if the hybrid is short enough (14, 35, 36), or the backtracked complex can be stabilized. The prolonged association of the backtracked transcript may lead to intrinsic (or GreA-induced) RNA cleavage by RNAP. The backtracked-paused complex may also slowly release the nascent RNA through further backtracking. The model depicted in Fig. 5 adds an additional "slow" pathway (red arrows and beige rectangle background) to the conventional abortive initiation model.

Although the paused–backtracked initiation intermediate was observed frequently even at high NTP concentration, its effect on transcription and promoter escape was most pronounced under limiting NTP concentrations. These findings could have direct implications for in vivo conditions. First, the intracellular concentrations of NTPs can vary by severalfold, even during normal bacterial growth from midlog to early stationary phase (68). Additionally, nutritional limitation (e.g., carbon sources) can lead to a significant decrease of intracellular NTPs that directly affect the efficiency of transcription initiation and start-site selection at promoters of ribosomal and pyrimidine biosynthetic operons [reviewed by Turnbough (69)]. Moreover, decreases in nucleotide pools become even more dramatic during metabolic stress induced by antibiotics, oxidative stress, etc. (67). Therefore, under certain growth and environmental conditions, NTP concentrations may become limiting and thus are likely to affect abortive synthesis, early transcription, pausing, and promoter escape. Second, a number of in vitro studies have shown that the rate-limiting steps in transcription initiation include promoter DNA recognition/binding, open complex formation, abortive RNA synthesis/release, and promoter escape, all of which are targets for regulation (53). In vivo analysis of the chromosomal distribution of bacterial RNAPs by ChIP-chip, ChIP-seq, and DNA footprinting demonstrated that even under normal growth conditions (without NTP deprivation), a substantial fraction of RNAP accumulated at promoters of both transcriptionally active and inactive regions, engaged in either "poised" or open promoter complexes or in initial transcribing and paused complexes (43, 71–74). These results indicate that abortive initiation and promoter escape are the major rate-limiting steps in vivo.

Mechanisms of transcription by cellular RNA polymerases are broadly conserved (75). For example, scrunching (3, 4, 76) and trigger loop function during catalysis (37, 77–80) are similar, as are some of the mechanistic roles of σ70 and TFIIB (1, 81). It will be important to determine whether similar backtracked states are adopted during transcription initiation by eukaryotic RNA polymerases. Mammalian RNA polymerase II (Pol II) pauses during early stages of transcription, and this represents a common regulatory intermediate (82). Potentially, a mechanistic intermediate of paused mammalian Pol II enzymes may involve RNA backtracking; such backtracked intermediates may help explain why TFIIS, a eukaryotic convergently evolved analog of GreA, has been linked to transcription initiation and assembles with Pol II at the promoter (42, 83, 84). Finally, we emphasize that, because RNAP transcription initiation is rate limiting at many genes and highly regulated in vivo, this previously hypothesized yet uncharacterized backtracked-paused RNAP state may lead to potential new strategies for molecular therapeutics and to the development of novel antibiotics.

## Methods

### Transcription Quenched Kinetics Assay.

***Preparation of a stable $RP_O$, $RP_{ITC=2}$.*** $RP_O$ solution is prepared with 3 μL *E. coli* RNAP holoenzyme (NEB, M0551S; 1.6 μM), 10 μL 2× transcription buffer [80 mM Hepes KOH, 100 mM KCl, 20 mM $MgCl_2$, 2 mM dithiotreitol (DTT), 2 mM 2-mercaptoethylamine-HCl (MEA), 200 μg/mL BSA, pH 7], 1 μL of 1 μM lacCONS (65) or T5N25 promoter (5) (sequence in Fig. 1*A* and Fig. S1), and 6 μL of water. $RP_O$ is then incubated in solution at 37 °C for 30 min. To remove unreacted and nonspecifically bound RNAP, 1 μL of 100 mg/mL Heparin-Sepharose CL-6B beads (GE Healthcare) is added to $RP_O$ solution together with 10 μL of prewarmed 1× transcription buffer [Heparin challenge (3, 65)]. The mixture is incubated for 1 min at 37 °C and centrifuged for at least 45 s at 6,000 rpm. A total of 20 μL of the supernatant containing $RP_O$ is transferred into a new tube with initiating dinucleotide, for sP-DNA complexes (40–42). In this context, the Heparin-challenged $RP_O$ solution is incubated with 1.5 μL of 10 mM adenylyl(3′–5′) adenosine or adenylyl(3′–5′) uridine ($A_pA$ or $A_pU$; Ribomed) at 37 °C for 20 min to form stable $RP_{ITC=2}$ solution. This $RP_{ITC=2}$ solution is used as a stock for all transcription reactions (for more information regarding the use of initiating dinucleotides, see *Supporting Information*, *Initiating Dinucleotide With/Without a 5′-Triphosphate Group*). A total of 2 μL of RNase inhibitor (NEB; M0314S) is added into the $RP_{ITC=2}$ solution to prevent degradation of newly transcribed RNA molecules.

