**Title**
Investigation of Bacterial Transcription using Single Molecule Techniques



**Author**
Chung, SangYoon

**Publication Date**
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Investigation of Bacterial Transcription using Single Molecule Techniques

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemistry

By

Sang Yoon Chung

2016

ABSTRACT OF THE DISSERTATION

Investigation of Bacterial Transcription using Single Molecule Techniques

by

Sang Yoon Chung

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2016

Professor Shimon Weiss, Chair

The numerous complex molecular processes occurring inside living cells are primarily carried out by proteins and other biopolymers, such as ribonucleic acids (RNA). The identity and quantity of the different proteins and RNA determine the cell's phenotype and changes in response to the environment. Therefore, the internal composition of the cell in terms of the type and concentration of proteins and RNA is tightly regulated. Gene expression is the process of using the DNA sequence information to produce these biopolymers. Transcription, the initial step in gene expression, where one strand of DNA is used as template by the enzyme RNA polymerase (RNAP) for synthesizing a complementary RNA or transcript. Since cell phenotype is mostly determined by transcription, a complex regulatory mechanism exists involving a large number of factors to control the level of transcription of a gene. Although most studies are focused on multiple cycles of either transcription or association of DNA and RNA Polymerase (RNAP) to make RNAP-Promoter open complex ($RP_O$), single round transcription studies are crucial in elucidating the mechanism of sophisticated RNAP-DNA interactions and its kinetics in

transcription. In this context, we have developed a novel *in vitro* quenching based single round transcription assay using single molecule detection. Using this, we could successfully dissect initiation kinetics starting from different initial transcribing stages and found that transcription initiation doesn't follow a sequential model (as commonly believed). Instead, we identified a previously uncharacterized state that is unique to initial transcribing complexes and associated with the backtracked RNAP-DNA complex. Also, we have investigated the size/concentration effects of various osmolytes and macromolecular crowding agents, which mimic the crowded cellular environment, on actively-transcribing RNAP and found enhancement in transcription kinetics by larger crowding agents at the same viscosity.

The dissertation of Sang Yoon Chung is approved.

William M. Gelbart

Xiangfeng Duan

Pei-Yu Chiou

Shimon Weiss, Committee Chair

University of California, Los Angeles

2016

I was someone with nervousness and anxiety. However, everything has been changing since I met Jesus Christ as my lord. He has provided me comfort, peace and made me more confident and have a dream for the future. Without him, I wouldn't have been in this position. He always fills my life with a full of grace, which I don't deserve.

"Do not be anxious about anything, but in every situation, by prayer and petition, with thanksgiving, present your requests to God. And the peace of God, which transcends all understanding, guard your hearts and your minds in Christ Jesus" (Philippians 4:6-7)

## Table of Contents

## List of Tables and Figures

**Acknowledgements**

First of all, I would like to thank my advisor Dr. Shimon Weiss from the bottom of my heart for all of his scientific instruction and moral support. His patience and encouragement make me stand up and keep trying hard to be a better scientist and a better human being. Without him, I would have already given up my graduate study, which has been my dream for a long time.

I appreciate my committee members, Dr. Willliam Gelbart, Dr. Xiangfeng Duan, and Dr. Pei-Yu Chiou for being my committee members and their helpful criticism and advice on my research. Especially I thank Dr. Gelbart for his moral support and scientific discussion throughout my graduate study.

I also thank Dr. Dylan Taatjes for his scientific knowledge and guidance. I felt I learned a great deal from Dr. Taatjes after every teleconference meeting.

It is truly blessing to work with great colleagues. I appreciate formal and current Weiss lab members, Dr. Xavier Michalet, Dr. Jack Li, Dr. Young-Gyu Kim, Dr. Yuval Ebenstein, Dr. Devdoot Majumdar, Dr. Ron Lin, Dr. Jianmin Xu, Dr. Soohong Kim, Dr. Kyungwon Park, Dr. Sara Weitz, Dr. Antonino Ingargiola, Dr. Eitan Lerner, Zhiping Feng, Yazan Alhadid, Xiyu Yi, Yung Kuo, and Robert Boutelle for their friendship and precious scientific discussions. Whenever I needed help, they always came to me and willing to helping me.

I especial thank Dr. Young-Gyu Kim, Dr. Yuval Ebenstein, Dr. Devdoot Majumdar, Dr. Soohong Kim, Dr. Eitan Lerner, and Dr. Antonio Ingargiola for their mentorship. I believe that I have been getting closer to becoming an independent researcher with their scientific knowledge and enthusiasm toward the science.

Also, I would like to mention here how much I am grateful to my friend, Yazan Alhadid, Glen Chung, and Won-Gyu Lee. I deeply appreciate their support and all the delightful moments they brought me during my graduate school years.

I thank my family for the support and trust they placed in me throughout my life. My wife, Jihye Ju, is the woman who is more than I could have asked for. Her love and companionship help me to overcome and move forward every time I have an obstacle. My daughter, Grace Hyeyoon Chung, is truly the grace of God. I felt that God exists and was blessing me every time I saw my daughter.

Lastly, I wish to express my sincere appreciation to everyone that I shared my unforgettable graduate school life with.

# VITA

**Education**

**B.Sc.**, Chemistry (2006)
Kyung Hee University, Korea

**M.Sc.**, Physical Chemistry (2008)
Thesis: Structures and Electronic Spectra of CdSe–Cys Complexes: Density Functional Theory Study of a Simple Peptide Coated Nanocluster (Advisor: Prof. Dr. Sungyul Lee)
Kyung Hee University, Korea

**Research**

**2008-2016** **Shimon Weiss Group, UCLA.** Single molecule study on Bacterial transcription kinetics in various environments

**2005-2008** **Sungyul Lee Group, Kyung Hee University.** Theoretical study on the role of solvent molecules in the reaction and Quantum chemical analysis for structures and electronic properties of bio-inorganic clusters

**Teaching Experience**

**2008-2013** **Teaching Assistant/Associate**
**Department of Chemistry**
**University of California, Los Angeles**
**Los Angeles, CA**

**Publications**

1. M. Sakin, A. Grunwald, S. Kim, N. Gassman, J. Antelman, Y. Kim, S. Ho, R. Samuel, X. Michalet, A. Gottfried, R. Lin, T. Dertinger, A. Kim, **S. Chung**, R. Colyer, E. Weinhold, S. Weiss, Y. Ebenstein (2014), Towards single-molecule optical mapping of the epigenome, *ACS Nano*, 8(1), 14.

2. S. Kim, A. Gottfried, R. R. Lin, T. Dertinger, A. S. Kim, **S. Chung**, R. A. Colyer, E. Weinhold, S. Weiss, and Y. Ebenstein (2012), Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. *Angew. Chem. Int.*, 51(15), 3578.

3. A. Ingargiola, E. Lerner, **S. Chung**, S. Weiss, X. Michalet, FRETBursts: Open Source Burst Analysis Toolkit for Confocal Single-Molecule FRET (2016), *BioRxiv*

4. C. Liu, **S. Chung**, S. Lee, S. Weiss, and D. Neuhauser (2009), Adsorbate-induced absorption redshift in an organic-inorganic cluster conjugate: Electronic effects of surfactants and organic adsorbates on the lowest excited states of a methanethiol-CdSe conjugate, *J.Chem. Phys*., 131(17), 174705.

5. **S. Chung**, S. Lee, C. Liu, D. Neuhauser (2009), Structures and Electronic Spectra of CdSe–Cys Complexes: Density Functional Theory Study of a Simple Peptide Coated Nanocluster, *J. Phys. Chem. B*, 113(1), 292.

6. H.-S. Kim, D.-S. Ahn, **S. Chung**, S. K. Kim, S. Lee (2007), Tautomerization of Adenine Facilitated by Water: Computational Study of Microsolvation, *J. Phys. Chem A*, 111, 32, 8007.

7. S. Choi, K.-W. Choi, S. K. Kim, **S. Chung**, S. Lee (2006), Vibrational structures of Dimethyl sulfide and Ethylene sulfide Cations Studied by Vacuum-Ultraviolet Mass-Analyzed Threshold Ionization (MATI) spectroscopy, *J. Phys. Chem. A*. 110, 13183.

# Chapter 1:

# Introduction

**Bacterial RNA Polymerase**

RNA Polymerase (RNAP) is the enzyme that carries out transcription in all living organism and its function is to use the nucleotide base sequence information in DNA to synthesize a complementary RNA strand, which is primarily used in the process of protein synthesis called Translation. Although in eukaryotes, three distinct RNAPs (RNA pol I, II, III) transcribe different classes of genes, in prokaryotes, a single RNAP is devoted to transcription of all genes[1]

The core, which is composed of 5 subunits, of the multisubunit RNAPs is conserved in bacteria, archaea, and eukaryotes[1,2]. Among this family, the simplest one is bacterial RNAP with 5 subunits ($\alpha_I$, $\alpha_{II}$, β, β', ω) while eukaryotic RNAP pol I, II, III are composed of 14, 12, and 17 subunits respectively[1,2].