***Design and measurement of the transcription kinetics.*** To produce runoff transcripts, high-purity ribonucleotide triphosphates (NTPs) (GE Healthcare) were used in all transcription reactions at 100 μM each. To obtain a specific initiation or elongation state, only a partial set of NTPs was used. The choice of the partial set of NTPs depended on the sequence of ITSs being used (Fig. 1*A* and Fig. S1). To exit from the initiation/elongation NTP-starved state the reaction mixture was complemented with all four NTPs. The nontemplate strands of all promoter DNAs have the sequence of 20 dA (20 consecutive A), the complementary sequence of ssDNA FRET probe (20 dT), at the end (Fig. 1*A* and Fig. S1 for probe target sequence). The ssDNA FRET probe is doubly end labeled with a pair of fluorophores suitable for smFRET: a donor, tetramethylrhodamine, at the 5′ end (5′ TAMRA modification) and an acceptor, Alexa Fluor 647, at the 3′ end (3′ Alexa Fluor 647 modification) [ordered from IDT (65)].

For kinetics, the reaction mixture is incubated with the partial set of NTPs for a constant duration of 40 min at 37 °C. All four NTPs are then added to the reaction mixture and incubated for various times, $t_{exit}$, at 37 °C. The reaction is quenched by addition of 0.5 M GndHCl. Subsequently, a ssDNA FRET probe is added to the quenched reaction mixture and incubated for 20 min at room temperature to detect runoff RNA transcripts produced during $t_{exit}$ (see Fig. S6 for hybridization kinetics in 0.5M GndHCl). The quenched-probed reaction mixtures were then used for μsALEX measurements. An example of the quenched kinetic assay FRET results is shown in Fig. 1 *C* and *D*.

Each kinetic curve was measured in two or three separate repeats to show reproducibility (Fig. S7). The concentration of the ssDNA FRET probe in each

experimental repeat was designed so that the steady-state level of a kinetic curve would be 0.9 ± 0.05. To make all repeats of kinetic curves reach a steady-state level of exactly 0.9 and not close to it, these repeats had to be rescaled (as shown in Fig. S14 and a detailed explanation in *Supporting Information, RNAP–DNA Complex Concentration Adjustment and Rescaling for Transcription Kinetics Assays*).

Each time point in the quenched kinetics assay is measured for duration of 10–15 min, using a setup described in Panzeri et al. (85), using Perkin-Elmer SPADs and 532-nm and 638-nm CW lasers operating at powers of 170 μW and 80 μW, respectively.

For transcription kinetics experiments with GreA, 1 μM GreA is added to transcription complexes in NTP-starved initiation or elongation states and incubated for 15 min before adding all four NTPs to initiate transcription reactions for $t_{exit}$.

Each kinetic measurement was performed at least in duplicates, using different preparations obtained on different days (Fig. S7). For each batch, we made sure of the following:

*i*) The FRET probe in the presence of transcription complexes without NTPs yielded a large high-FRET population with no or a negligible amount of low-FRET population (no NTP, negative control).

*ii*) After 20 min incubation of $RP_{ITC=2}$ with all four NTPs (steady state for $RP_{ITC=2}$; Figs. 1*D*, 2, and 3 *A* and *B* and Figs. S3, S4*A*, S7, S9, and S14), the fraction of hybridized probe reaches 90 ± 5% (positive control). This control is performed daily on the same batch used to prepare NTP-starved RNAP states.

*iii*) With an extralong incubation time (typically several hours) after quenching the reaction, the measurement yielded the same hybridized fraction as measured without the extra incubation (quenching does work).

The results of the negative control (no NTP) serve as the "t = 0" time point. The positive control shows that the prepared transcription complexes are active and produce a lower amount of transcripts than the detection limit of the assay (overall amount of transcripts is lower than the amount of ssDNA FRET probe) at very long times, "t = ∞."

All NTP-starved states were prepared from $RP_{ITC=2}$ stock solution. Because the other experimental conditions (concentrations, temperature, etc.) are identical, any changes in activity that may be caused solely due to the starvation of NTPs will show a change in the hybridized fraction in long time points of the kinetic trace. Such comparisons were routinely performed and have never shown a difference in the long time point baseline between the kinetics from $RP_{ITC=2}$ and from NTP-starved states (within 5% error). Therefore, we conclude that our experimental conditions (e.g., NTP starvation) did not alter the transcription activity but only the kinetics. In this regard, the same quantity of the stock solution could be used for kinetics assays for NTP-starved states as the one used for the abovementioned positive control.

μsALEX-smFRET data were analyzed as described in *Supporting Information, μsALEX-smFRET Analysis for the Quantification of Transcription Kinetics*.

**Transcription Assays Visualizing Abortive Product Formation Using Urea-Denaturing PAGE Analysis of [$^{32}$P]-Radiolabeled RNA Products.** Abortive transcription assays were run using the lacCONS promoter, having its probe target 20A sequence replaced by the WT lacUV5 sequence at registers from +20 to +39 (Fig. S8). Three units of RNAP holoenzyme (NEB; M0551S) was mixed with 50 nM promoter DNA in 1× transcription buffer in a final volume of 20 μL. The reaction was then incubated at 37 °C for 20 min to form $RP_O$, followed by addition of 1 μL of 100 mg/mL Heparin-Sepharose beads and 10 μL of transcription buffer. The mixture was incubated for ∼1 min and centrifuged, and 20 μL of the supernatant was removed and added to 10 μL prewarmed transcription buffer. After incubating an additional 10 min, $A_pA$ was added at a final concentration of 1.3 mM and incubated for 30 min to form the $RP_{ITC=2}$. The $RP_{ITC=2}$ was then diluted to 400 μL with transcription buffer containing SUPERaseIN (AM2696; Thermo Fisher Scientific) to final concentrations of 1.7 nM template, 112 μM $A_pA$, and 0.3 units/μL SUPERaseIN. This solution was stored at room temperature and used as a stock for each time course.

For time-course experiments, 90 μL of the stock solution was briefly incubated to bring it to 37 °C. To analyze the production kinetics of abortive products from $RP_{ITC\leq7}$, stock solution was mixed with 10 μL of 200 μM UTP+GTP mixture supplemented with ∼10 μCi [$\alpha^{32}$P]UTP. At each time point, 10 μL aliquot was then removed and mixed with an equal volume of formamide gel loading buffer. To analyze abortive product formation from RNAP that was not stalled, the UTP+GTP mixture was replaced by a complete set of NTPs. In experiments looking at the effects of GreA on abortive product formation, additional 15-min incubation at 37 °C was performed before the addition of NTPs, either in the presence or in the absence of 1 μM GreA. The stopped reaction aliquots were stored at −20 °C until running the urea-denaturing PAGE.

Samples were heated for 3 min at 90 °C and loaded on a 23% (wt/vol, 19:1 acrylamide:bis-acrylamide), 0.4-mm-thick urea-denaturing polyacrylamide gel. The gel usually ran for 5–6 h at 1,500 V in 1× TBE with an additional 0.3 M sodium acetate in the bottom well. The gels were then removed, dried, and exposed on a phosphor-storage screen about 2 d. Screens were visualized using a Typhoon PhosporImager.

**Magnetic Trapping Assay.**

***Single-molecule experiments.*** For the single-molecule experiments we used the DNA constructs described in *Supporting Information, DNA Constructs for the Magnetic Trapping Assay*.