The core enzyme of bacterial RNAP, known as the apo-RNAP complex, cannot properly initiate transcription alone as it requires another factor, σ, to properly orient and position the catalytic core at gene promoters. Association of the core RNAP with σ factor forms the holo RNAP enzyme, which is competent for transcription initiation at specific promoters. Since σ factors are responsible for promoter recognition, the core enzyme would form a complex with one of several different σ factors, each conferring specificity to different promoter DNA sequences[1–4]. In Escherichia coli (*E. coli*) $\sigma^{70}$ is the σ factor responsible for recruiting the RNAP core to the promoters of most house-keeping genes, which are active during normal growth. Structural and sequence analysis of σ factors identified four conserved regions or domains (σR1, σR2, σR3, σR4) connected with flexible linkers[5,6]. The mechanism of σ factor action and function in transcription initiation has been extensively studied and involves a multitude of processes; 1) promoter recognition, 2) un-zipping the double stranded DNA to

enable the template strand to insert into the active site of RNAP, 3) facilitating the stabilization of the short RNA, 4) blocking the RNA exit channel, where a long nascent RNA (separated from RNA-DNA hybrid) releases, and 5) abortive initiation[6,7]. During promoter escape and transition into elongation, σ factor breaks its interaction with promoter DNA and is weakly associated with the elongating core RNAP until it is released from the core RNAP in a stochastic manner. However, sometimes σ factor associated with elongating RNAP stalls the polymerase in a process called promoter-proximal pausing if the elongating RNAP encounters a -10 like sequence before σ factor dissociates[7–9].

The bacterial RNAP holoenzyme has a crab-claw like shape with its β, β' subunits forming the two pincers of the claw. The highly conserved primary (or active-site) channel, which is the cleft between β, β' subunits, accommodates downstream DNA and the RNA-DNA hybrid[10,11]. The active site in the primary channel is where ribonucleotide monophosphate incorporation into a nascent RNA chain occurs along with the release of pyrophosphate from ribonucleotide triphosphate substrates (NTP)[5,10,11]. Two $Mg^{2+}$ ion located in the active site are necessary for RNAP catalytic activity. One $Mg^{2+}$ ion (Mg I) is strongly coordinated by three highly conserved β' active site aspartic residues and plays an important role for catalysis of the phosphodiester bond formation and RNA extension[10,11]. The other $Mg^{2+}$ ion (Mg II) is necessary for proper positioning of an incoming NTP in the active site[11]. Another channel called the "Secondary" channel (10~12 Å) formed at the interface of the β and β' subunits provides a direct conduit for diffusing substrate NTPs to enter the active site[11–13]. Also, abortive product are released through the secondary channel and a class of transcription factors, known as secondary channel binding factors, such as GreA, GreB and DksA regulate the activity of RNAP[11,12,14,15]. Inside the secondary channel, highly conserved structural features in the β'

subunit known as the trigger loop and bridge helix facilitate NTP incorporation and RNAP translocation along the DNA template during the polymerization reaction[13,16].

The α-subunit is comprised of three different domains: well-structured N-terminal domain (αNTD), disordered flexible linker and C-terminal domain (αCTD). Two α subunits dimerize and then the β and β' subunits assemble through their specific interactions with the αNTD of their partner α-subunit. The αCTD of RNAP is oriented in the upstream direction following promoter binding and is capable of binding DNA, such as to the UP element, and interacting with transcription factors, such as catabolite activating protein, to regulate transcription[11,17].

The ω subunit is the smallest subunit with only 91 amino acids (~10kDa) and is associated with the β' subunit in the holo RNAP. Despite being discovered several decades ago, the function of the ω subunit is still ambiguous and controversial. Several researches have showed that the ω subunit is important for maintaining the β' subunit conformation and for regulation during initiation in response to ppGpp in a stress condition[18,19].

**Transcription cycles in Escherichia coli**

Transcription is the first major process of gene expression and is also the most regulated. Transcription in Escherichia coli (*E. coli*) can be described in three separate steps: (I) Initiation, (II) Elongation and (III) termination. Initiation can be divided into three sub-steps[3,4]: (I-i) formation of RNAP holoenzyme following $\sigma^{70}$ association with the core RNAP enzyme; (I-ii) holoenzyme identifying the promoter and binding tightly to two well-defined DNA elements upstream to the transcription start site (TSS), the -10 and -35 hexamer elements in the promoter recognition sequence (PRS); (I-iii) RNAP melting, through successive isomerization steps, the DNA double helix from -10 to +2 to generate the RNAP open complex, with the melted region called the transcription bubble. The separation of template and non-template strands of DNA facilitates the positioning of the TSS in close proximity to the active site, thus preparing the transcription machinery for RNA synthesis. After formation of the transcription bubble, a rate limiting step in formation of catalytically active RNAP, conformational changes stabilize the RNAP-Promoter (RP) complex to form the RNAP-Promoter open complex ($RP_O$), which is ready to begin catalyzing incorporation of nucleoside triphosphates (NTPs) into an RNA transcript; (I-iv) Formation of the first phosphodiester bond changes the RNAP open complex to the RNAP initial transcribing complex ($RP_{ITC}$). As RNA is synthesized by the $RP_{ITC}$, downstream DNA is pulled into the primary channel causing compression of accumulated single stranded DNA in a process called DNA scrunching[20,21]. The $RP_{ITC}$ continues synthesizing RNA products of different lengths and releasing them in a process called abortive initiation (AI), which undergoes multiple RNA synthesis-release cycles until the RNAP clears the promoter. Since most of these cycles are unsuccessful, RNAP aborts initiation after releasing the short nascent RNA and forming the $RP_O$ again for a new synthesis trial[22,23]; (I-v) It is only when enough strain is built in the enzyme that

the fifth step – promoter clearance – ensues. In this step the holoenzyme loses its grip on the PRS ($\sigma^{70}$ breaks contact with the -10 and -35 hexamers) and the blockage of the RNA exit channel is removed, marking the transition of RNAP to an elongation complex.

In AI, the strong interactions between RNAP and the PRS, mediated by the $\sigma^{70}$ subunit, limit the lengths of abortive RNA products. The stronger the promoter is (RP recognition site interaction), the greater amount of time RNAP will spend cycling in AI[24] and the distribution of abortive transcript lengths will be skewed towards longer ones[25]. These findings may be the reason for why, after establishing tight promoter recognition interactions, transcription initiation is much slower than elongation.

In elongation (II), after σR3.2 is removed from the exit channel, RNAP undergoes a global conformational change, $\sigma^{70}$ loses its grip on the PRS, and RNAP processively and rapidly elongates the RNA transcript until it reaches the termination signal[26]. Transcription elongation is highly stable and processive (10~20 bp/sec) process[16]. However, the RNAP elongation complex is weakly associated with $\sigma^{70}$, which can cause transcriptional pausing upon encountering specific sequences[27]. $\sigma^{70}$ is stochastically released from the elongating polymerase[9], but can be recruited back whenever a PRS-like sequence re-occurs[28,29]. Some population of paused RNAP exist in a backtracked state where the RNAP active site is positioned upstream of the RNA 3' end. The single stranded RNA that was hybridized to the template strand becomes lodged in the secondary channel[27], where NTPs enter[5,11–13]. Pausing not mediated by $\sigma^{70}$ is also possible due to backtracking caused either by specific DNA sequences or the incorrect NTP being incorporated into the 3' end of the RNA resulting in misalignment 3' end of the transcript. These backtracked-paused complex could be rescued and made into active elongation complexes by cutting the RNA,

upstream of the 3' RNA end, and regenerating a 3'-hydroxyl correctly positioned in the RNAP active site. RNAP has an intrinsic endonucleolytic activity that is greatly stimulated by transcription elongation factors[12,30,31].

Transcription is terminated in the last stage (III), when RNAP encounters either 1) a termination sequence (intrinsic termination), which generally has a hairpin loop secondary structure followed by U-rich region in RNA transcript, or 2) termination factor, ρ. In termination, the full RNA transcript is released from RNAP. Also, RNAP is dissociated from DNA and could be reusable, with one of the σ factors, for another transcription reaction[32]

**Biological reactions in crowded environments**

Biological reactions in the cellular environment differ physicochemically from those performed in dilute buffer solutions. Inside the cell, many kinds of macromolecules such as lipids, sugars, nucleic acid, and proteins coexist and can arrange in higher order structures like the cytoskeleton [33]. The concentration of those molecules in the cytoplasm is 50~400mg/ml, which is corresponding to a volume occupancy of 5~40%[34].

Such crowded environments undoubtedly affect the physical and chemical properties of biomolecules so that the biological reactions under crowded conditions exhibit different properties from those in the buffer[35]. Two main effects of crowded environments on the biological reactions are viscosity effects, affecting diffusion coefficients, and crowding effects due to the volume exclusion of macromolecules.

The high viscosity of crowded conditions slows down diffusion of reactants as well as the chemical reaction itself. A simple Kramer's theory predicts the rate constant is inversely proportional to the viscosity of the medium[36].

The crowding effects, due to the volume exclusion of highly dense macromolecules, reduce the volume available for reactants in the solution, which results in increasing their local concentrations[35]. In the same manner, the macromolecular crowding could modulate the binding affinities of biomolecules and increase their association rate[35,37,38]. Also, the crowded environment affects the structure of biomolecules. Recent literature have reported that macromolecular crowding generally makes the shape of biomolecules more compact[35,37], stabilizing their native states while destabilizing the denatured state[35,37–39].

**An Overview of this Dissertation**

This dissertation summarizes my efforts to elucidate the bacterial transcription kinetics in initiation and transcription under crowding conditions as a mimic system for the cell environments. Chapter two states the development of *in vitro* quenching based single round transcription assay using single molecule detection and their use for investigation of the effects of size and concentration of crowders on bacterial transcription. The results suggest that crowding agents affect transcription reaction kinetics after $RP_O$ formation through volume exclusion. Chapter three describes the efforts to monitor the kinetics of transcription in initiation by using various ensemble and single molecule techniques. Contrary to expectations, we found that transcription kinetics from later stages of initiation (e.g. from a 7-base transcript) was significantly slower than from earlier stages, and this newly identified pathway in initiation is attributed to the RNAP backtracking.