Functionalized 2.2-kbp DNA molecules were first attached to 1-μm-diameter streptavidin-coated magnetic beads (MyOne Streptavidin C1; Life Technologies) and then tethered to a modified glass capillary surface coated with anti-digoxigenin (Roche) (86). Experiments were carried out on a homemade magnetic tweezer microscope to extend and supercoil the DNA, running the PicoTwist software suite to track and analyze the position of the magnetic bead. This position marks the free end and thus the extension of the functionalized DNA. Data were analyzed using custom routines in the Xvin software suite. Experiments were carried out in standard buffer at 34 °C, using 100 pM RNAP saturated with σ70 (prepared as in ref. 87) and 100 μM $A_pA$ (for experiments on lacCONS promoter) (we used 100 μM $A_pU$ for experiments on T5N25 promoter) and 100 μM each of ATP, UTP, GTP, and CTP. When added, GreA is at 1 μM.

Raw time traces were filtered (∼1 s) for analysis. The lifetime of the initiation state is defined as the total time elapsed between initial promoter unwinding and the transition to promoter escape and formation of the elongation complex ($RD_E$). The extent of apparent unwinding is determined by averaging the DNA extension over the entire lifetime of the initiation state as defined above and taking the difference between this average extension and that observed in the baseline high-extension state that separates the succession of transcription events. Importantly, our analysis excluded unproductive transcription initiation events in which RNAP spontaneously dissociated from DNA without forming the elongation state $RD_E$. Finally, we note that all elongation complexes formed, whether from $RP_{ITC}$ or $RP_{ITC}$*, displayed the expected elongation rate, namely ∼12 bases/s for 100 μM NTPs at 34 °C.

***Interpretation of apparent DNA unwinding during scrunching and backtracking and its relation to the length of transcribed RNA as seen in the quenched kinetics assay.*** In the supercoiling transcription assay where plectonemic supercoils are present (+4 positive supercoils throughout), the extension changes of the DNA construct report on the number of supercoils. Specifically, the DNA typically contracts by ∼55 nm for every additional supercoil when extended at low force (F = 0.3 pN) as in these experiments. DNA unwinding by RNAP is sensitively reported via its effect on overall DNA supercoiling: Conservation of linking number means that topological unwinding of 10.5 bp results in an ∼55-nm decrease in DNA extension.

During initial transcription and scrunching, an "N-2" rule has been observed (88), relating the length of the RNA and the extent of DNA unwinding in the bubble. Because an RNA 2-mer can be formed in RPo without a need for additional DNA unwinding and scrunching, an RNA N-mer can be formed by additional unwinding and scrunching of N-2 bases. However, during backtracking, this linear relationship is lost, although the simplest hypothesis is that backtracking by 1 bp dehybridizes 1 bp of the RNA–DNA hybrid and reduces the bubble size by the same amount. Thus, in this work the single-molecule assay can report only on the apparent bubble size, which results from the final backtracked state. This measurement provides information complementary to that obtained from the quenched kinetics assay, which instead provides insight into the relation between accessible RNA lengths and likelihood of entry into a state that is not competent for promoter escape.

***Separation of correlative lifetime/amplitude data into quadrants.*** We separate short- and long-lifetime initiation events by setting a boundary corresponding to the short mean lifetime plus 1 SD (as it is a single-exponential distribution this essentially means two times the short lifetime). We separate small and large bubbles by setting a boundary corresponding to the mean unwinding amplitude for the short-lifetime state in the absence of GreA plus 1 SD. Similar results are obtained if one uses the mean and SDs of amplitude data obtained in the +GreA condition and for which lifetime is homogeneously short. These boundaries determine the fractions of events categorized and cited as short or long lived and displaying small or large bubble sizes.

**Illustrations.** All illustrations of RNAP transcription initiation and elongation states have been prepared in Adobe Illustrator CC 2015.

**Note Added in Proof.** A study by Duchi et al. (89) that describes similar findings (but with much shorter paused initiation state), using somewhat different methodologies, was published while this manuscript was under final revisions.