# Chapter 2:

## *E.coli* RNA polymerase activity under crowding

## Introduction

Gene expression starts from the process called Transcription, where the information in DNA is copied into more functional nucleic acid, RNA by the enzyme RNA polymerase (RNAP). Since the phenotypes of the cells are determined by transcription, it is extremely crucial to understand RNAP-DNA interactions in the transcription processes in order to figure out how cells develop their functionalities and how they survive.

Unfortunately, understanding transcription in the cell is highly challenging since transcription is highly regulated and complicated process involving many different transcription factors[14,40,41].

One of highly abundant transcription factors, DksA, which binds to the secondary channel of RNAP to disrupt RNAP-Promoter DNA interaction. DksA acts primarily to downregulate many genes at the level of transcription initiation in conjunction with guanosine tetraphosphate (ppGpp), which is produced when there insufficient amount of amino acids in the cell by the ribosome, as a feedback mechanism to avoid wasting amino acids on proteins that are not essential for cell function[14,18,40,42,43].

The *E. coli* cellular environment is significantly different from that in dilute buffer since the cell has macromolecules such as nucleic acid, sugar, and protein that make a highly condensed environment[33,44]. The concentration of an individual macromolecule is not so high, but the overall macromolecules take up a huge amount of the volume in the cytoplasm (~40% in Escherichia coli[45]). This compact and dense environment in the cell could alter biological reactions significantly. First, dense cellular environments cause a high viscosity. Increase in viscosity slows down the translational motions of the molecules resulting in slower kinetics. Kramer theory has successfully described the viscosity effects on kinetics of many reactions[36,46]. Many theoretical

and experimental researches have shown that viscosity affect the kinetics of reaction rather than its thermodynamic properties[36,38,47,48].

Secondly, the large volume occupancy of macromolecules in the cell could profoundly affect thermodynamic or/and kinetics of chemical reactions and properties of the biologically important materials such as proteins and nucleic acids. As highly concentrated macromolecules occupy a larger space, the space for other molecules decreases. This is because the space taken up by a macromolecule is inaccessible to other molecules, therefore, the volume occupied by a macromolecule excludes other molecules from its vicinity. Therefore, this inaccessible volume is called the excluded volume of a macromolecule[35,44,49,50]. The excluded volume of a macromolecule function of the shape and size of the macromolecule itself and that of other molecules. The volume exclusion of macromolecules significantly affects the thermodynamics and/or kinetics of reacting molecules especially when the reaction causes the change in the volume of reactants. It was previously shown that volume exclusion of macromolecules in the crowded environment considerably increases the local concentration of reactants, as well as changes in the size of globular proteins, stability of native protein structure, and the folding behavior of RNA[38,39,44].

Recently, several researches have reported that transcription reactions could be affected by macromolecule crowding[51–54]. For example, many studies on Cell free protein expression (CFPE) have shown that transcription reactions by T7 RNA polymerase (T7 RNAP) are markedly enhanced under crowding conditions, specifically when the size of crowders is comparable to that of the RNAP[51]. This enhancement is attributed to increased association between promoter DNA and T7 RNAP[51,52]. However, since most studies on transcription under

Crowding are focused on multiple cycles of either transcription or association of DNA and RNA Polymerase (RNAP) to make RNAP-Promoter open complex (RPo), the effect of the crowded environment on a *single* transcription run is still poorly understood.

In this context, we have developed a novel *in vitro* quenching based single round transcription assay using single-molecule detection. Our assay is based upon hybridization of a probe DNA with RNA, produced by transcription reactions, and whose sequences are controlled by the template DNA sequences we designed. Using this assay, we were able to count the number of RNAs transcribed during a given time by transcription in the presence of various crowding environments, as a model system of the crowded cellular environment, and investigate the size/concentration effects of various osmolytes and macromolecular crowding agents on actively-transcribing RNAP. Our results demonstrate an expected slowdown of transcription kinetics due to increased viscosity, and unexpected enhancement in transcription activity by larger crowding agents at the same viscosity. These findings suggest that crowding agents affect the transcription reaction after RNAP-Promoter open bubble ($RP_O$) formation through volume exclusion.

**Results**

Single round transcription experiments starting from RPo, in the presence of various crowding conditions, were performed using an *in vitro* quenching based single molecule FRET assay quantifying transcripts to determine the efficiency of transcription. Our assay utilizes single strand DNA (ssDNA) probes labeled with donor and acceptor dyes targeting RNA transcription products through hybridization. The number of hybridized and unhybridized probe to transcripts is accurately detected and counted using alternating-laser excitation (ALEX)-based fluorescence-aided molecule sorting (ALEX-FAMS)[1,2]. As ALEX-FAMS is based on single molecule Förster Resonance Energy Transfer (FRET)[3], identification of sub-populations with different conformations (*i.e.* different FRET efficiencies) is possible. Analysis of the number of single molecule events for the hybridized and unhybridized probe populations, each with a particular FRET efficiency distribution, gives the transcription efficiency during a given time (900 seconds, in this study). Since the concentration of probe and RPo is identical in all measurements, a greater fraction of hybridized probe indicates higher number of transcripts, that is, higher efficiency of transcription per a given time.[1,2,4].

The design of the *in vitro* quenching based single round transcription assay is described below and depicted in Fig. 1. The DNA template used for the transcription reaction was designed to produce a transcript containing as sequence complementary to the probe sequence (Fig. 2). Thus, an RNA produced by the transcription reaction would be efficiently hybridized with the ssDNA probe.

Our transcription detection probe is ssDNA, doubly labeled with a donor-acceptor FRET pair - a donor (TMR) at its 5' end and an acceptor (Alexa Fluor 647) at its 3' end. In order for efficient

hybridization with transcripts, we designed the ssDNA probe with poly dT (20 dT) to be unstructured (without having a stable secondary structure) in the solution[5]. Since the persistence length of unstructured ssDNA is small (~1.5 nm in 2 M NaCl, ~3 nm in 25 mM NaCl[6]), ssDNA probe with 20nt (> 6nm), and yields a single FRET population with peak FRET efficiency of ~0.75. However, when hybridized to the run-off mRNA transcripts, the probe becomes stretched and much more rigid due to the long persistence length of dsDNA-RNA hybrid (for dsDNA ~50nm[7], for dsRNA ~64nm[8]), and gives rise to another FRET sub-population with a lower peak FRET efficiency of ~0.3.

Experiments were performed with low concentration of FRET probe (~100 pM), which is the concentration range needed for ALEX-FAMS. At lower concentration than RNAP-Promoter binding affinity ($K_d$ = 1 ~ 10nM at physiological ionic strength[9]), only single round run-off transcription reactions occur on a given template due to the unlikely event of re-association of RNAP and template following RNAP dissociation from the template. By detection of RNA from single round run-off transcription, we can accurately quantify the transcription efficiency, starting from $RP_O$, during a given time. By comparing the transcription activity with respect to the concentrations/types of osmolytes, we could investigate the physical and chemical effects of various osmolytes on the transcription reaction.

In order to validate the assay suitable for detection of transcripts per given time, two important controls were done: the first ensuring the probe should hybridize specifically with RNA produced by the expected transcription reaction (No non-specific binding), the second ensuring that the reaction is stopped immediately by a quencher (fast/efficient quenching). Transcription reactions, done by addition of NTPs to RPo, that were measured by ALEX FAMS exhibit a shift in the

population of high FRET to low FRET (indicating probe hybridization to transcript). Only a high FRET population was observed for the probe alone (Fig 3A) or transcription reactions excluding NTP (Fig 3B), confirming that the probe only binds to transcribed RNA with the target sequence. The transcription reaction was stopped or quenched by addition of $MgCl_2$ to a final concentration of 500 mM, due to RNAP-Promoter association decreasing drastically with increasing ionic strength[9]. Fig. 4 shows that simultaneous addition of $MgCl_2$ and NTP results in no low FRET shift, suggesting the reaction was completely quenched. To address the concern that addition of $MgCl_2$ does not quench the transcription reaction, but instead slow down the reaction, measurements were done following longer incubation times after adding $MgCl_2$ in the reaction mixture. If 500 mM $MgCl_2$ only slows down the reaction, any progression in the reaction (more low FRET population) would be observed after long incubation even in the presence of $MgCl_2$. As shown in Fig. 4, incubations up to 3 hours following NTP and $MgCl_2$ addition to RPo demonstrated no FRET population shifting toward low FRET. Therefore, confirming that 500 mM $MgCl_2$ works as a fast and efficient quencher that completely stops the transcription reaction from proceeding.

With the *in vitro* quenching based single round transcription assays, transcription reactions kinetics following $RP_O$ formation was tested as a function of various crowding conditions. All transcription reactions were incubated for 15 minutes under various size and concentrations of osmolytes (See Methods and Material). Our results demonstrate that, regardless of the size of osmolytes, transcription efficiency by *E. coli* RNAP during a given time decreases as the concentrations of osmolytes increase. This is not surprising as the reaction kinetics is heavily affected by viscosity. According to the simple Kramer theory, the viscosity slows down the kinetics of reactions and has inverse relationship with a kinetic constant[10–14].