**ACKNOWLEDGMENTS.** We thank Prof. William Gelbart, Prof. Charles Knobler, Dr. Cathy Yan Jin, and Xiyu Yi for fruitful discussions and Maya Lerner for preparation of illustrations. We thank Prof. Richard Ebright for generously providing us with the triphosphate dinucleotide pppApA, as a gift. The T.R.S. laboratory wishes to thank the CNRS, the University of Paris Diderot, and the Programme Equipe Labellisees of the French Ligue Contre le Cancer for core funding; this work was further made possible by ANR Grant "RepOne". This work was funded by the National Institutes of Health (Grant GM069709 to S. Weiss and Grant GM095904 to X.M. and S. Weiss) and the National Science Foundation (Grant MCB-1244175 to S. Weiss and D.J.T.).

1. Murakami KS, Darst SA (2003) Bacterial RNA polymerases: The wholo story. *Curr Opin Struct Biol* 13(1):31–39.
2. Young BA, Gruber TM, Gross CA (2002) Views of transcription initiation. *Cell* 109(4): 417–420.
3. Kapanidis AN, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314(5802):1144–1147.
4. Revyakin A, Liu C, Ebright RH, Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314(5802):1139–1143.
5. Hsu LM (2002) Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta* 1577(2):191–207.
6. Luse DS, Jacob GA (1987) Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *J Biol Chem* 262(31):14990–14997.
7. Carpousis AJ, Gralla JD (1980) Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry* 19(14): 3245–3253.
8. Ellinger T, Behnke D, Bujard H, Gralla JD (1994) Stalling of Escherichia coli RNA polymerase in the +6 to +12 region in vivo is associated with tight binding to consensus promoter elements. *J Mol Biol* 239(4):455–465.
9. Hsu LM (2009) Monitoring abortive initiation. *Methods* 47(1):25–36.
10. Abbondanzieri EA, Shaevitz JW, Block SM (2005) Picocalorimetry of transcription by RNA polymerase. *Biophys J* 89(6):L61–L63.
11. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438(7067):460–465.
12. Adelman K, et al. (2002) Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc Natl Acad Sci USA* 99(21):13538–13543.
13. Wang MD, et al. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science* 282(5390):902–907.
14. Epshtein V, et al. (2002) Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol Cell* 10(3):623–634.
15. Zhang G, et al. (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* 98(6):811–824.
16. Hansen UM, McClure WR (1980) Role of the sigma subunit of Escherichia coli RNA polymerase in initiation. II. Release of sigma from ternary complexes. *J Biol Chem* 255(20):9564–9570.
17. Kapanidis AN, et al. (2005) Retention of transcription initiation factor sigma70 in transcription elongation: Single-molecule analysis. *Mol Cell* 20(3):347–356.
18. Korzheva N, Mustaev A, Nudler E, Nikiforov V, Goldfarb A (1998) Mechanistic model of the elongation complex of Escherichia coli RNA polymerase. *Cold Spring Harb Symp Quant Biol* 63:337–345.
19. Zuo Y, Steitz TA (2015) Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol Cell* 58(3):534–540.
20. Campagne S, Marsh ME, Capitani G, Vorholt JA, Allain FH (2014) Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat Struct Mol Biol* 21(3):269–276.
21. Campbell EA, et al. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* 9(3):527–539.
22. Mekler V, et al. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* 108(5):599–614.
23. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* 296(5571):1285–1290.
24. Pupov D, Kuzin I, Bass I, Kulbachinskiy A (2014) Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res* 42(7):4494–4504.
25. Samanta S, Martin CT (2013) Insights into the mechanism of initial transcription in Escherichia coli RNA polymerase. *J Biol Chem* 288(44):31993–32003.
26. Lee DN, Landick R (1992) Structure of RNA and DNA chains in paused transcription complexes containing Escherichia coli RNA polymerase. *J Mol Biol* 228(3):759–777.
27. Nudler E, Avetissova E, Markovtsov V, Goldfarb A (1996) Transcription processivity: Protein-DNA interactions holding together the elongation complex. *Science* 273(5272):211–217.
28. Sidorenkov I, Komissarova N, Kashlev M (1998) Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell* 2(1):55–64.
29. Komissarova N, Kashlev M (1998) Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci USA* 95(25):14699–14704.
30. Travers AA, Burgessrr (1969) Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222(5193):537–540.
31. Straney DC, Crothers DM (1985) Intermediates in transcription initiation from the E. coli lac UV5 promoter. *Cell* 43(2 Pt 1):449–459.