Is the viscosity the only factor that affects the transcription reaction under crowding? If so, then the same decrease in transcription efficiency should occur at a given viscosity for all osmolytes. The small osmolytes Ethylene glycol (EG), Polyethylene glycol 200 (PEG 200), and PEG 400 display a similar inhibitory effect not significantly different from that of Glycerol, a well-known viscogen. However, for large osmolytes such as PEG600 (~ 8000), the inhibitory effect on the reaction is noticeably smaller compared to that of small osmolytes. Therefore, we observed that the degree of inhibition decreases with increasing size of osmolytes at a given viscosity (Fig. 5).

To gain a greater understanding of the effect of osmolyte properties on the transcription reaction, the transcription efficiencies was investigated as a function of volume occupancies of osmolytes. The transcription efficiencies for all osmolytes generally decrease as the volume occupancies of osmolytes increase. In addition, transcription with large osmolytes also showed strong anti-correlation between the size of osmolytes and the degree of inhibition at a given volume occupancy of osmolytes, while small osmolytes do not show any distinct size effect on the transcription efficiency, as shown in Fig. 6.

**Discussion and Conclusions**

We have developed a novel *in vitro* quenching based single round transcription assay to investigate the transcription reactions by *E. coli* RNAP in the presence of various osmolytes. Since crowders were added following formation of RPo, the effect of crowders on RNAP-Promoter DNA association to form $RP_O$ was removed, which has been reported to be a step heavily influenced by the effect of crowders. Therefore, the effect of the size/concentration of osmolytes/crowders on transcription reactions following $RP_O$ formation was investigated.

Our results successfully demonstrate that 1) macromolecular crowding affects transcription reactions even after $RP_O$ formation, 2) small size osmolytes such as EG, PEG 200, and PEG 400 only act as viscogens, just like Glycerol, which increase the viscosity of the medium, 3) large osmolytes (PEG 600 ~ 8000) behave not only as viscogens, but also as macromolecular crowding agents that affect the transcription reaction by volume exclusion, and the effect depends the size of crowders.

How do crowders affect the transcription, will only kinetics be affected? As we conducted transcription assays only for a 900s time point, we don't have any experimental evidence to prove whether crowding acts solely on transcription kinetics. However, since many researches have reported that viscosity affects kinetics rather than thermodynamics[10–13], and small osmolytes, acting only as viscogens, showed the biggest inhibitory effect, we expect the effect of crowders mostly on the kinetics of transcription reactions (unless small osmolytes interacts with reactants directly) and crowding effect enhances transcription kinetics, and somewhat cancels out the opposite effect caused by viscosity. In other words, large osmolytes show a smaller inhibitory effect compared to small osmolytes due to the opposing action between crowding effect and

viscosity effect, which increase and decrease the kinetics respectively, while small osmolytes only have viscosity effect.

Which step of transcription kinetics will it be affected the most? According to the transcription kinetics model[15–18], promoter escape is the slowest and rate limiting step after formation of $RP_O$. In this regards, we anticipate that the transcriptional reaction enhancement, compared to the one, by small osmolytes, by larger osmolytes/crowders might be attributed to larger volume occupancies that alter the conformation of RNAP-Promoter initial transcribing complexes ($RP_{ITC}$), in early transcription stages, resulting in a reduction in the activation barrier for promoter escape.

## Methods and Materials

### 1. Preparation of promoter DNA:

90 nt template strand and non-template strand of *lac*CONS-20dA-4bp promoter designed in Fig. 2 were chemically synthesized (ordered from IDT, Coralville, IA, USA). Both strands are hybridized in the hybridization buffer (40 mM Tris-HCl, pH 8, 200 mM Magnesium Chloride ($MgCl_2$)) with a thermal cycler (increase temperature to 95C˚then gradually and slowly decrease temperature down to 21 C˚). Since our transcription assay uses hybridization to detect RNA transcripts, it is crucial to make sure that only hybridized double stranded promoter DNAs exist in the promoter DNA solution (Single stranded DNA leftover hinders the correct quantification of RNA transcripts: excess template strands would hybridize with transcripts, and excess non-template strands would hybridize with our probes).

In order to assure no single stranded DNA in the promoter DNA solution, we confirmed with ALEX-FAMS that only high FRET populations were observed for the mixture of promoter DNA solution and the probe having complementary sequence to single stranded DNAs before preparing RNAP-Promoter open complexes.

**2. Estimation of volume occupancies of the osmolytes:**

Volume occupancies of crowders/osmolytes were calculated by the equation below using hydrodynamic radius, which were obtained from Table 1.

$$\emptyset_C = \frac{4\pi(R_H)^3 N_A\, C}{3M_w}$$

where $\boldsymbol{R_H}$ is a hydrodynamic radius of an osmolyte, $\boldsymbol{N_A}$ is Avogadro's number, and **C** is a concentration of an osmolyte (w/v)

### 3. Preparation of a stable open complex, $RP_{ITC=2}$:

$RP_O$ solution is prepared with 3 μL *E. coli* RNAP holoenzyme (NEB, Ipswich, MA, USA, M0551S; 1.6 μM), 10 μL 2X 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 Bovine Serum Albumin (BSA), pH 7), 1 μL of *lac*CONS-20A-4bp promoter and 6 μL of water. $RP_O$ is then incubated in solution at 37˚C for 30 minutes. To remove nonspecifically-bound RNAP, 1 μL of 100 mg/mL Heparin-Sepharose CL-6B beads (GE Healthcare, Little Chalfont, Buckinghamshire, UK) is added to $RP_O$ solution together with 10 μL of pre-warmed 1X transcription buffer. The mixture is incubated for 1 minute at 37˚C and centrifuged for at least 45 seconds at 6000 rpm. 20 μL of the supernatant containing $RP_O$ formed on *lac*CONS-20A-4bp promoters are transferred into a new tube for an extra incubation with 1.5 μL of 10 mM Adenylyl(3′-5′) adenosine or Adenylyl(3′-5′)uridine ($A_pA$, Ribomed, Carlsbad, CA, USA) at 37˚ C for 20 minutes, respectively, to form $RP_{ITC=2}$ solutions. These $RP_{ITC=2}$ solution s are used as stock for all transcription reactions. 2 μL of RNAse inhibitor (NEB, Ipswich, MA, USA, M0314S) are added into the $RP_{ITC=2}$ solution to prevent degradation of newly synthesized RNA molecules.

## 4. *In vitro* quenching based single round transcription assays

In order to produce run-off RNA transcripts one has to add all four types of ribonucleotide triphosphates (NTPs). In all of our transcription reactions we have used 100 μM of high purity NTPs (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

After quenching a transcription reaction, 100pM of ssDNA FRET probe is added and incubated with the quenched reaction mixture for 30 minutes at room temperature. The quenched-probed reaction mixture are then used for μsALEX measurements (Fig. 1).

The stability of RNAP-Promoter DNA complexes changes as a function of conditions. Therefore, the concentration of the RNAP-DNA complexes is calibrated beforehand to yield a dynamic range of low FRET population-fraction between 0 and 0.9 with each change in conditions. Because the concentration of promoter DNA remains unchanged after the treatment with Heparin Sepharose beads, based on the promoter DNA concentration, we estimate that less than 1 nM of RNAP-DNA complexes were used for all transcription assays.

For all data for *in vitro* quenching based transcription assays were acquired for a duration of 10-15 minutes using a setup described already in Panzeri et. al.[69] with the Perkin Elmer SPADs and 532 and 638 nm CW lasers operating at powers of 170 and 80 μW, respectively.

The results shown in Fig. 5, 6 are all averages of repeated measurements. The error bars reported these figures represents the standard deviation of averages of measurement repeats

**5. µsALEX analysis for transcription assays:**

A dual channel burst search (DCBS; intersection of bursts from donor excitation burst search and acceptor excitation burst search)[70] was implemented with m=10 and F=6 (Each burst is identified according to a criteria of that consecutive m(=10) photons detected with a photon count rate, which is F(=6) times higher than the background rate[71,72]), in order to isolate the FRET-only sub-population for further analysis. After DCBS, we further selected the bursts that have more than 25 photons during donor excitation period ($N_{DA} + N_{DD} \geq 25$) and more than 25 photons from the acceptor channel during acceptor excitation period ($N_{AA} \geq 25$).

Following, the FRET efficiency value (E) and the Stoichiometry value (S) for each burst are tabulated in a 2D scatter plot of E vs. S for all burst events. The populations associated with the free ssDNA probe (high FRET efficiency) and the hybridized probe (low FRET efficiency), are identified in the 2D E-S scatter plot. 1D-FRET histograms are extracted from the 2D E-S scatter plot (as 1D projection on the E axis). All 1D-FRET histograms from transcription assays are fitted to a sum of two Gaussians with the constraints for their means and standard deviations.

All data analysis including burst search, burst selection, and 1D-FRET histogram fitting in this work have been done by using Python-based open source burst analysis toolkit for confocal single-molecule FRET, FRETBursts[73].

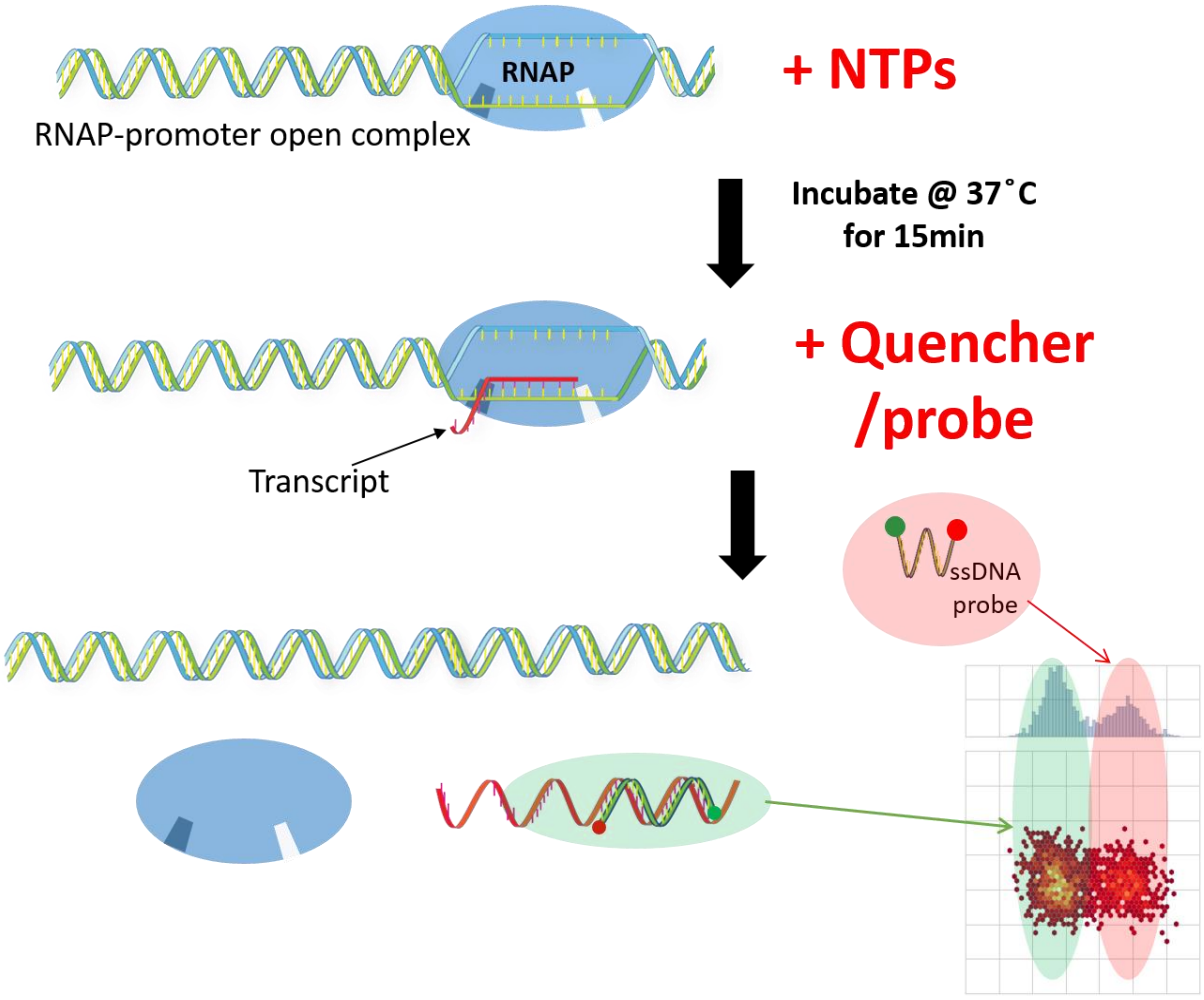



**Figure 1. A schematic of the *in vitro* single round quenching based transcription assay.** Transcription from RNAP-Promoter open complex ($RP_O$) is started with addition of NTPs, and stopped by a reaction quencher after 15 minutes incubation at 37˚C. Then, RNA transcripts produced by single-round transcription reactions for 15 min are hybridized with the ssDNA probe that has the complementary sequence to the part of the transcript.

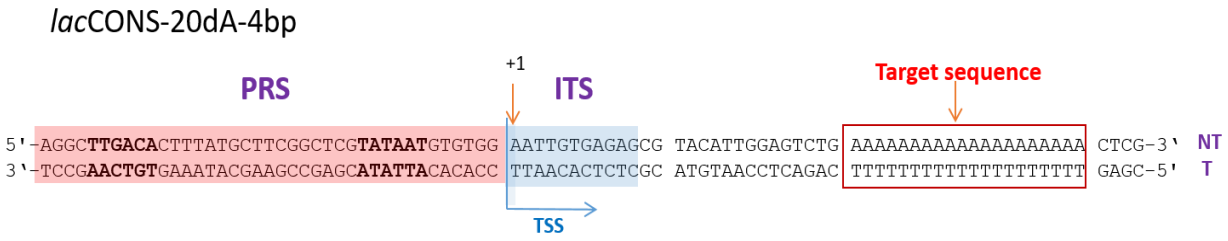



**Figure. 2: *lac*CONS-20dA-4bp promoter sequence used in Chapter 2.** Highlighted are the Promoter Recognition Sequence (PRS), the Transcription Start Site (TSS), the Initially Transcribed Sequence (ITS), and the elongation sequence. The elongation sequence contains 20 consecutive T's (20dT) that transcribes into RNA containing 20 consecutive A's (20dA). This 20dA in RNA is detected by the doubly labeled ssDNA FRET probe.

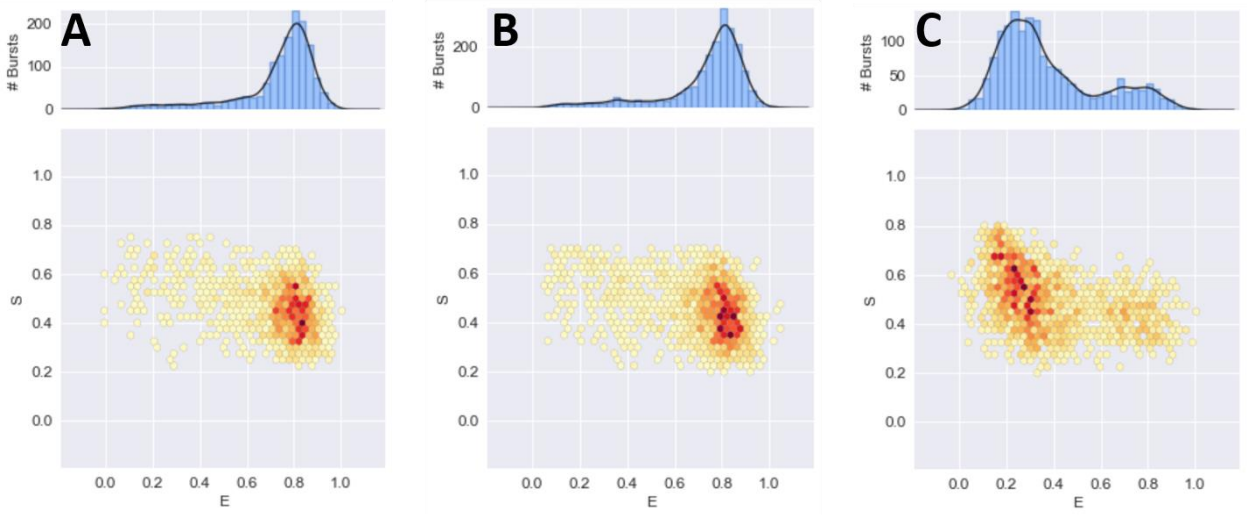



**Figure. 3: Validation for detection of transcripts using the *in vitro* single round quenching based transcription assay.** The high FRET population shifts toward low FRET only when probes hybridize with RNA produced by transcription reactions. A) no RNAP negative control; the solution containing ssDNA probe, promoter DNA, and NTP was incubated for 15 min at 37℃, then quenched by 500 mM $MgCl_2$. B) no NTP negative control; the solution containing ssDNA probe, RNAP-Promoter open complexes was incubated for 15 min at 37˚C, then quenched by 500 mM $MgCl_2$. C) Positive control; the solution contains ssDNA probe, RNAP-Promoter open complexes, NTPs was incubated for 15 min at 37˚C, then quenched by 500 mM $MgCl_2$.