32. Krummel B, Chamberlin MJ (1989) RNA chain initiation by Escherichia coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry* 28(19):7829–7842.
33. Straney DC, Crothers DM (1987) A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *J Mol Biol* 193(2): 267–278.
34. Hsu LM, et al. (2006) Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry* 45(29):8841–8854.
35. Weixlbaumer A, Leon K, Landick R, Darst SA (2013) Structural basis of transcriptional pausing in bacteria. *Cell* 152(3):431–441.
36. Martinez-Rucobo FW, Cramer P (2013) Structural basis of transcription elongation. *Biochim Biophys Acta* 1829(1):9–19.
37. Bar-Nahum G, et al. (2005) A ratchet mechanism of transcription elongation and its control. *Cell* 120(2):183–193.
38. Foster JE, Holmes SF, Erie DA (2001) Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* 106(2):243–252.
39. Holmes SF, Erie DA (2003) Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* 278(37):35597–35608.
40. Nedialkov YA, et al. (2003) NTP-driven translocation by human RNA polymerase II. *J Biol Chem* 278(20):18303–18312.
41. Sousa R (2001) A new level of regulation in transcription elongation? *Trends Biochem Sci* 26(12):695–697.
42. Guglielmi B, Soutourina J, Esnault C, Werner M (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci USA* 104(41):16062–16067.
43. Reppas NB, Wade JT, Church GM, Struhl K (2006) The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Mol Cell* 24(5):747–757.
44. Borukhov S, Lee J, Laptenko O (2005) Bacterial transcription elongation factors: New insights into molecular mechanism of action. *Mol Microbiol* 55(5):1315–1324.
45. Toulmé F, et al. (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J* 19(24):6853–6859.
46. Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H (2004) The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nat Struct Mol Biol* 11(6):551–557.
47. Shaevitz JW, Abbondanzieri EA, Landick R, Block SM (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* 426(6967): 684–687.
48. Stepanova E, et al. (2007) Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *J Bacteriol* 189(24):8772–8785.
49. Stepanova E, Wang M, Severinov K, Borukhov S (2009) Early transcriptional arrest at Escherichia coli rplN and ompX promoters. *J Biol Chem* 284(51):35702–35713.
50. Feng GH, Lee DN, Wang D, Chan CL, Landick R (1994) GreA-induced transcript cleavage in transcription complexes containing Escherichia coli RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J Biol Chem* 269(35):22282–22294.
51. Hsu LM, Vo NV, Chamberlin MJ (1995) Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc Natl Acad Sci USA* 92(25):11588–11592.
52. Borukhov S, Sagitov V, Goldfarb A (1993) Transcript cleavage factors from E. coli. *Cell* 72(3):459–466.
53. Hsu LM (2008) Promoter escape by Escherichia coli RNA polymerase. *Ecosal Plus* 3(1).
54. Goldman SR, Ebright RH, Nickels BE (2009) Direct detection of abortive RNA transcripts in vivo. *Science* 324(5929):927–928.
55. Skancke J, Bar N, Kuiper M, Hsu LM (2015) Sequence-dependent promoter escape efficiency is strongly influenced by bias for the pretranslocated state during initial transcription. *Biochemistry* 54(28):4267–4275.
56. Kapanidis AN, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci USA* 101(24):8936–8941.
57. Kapanidis AN, et al. (2005) Alternating-laser excitation of single molecules. *Acc Chem Res* 38(7):523–533.
58. Förster T (1948) Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Ann Phys* 2(1-2):55–75.
59. Deniz AA, et al. (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proc Natl Acad Sci USA* 96(7):3670–3675.
60. Deniz AA, et al. (2001) Ratiometric single-molecule studies of freely diffusing biomolecules. *Annu Rev Phys Chem* 52:233–253.
61. Mekler V, Pavlova O, Severinov K (2011) Interaction of Escherichia coli RNA polymerase σ70 subunit with promoter elements in the context of free σ70, RNA polymerase holoenzyme, and the β'-σ70 complex. *J Biol Chem* 286(1):270–279.
62. Mekler V, Minakhin L, Kuznedelov K, Mukhamedyarov D, Severinov K (2012) RNA polymerase-promoter interactions determining different stability of the Escherichia coli and Thermus aquaticus transcription initiation complexes. *Nucleic Acids Res* 40(22):11352–11362.
63. Margeat E, et al. (2006) Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys J* 90(4):1419–1431.
64. Mukhopadhyay J, et al. (2001) Translocation of sigma(70) with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* 106(4):453–463.
65. Kim S, et al. (2011) High-throughput single-molecule optofluidic analysis. *Nat Methods* 8(3):242–245.