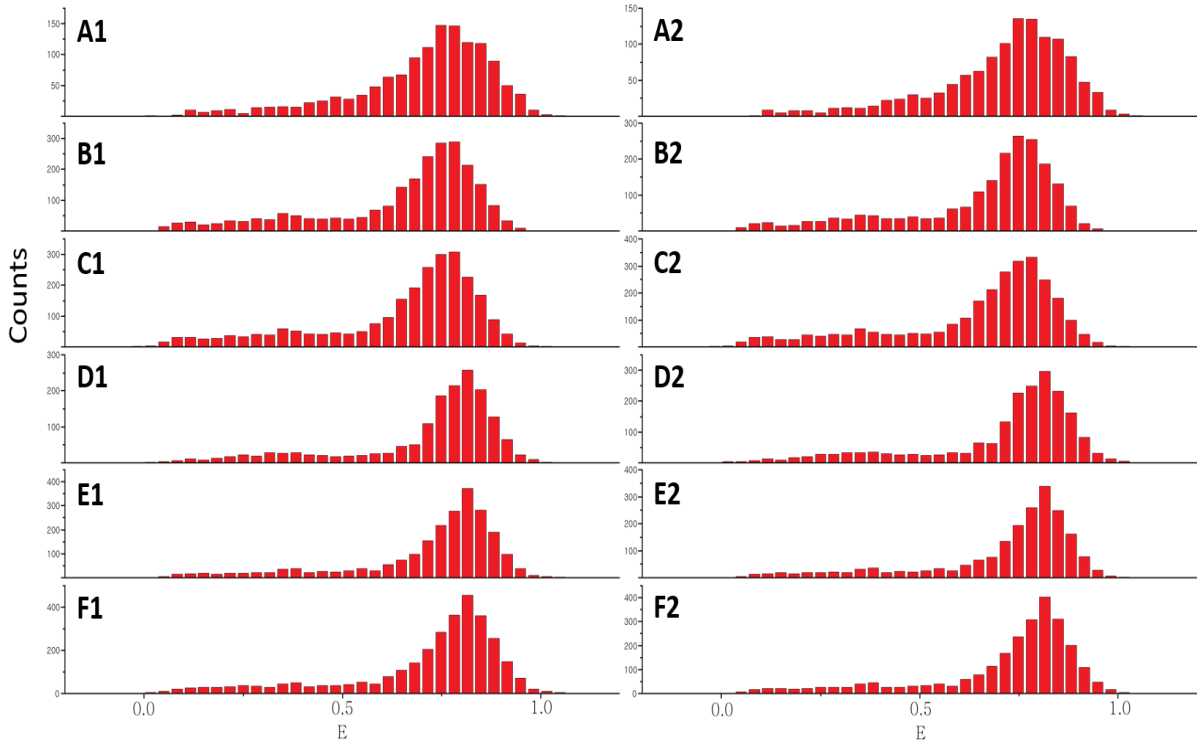



**Figure. 4: 500 mM $MgCl_2$ as a quencher for the *in vitro* single round quenching based transcription assay under various crowding environments.** A-G) Transcription reactions were immediately quenched by 500 mM $MgCl_2$ after started by addition of NTPs (0 min incubation). A1-2) transcription mixture **in buffer** was measured right after quenching (A1), or measure 3 hours after quenching (A2). B1-2) transcription mixture **in 25% Glycerol** was measured right after quenching (B1), or measure 3 hours after quenching (B2). C1-2) transcription mixture **in 30% PEG 400** was measured right after quenching (C1), or measure 3 hours after quenching (C2). D1-2) transcription mixture **in 20% PEG 1000** was measured right after quenching (D1), or measure 3 hours after quenching (D2). E1-2) transcription mixture **in 15% PEG 3350** was measured right after quenching (E1), or measure 3 hours after

quenching (E2). F1-2) transcription mixture **in 15% PEG 8000** was measured right after quenching (F1), or measure 3 hours after quenching (F2).

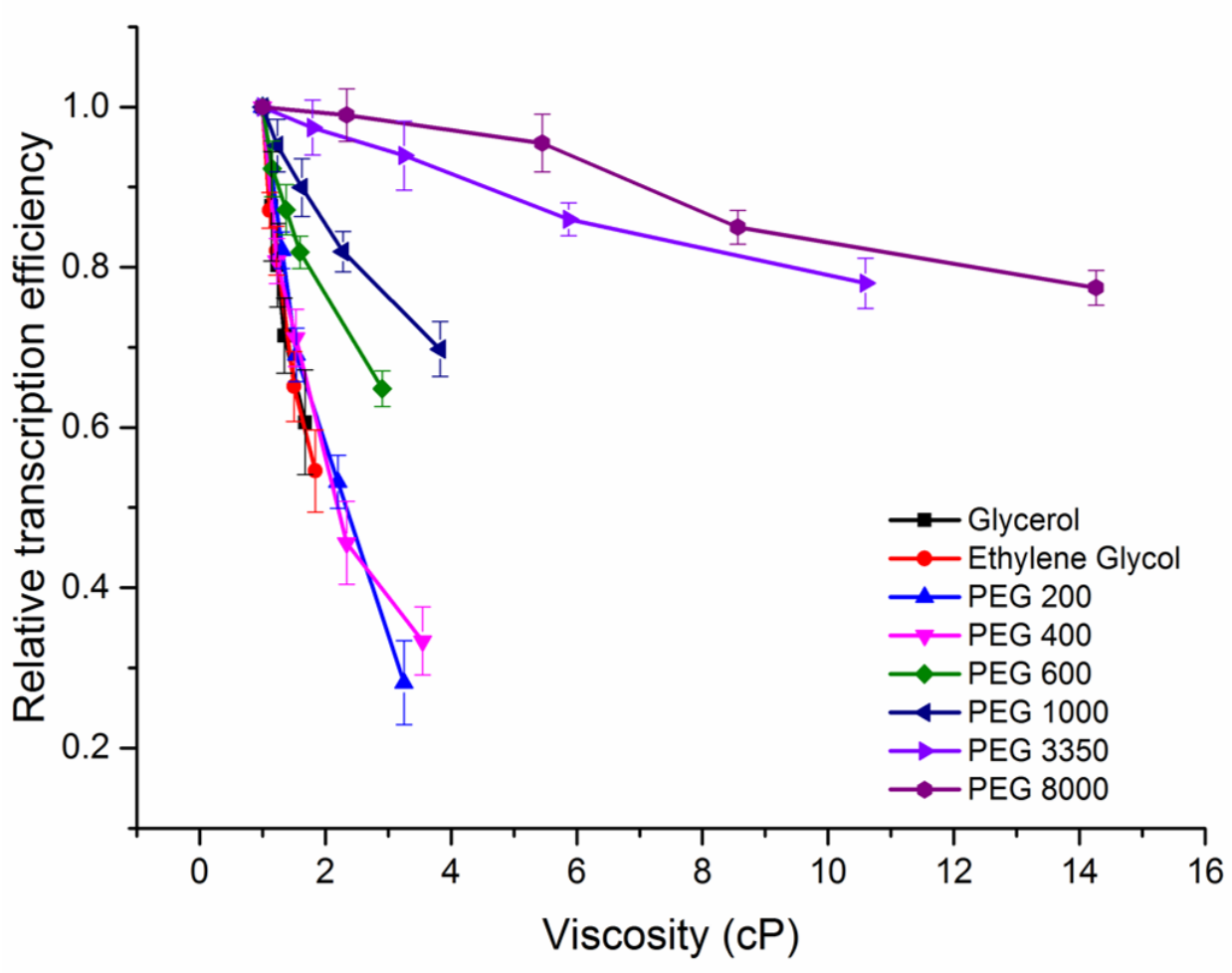



**Figure. 5: Relative transcription efficiencies (in reference to the transcription efficiency in the buffer per 900s) at a given time point (900s) plotted against viscosity of the solution with various crowding conditions.** Small osmolytes (smaller than PEG 600) show the same inhibitory effect (no significant difference) on transcription reactions with Glycerol, which only act as a viscogen. On the other hand, for large osmolytes (PEG 600 ~ 8000), decrease in transcription efficiency is much lesser than the ones for all small osmolytes at a given viscosity and this inhibitory effect depends on the size of osmolytes.