66. Laptenko O, Lee J, Lomakin I, Borukhov S (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J* 22(23): 6322–6334.
67. Belenky P, et al. (2015) Bactericidal antibiotics induce toxic metabolic perturbations that lead to cellular damage. *Cell Reports* 13(5):968–980.
68. Buckstein MH, He J, Rubin H (2008) Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *J Bacteriol* 190(2):718–726.
69. Turnbough CL, Jr (2008) Regulation of bacterial gene expression by the NTP substrates of transcription initiation. *Mol Microbiol* 69(1):10–14.
70. Tang GQ, Roy R, Ha T, Patel SS (2008) Transcription initiation in a single-subunit RNA polymerase proceeds through DNA scrunching and rotation of the N-terminal subdomains. *Mol Cell* 30(5):567–577.
71. Herring CD, et al. (2005) Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J Bacteriol* 187(17):6166–6174.
72. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ (2005) Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA* 102(49):17693–17698.
73. Kusuya Y, Kurokawa K, Ishikawa S, Ogasawara N, Oshima T (2011) Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in Bacillus subtilis cells. *J Bacteriol* 193(12):3090–3099.
74. Hatoum A, Roberts J (2008) Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. *Mol Microbiol* 68(1):17–28.
75. Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9(2):85–98.
76. Fazal FM, Meng CA, Murakami K, Kornberg RD, Block SM (2015) Real-time observation of the initiation of RNA polymerase II transcription. *Nature* 525(7568):274–277.
77. Allison LA, Moyle M, Shales M, Ingles CJ (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42(2):599–610.
78. Hekmatpanah DS, Young RA (1991) Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. *Mol Cell Biol* 11(11): 5781–5791.
79. Thuillier V, Brun I, Sentenac A, Werner M (1996) Mutations in the alpha-amanitin conserved domain of the largest subunit of yeast RNA polymerase III affect pausing, RNA cleavage and transcriptional transitions. *EMBO J* 15(3):618–629.
80. Weilbaecher R, Hebron C, Feng G, Landick R (1994) Termination-altering amino acid substitutions in the beta' subunit of Escherichia coli RNA polymerase identify regions involved in RNA chain elongation. *Genes Dev* 8(23):2913–2927.
81. Sainsbury S, Niesser J, Cramer P (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493(7432):437–440.
82. Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nat Rev Genet* 13(10):720–731.
83. Kim B, et al. (2007) The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci USA* 104(41):16068–16073.
84. Stepanova EV, Shevelev AB, Borukhov SI, Severinov KV (2009) Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. *Biofizika* 54(5):773–790.
85. Panzeri F, et al. (2013) Single-molecule FRET experiments with a red-enhanced custom technology SPAD. *Proc SPIE Int Soc Opt Eng*, 8590.
86. Revyakin A, Ebright RH, Strick TR (2005) Single-molecule DNA nanomanipulation: Improved resolution through use of shorter DNA fragments. *Nat Methods* 2(2): 127–138.
87. Howan K, et al. (2012) Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* 490(7420):431–434.
88. Revyakin A, Ebright RH, Strick TR (2004) Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc Natl Acad Sci USA* 101(14):4776–4780.
89. Duchi D, et al. (2016) RNA polymerase pausing during initial transcription. *Mol Cell* 63(6):939–950.