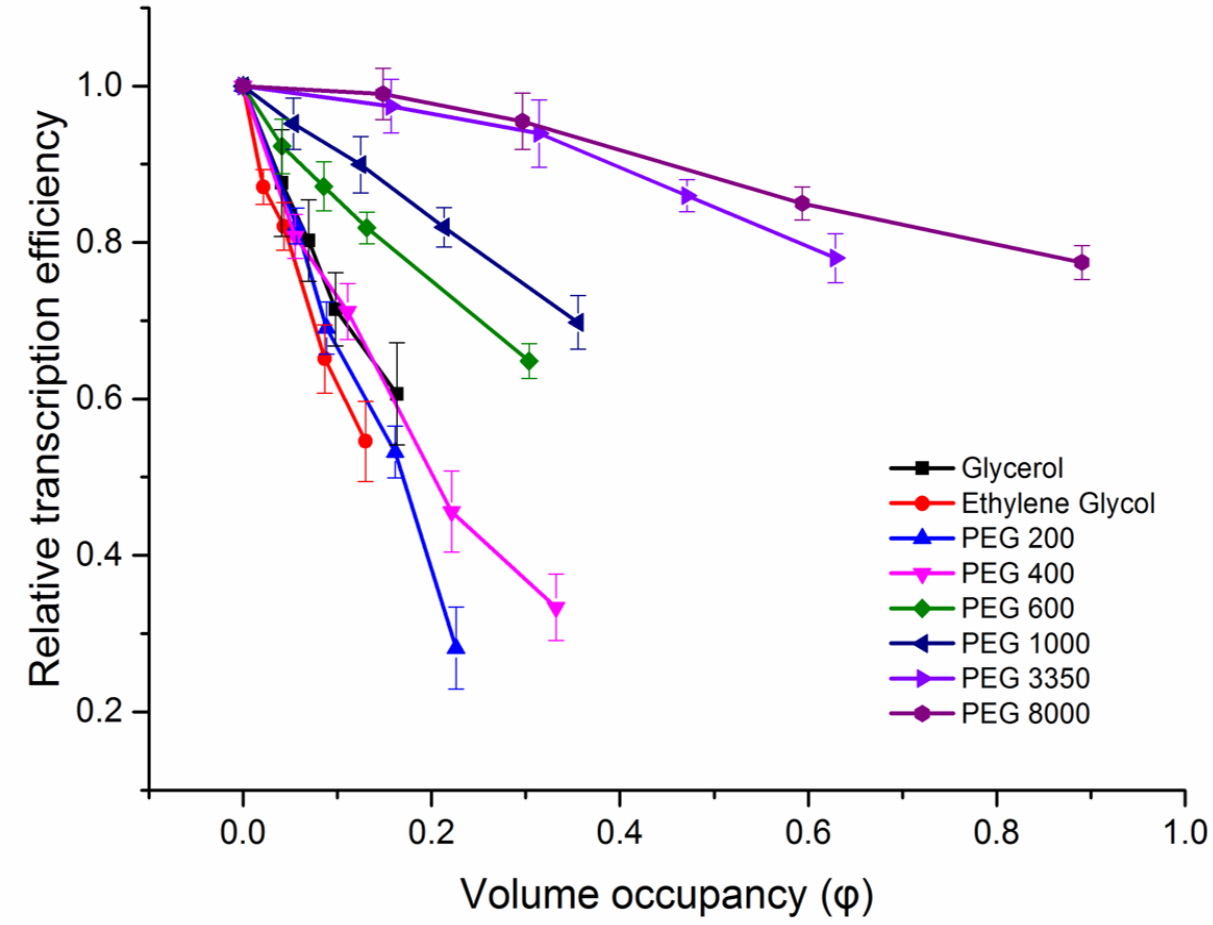



**Figure. 6: Relative transcription efficiency (in reference to the transcription efficiency in the buffer per 900s) at a given time point (900s) plotted against volume occupancies of the various osmolytes in the solutions.** For all osmolytes, transcription efficiency per a given time (900s) decreases as the volume occupancy of the osmolytes increase. Large osmolytes (PEG 600 ~ 8000) show lesser inhibitory effect than small osmolytes at a given volume occupancy and this effect decreases as the size of osmolytes increases. All small osmolytes show no significant difference in the inhibitory effect on transcription reactions.

**Table1. Hydrodynamic radii of osmolytes used in Chapter 2**

| Molecular Crowders | Average Molecular Weight (g/mol) | Hydrodynamic radius (Å) |
|---|---|---|
| Glycerol | 92.09 | 3.1 |
| Ethylene Glycol | 62.07 | 2.2 |
| PEG 200 | 200 | 4.0 |
| PEG 400 | 400 | 5.6 |
| PEG 600 | 600 | 6.9 |
| PEG 1000 | 1000 | 8.9 |
| PEG 3350 | 3350 | 16.1 |
| PEG 8000 | 8000 | 26.6 |

All data were obtained from references (ref.[66,67] for EG and PEGs, ref.[68] for Glycerol)

# Chapter 3:

# Pausing in Escherichia coli transcription initiation

**Introduction:**

Transcription is an essential and highly regulated process in gene expression[3,4]. Transcription of house-keeping gene in Escherichia coli **(***E.coli*) is composed of three main kinetics steps: Initiation, Elongation, and Termination[3,4,32]. After RNA Polymerase (RNAP) holoenzyme binds with the promoter DNA, undergoing several isomerizations, RNAP-Promoter DNA complex becomes a polymerization-competent open complex (RPo)[4]. Then, in the presence of nucleoside triphosphates (NTPs), the RNAP-Promoter initial transcribing complex ($RP_{ITC}$) engages in Abortive Initiation (AI), a process in which RNAP cycles between synthesis and release of short RNA transcripts[20,21,25]. Once enough strain is accumulated in the $RP_{ITC}$ complex, then RNAP holoenzyme can escape from the promoter recognition sequence (PRS) region and transit into elongation[4,20–22]. In this step the holoenzyme loses its grip on the PRS ($\sigma^{70}$ loses contact with the -10 and -35 hexamers), the blockage of the RNA exit channel is relieved and become an efficient, resistant, and processive elongation complex[4,11]. Transcription is terminated, when RNAP encounters a termination sequence or a termination factor at which the full RNA transcript is released and RNAP dissociates from DNA[32]. We note that in some genes, during elongation, RNAP encounters specific sequences that cause transcriptional pausing, sometimes together with backtracking due to 3' end of the RNA chain intruding into the secondary channel. Pausing and un-pausing in elongation constitute additional steps[8,28,29].

While the biochemical steps of RNA polymerization are most probably the same both in initiation and in elongation, their rates are far slower in initiation due to the strong RNAP-PRS interactions (absent in elongation) and due to the blockage of the RNA exit channel by the acidic tip of $\sigma^{70}$ region 3.2 ($\sigma$R3.2)[5,9,11,74]. These biochemical steps include NTP entrance and insertion, phosphodiester bond formation, pyrophosphate removal, and translocation (in order to allocate

space for the next NTP to be incorporated). RNA sequences of up to 9 nucleotides (nt's) are stabilized by hybridization to the melted template DNA in a RNA:DNA hybrid[75–77]. Although RNAP can accommodate a RNA molecule of up to 14 nt's long (9 in RNA:DNA hybrid + 5 in RNA exit channel)[78], the transition into elongation, whereby the RNA exit channel blockage by σR3.2 is relieved, only begins when RNA reaches the length of 11 nt's[22,79,80].

In what ways can RNAP overcome the barrier between initiation and elongation? Experimental data has accumulated to suggest that in AI, RNAP scrunches downstream DNA into its active site[20,21] increasing the size of the transcription bubble up to 25 bases.

Because σ70 holds the -10 and -35 hexamers tightly, the DNA region upstream to the transcription bubble (upstream of position -12) stays as a duplex, while up to 19 bases of DNA in the bubble (non-template DNA bases -3 to +12; template DNA bases -7 to +12) may become compressed and/or bulged-out, which adds into a strain buildup and stabilizes what is known as the 'stressed intermediate'[81–83]. Simultaneously, the longer the initially transcribed nascent RNA is, the closer its negatively charged 5' end will get to the acidic tip of σR3.2. At the same time, it will achieve more contacts with the template DNA in the RNA:DNA hybrid, which will, in turn, counteract the strain from the compressed bubble DNA and from the repulsion from σR3.2[84]. Abortive transcripts are held in RNAP until reaching a certain length, above which the strain is relieved either through RNA dissociation and release (abortion), or by pushing σR3.2 and hence unblocking the exit channel (transitioning into elongation)[25,84].

This description, however, does not agree with the abortive transcript release being the rate-limiting step in the context of AI[21,85]. Additionally, the opening or closing of the active site trigger loop, which serves as a controller for translocation, is likely to control the strain release in a step-

wise fashion, allowing the nascent RNA to be stabilized in a newly occupied register. In order to provide an answer to this puzzle, we decided to study the kinetics and mechanism of AI using a model promoter and RNAP from *E. coli*, with the hope that such measurements would shed new light on why abortive transcript release is the rate-limiting step in AI.

The kinetic experiments described below utilized the combination of NTP starvation and promoter sequence design in the initially transcribed sequence (ITS) in order to study intermediates in transcription initiation. From the standpoint of the nucleotide addition cycle, reaching an $RP_{ITC}$ state through NTP starvation means that RNAP experiences long periods of time whereby it waits while being introduced with only the wrong NTPs. Similar situations in elongation lead to stabilization of RNA in the pre-translocated state[86–88] eventually causing the nascent RNA transcript to backtrack on its way to be abortively released by $RP_{ITC}$[89].

Many *in vitro* transcription initiation experiments harnessing NTP starvation were performed under equilibrium conditions after long incubation times of all components. Such experiments assumed that AI behaves according to a sequential model that is described as a continuous, steady-state process characterized by sequential transitions between states. Is AI the only major process characterizing initiation or is it just one process among other alternative ones? To address this question, we have developed a novel *in vitro* solution-based "single-run" quenched kinetics transcription assay to measure kinetics into- and out- of NTP-starved states, whether in initiation or in elongation, with the *E. coli* wild type transcription components. We identified deviations from the sequential model in the initiation that highlight the existence of a previously undetected -pathway state, serving as a kinetic trap in initiation. To better understand this uncharacterized state in initiation, we measured transcription kinetics in the absence and presence of GreA, a

transcription elongation factor that catalyzes the cleavage of RNA backtracked into the secondary channel[90–95].

Additionally, we performed single molecule magnetic tweezers steady-state measurements to monitor temporal trajectories of single RNAP-Promoter DNA complexes in initiation. Lastly, we also performed quenched kinetics assays with $^{32}$P-radiolabeled NTPs using denaturing poly acryl amide gel electrophoresis (PAGE) as a control for our assay. As for full transcript detection, we were not able to achieve single-run transcription conditions and high enough signal to background ratio. However, focusing on abortive transcripts, we were able to study AI under NTP starvation conditions.

Altogether, our findings indicate that a previously uncharacterized RNAP state in transcription initiation exists, and is attributed to RNAP backtracking. In addition, we demonstrate that this paused-backtracked state can be reached even when all four NTPs are present, and that the larger the imbalance in the concentration of all four NTPs is, the higher is the flux into this stable backtracked state in transcription initiation.

**Results**

Because AI is the rate-limiting step in transcription after open bubble formation[85], measuring the production rate of run-off RNA transcripts under NTP starvation (or low and non-equimolar NTP concentrations) would report on the AI kinetics. We therefore extended our idea of an *in vitro* quenching based transcription assay to develop the quenched kinetics assay for transcription initiation kinetics.

The design of the quenched kinetics assay is described below and depicted in Fig. 1. The promoter portion of the DNA template strand was designed to include a sequence of 20 dT in its downstream elongation part (Fig. 2) so that the run-off transcript, that includes a 20A sequence, could be detected by hybridization with a 20dT ssDNA probe. The ssDNA probe is doubly labeled with a donor-acceptor FRET pair - a donor (Tetramethylrhodamine – TMR) at its 5' end and an acceptor (Alexa Fluor 647) at its 3' end.

For accurate measurement of kinetic data the assay requires: (i) formation of a stable initial state; (ii) addition of NTPs at zero time point; (iii) rapid quenching of the reaction at pre-determined times; (iv) efficient hybridization of the ssDNA FRET probe to transcripts; (v) prevention of RNA degradation. We designed the initially transcribed sequence (ITS) part of *lac*CONS[96,97] or T5N25[21] promoters so that after addition of a partial set of NTPs, RNAP will be able to transcribe abortive transcripts of maximal lengths of 4 and 7, or 6 and 11 nucleotides (Fig. 2).

To stabilize open transcription bubble in $RP_{ITC=2}$ formed at *lac*CONS or T5N25 promoters, the corresponding open promoter complexes were supplemented with initiating dinucleotides $A_pA$ and $A_pU$, respectively (Fig. 2). In order to stabilize a particular $RP_{ITC}$ state ($RP_{ITC\leq i}$, $i \in \{4,6,7\}$)

or an elongation complex ($EC_{i=11}$), $RP_{ITC=2}$ is incubated for a given time, $t_{entrance}$, with a partial set of NTPs.

The all four NTPs are then added and the system is incubated for another time period, $t_{exit}$. The transcription reactions are then quenched by addition of 0.5 M Guanidinium Chloride (GndCl) that has been shown to serve both as a successful transcription reaction quencher (Fig. 3) and as an enhancer of probe hybridization to RNA (Fig. 4). After quenching the transcription reaction, 20dT ssDNA FRET probe is added to the mix and hybridized with the target run-off transcript. Finally, FAMS-ALEX measurements are performed and the number of run-off transcripts per time point is determined through the fraction of ssDNA FRET probe that is hybridized (Fig. 1).

The exit kinetics from the NTP-starved state was measured by keeping a constant $t_{entrance}$ and by varying $t_{exit}$ incubation times ('Exit Kinetics'; Fig. 6). The kinetics of entrance into NTP-starved states was measured indirectly by varying $t_{entrance}$, while keeping the $t_{exit}$ incubation constant ('Entrance Kinetics'; Fig. 6). The results of 'exit kinetics' from NTP-starved states using the *lac*CONS promoter are described in Fig. 7.

In order to assess the contribution of elongation kinetics on our quenched kinetics assay, we compared exit kinetics from $RP_{ITC=2}$ to a transcription bubble closure kinetics measured by a single molecule FRET (smFRET) assay using multi-spot excitation[98,99]. Fig. 5 shows that kinetics from $RP_{ITC=2}$ measured by both the quenched kinetics assay and the bubble closure assay were almost identical. We therefore conclude that for the *lac*CONS promoter, elongation is much faster than its initiation kinetics (*i.e.* kinetics for promoter escape) so that our quenched kinetics assay represents AI kinetics.

If the kinetics of transcription follows a sequential model, one would expect exit kinetics resumed from a late state in initiation, and certainly from an early state in elongation, to be faster (or at least with similar speed) than exit kinetics resumed from an earlier transcription initiation stage. This is indeed the case, the exit kinetics out of early elongation state, $RD_{E=11}$ is slightly faster than the exit out of stable open complex state, $RP_{ITC=2}$ (Fig. 7, black and magenta).

Surprisingly, and contrary to the sequential kinetics model, we found that the exit kinetics out of NTP-starved $RP_{ITC \leq i}$, ($i \in \{4, 6, 7\}$) states are significantly slower than both exits out of $RP_{ITC=2}$ and $RD_{E=11}$ (Fig. 7). Moreover, the longer the nascent RNA transcript in initiation is, the slower is the exit kinetics (Fig. 7). These results were obtained using two *lac*CONS promoter sequences with the same PRS and elongation sequences but with different ITS (Fig. 2). Our observations suggest that due to NTP starvation, the initiation complex (RNAP + promoter DNA + nascent RNA) is stabilized in a non-catalytic state (kinetic trap) through an unknown mechanism, causing the exit kinetics to slow down or to be delayed.

In order to quantify these results, exit kinetics data were analyzed using a simplified transcription model that includes the non-catalytic state in initiation (Fig. 17). The best global fit results using this model are presented as solid curves in Fig. 7; the best fit values are reported in Table 1. The model and the fitting results support the existence of an off-pathway $RP_{ITC \leq i}$ state with an additional off-on rate for transitioning back into the AI on-pathway. This newly identified state in transcription initiation is likely to be responsible for delays in the production of run-off RNA transcript. The fitting to the model also suggest that the population of non-catalytic states increases with increasing RNA chain length.

Entrance kinetics (see Fig. 6 for definition) into NTP-starved states have shown the opposite trend, that is, the longer the maximal length the nascent RNA in $RP_{ITC}$ is, the faster it takes to enter into the NTP-starved non-catalytic states (Fig. 8), as long as the NTP-starved states are in initiation rather than in elongation. Entrance kinetics into the early elongation state $RD_{E=11}$ showed no delay compared to $RP_{ITC=2}$, indicating no state that acts as a kinetic trap in $RD_{E=11,}$ as seen in initiation (under conditions when complexes are incubated with NTPs for 10 min). Entrance kinetics are fit with exponential curves and the best fit results are presented as solid curves in Fig. 8. The best fit values are reported in Table 2.

Altogether, the kinetics results of entrance into- and exit- out of NTP-starved initiation states suggest the presence of states not described by the sequential model in AI that reduce the kinetics of promoter escape. The dependence of the exit kinetics delay and the entrance kinetics acceleration on transcript length in AI suggests that at least to some extent, RNAP-Promoter DNA complex with a long RNA (e.g. 7-mer) is stabilized in this state in initiation that is not competent for transcription. Rescue from this state (by introduction of all four NTPs) is associated with a time delay for transiting back into productive initiation on-pathway and transition to elongation.

Inspired by a similar, well-studied pausing and subsequent backtracking mechanism in transcription elongation[91,100], we tested whether the observed non-catalytic state in AI is the manifestation of the stabilization of RNA-Promoter complex in a backtracked state. The elongation factor GreA is known to function through the secondary channel of RNAP as a catalyzer of the intrinsic endonucleolytic activity of RNAP to cleave 2 or 3 nucleotides (nt's) of 3'-end backtracked RNA[31,93,94,100]. We performed exit kinetics measurements exiting from $RP_{ITC=2}$ and $RP_{ITC\leq7}$ in the absence and presence of 1 μM GreA (Fig. 9). While, as expected, GreA has shown

only a marginal effect on the exit from $RP_{ITC=2}$, its effect on the exit from $RP_{ITC\leq7}$ is more pronounced – the kinetics is accelerated and closer to that of exit out of $RP_{ITC=2}$ in the absence of GreA (Fig. 9).

Best-fit results using our simplified transcription model (Fig. 17) show how GreA decreases the initial population of- and further excursion into the off-pathway state (Table 1). We therefore conclude that GreA releases RNAP from this non-catalytic state in transcription initiation. When the next-to-be-incorporated NTP is missing, RNA polymerization cannot proceed properly. The outcome of NTP starvation is that the transcribed RNA transitions into an alternative pathway that involves some backtracking followed by pausing, similar to NTP-starvation conditions in elongation[86–88,101].

Another way to study transcription initiation effects is through analysis of $^{32}$P-radiolabeled transcripts performed by denaturing PAGE. Attempting to compare run-off transcript bands was difficult due to very low signal to background ratio (S/B). In order to achieve an acceptable S/B for run-off bands, we had to run the gel assays at RNAP-Promoter complex concentration of [$\geq$ 7 nM] which is about one order of magnitude higher than the concentration used for the quenched kinetics assay based on smFRET. At this concentration range, however, run-off transcription production kinetics did not reach a clear plateau at long times, suggesting that single-round transcription conditions were not met (Fig. 19).

The bands of the short abortive transcripts, however, had excellent S/B due to the repetitive abortive cycle. This allowed us to follow the abortive transcript production kinetics at low RNAP concentrations (~1.7 nM). The abortive transcription kinetics with and without NTP starvation was compared against that with starvation in the presence and absence of GreA (Fig.10).

Starting from $RP_{ITC=2}$ state, NTP starvation (lack of ATP and CTP) leads to $RP_{ITC\leq7}$ state. The bands are enumerated by the most probable abortive product sizes knowing that a $^{32}P$ radiolabeled UTP ('hot' UTP) has been used. A high cold:hot UTP ratio allows most of the observed bands of short abortive products to have a single 'hot' U. Since a G nucleotide adds an extra bulginess in such gel-based assays, the 5- and 6-mer abortive products (AAUUG(U)) most probably occupy the same gel band and the 4-mer (AAUU) and 7-mer (AAUUGUG) bands are well-separated from this combined product band. Moreover, by accounting to the intensity of the bands, the kinetics of the individual abortive product bands shows a doubled intensity for the band assigned as AAUUG(U), in comparison to the other bands, suggesting it is indeed comprised of a mixture of AAUUG and AAUUGU (Fig. 10). This band assignment is supported by assessment using NTP analog 3' deoxy nucleotides (3'dNTPs, see Fig. 19).

Interestingly, in NTP starvation conditions, the 7-mer band appears, while when all four NTPs are present, this transcript does not appear (left panel in Fig. 10). A similar description arises in starvation conditions in the presence of GreA (right panel in Fig. 10). This result suggests that in $RP_{ITC\leq7}$, where all abortive products up to a 7-mer are possible, it is the 7-mer that experiences 3'-end cleavage catalyzed by GreA (the cleaved base is a 'cold' G), hence it is the 7-mer nascent RNA that backtracks by a single nucleotide and pauses.

The kinetics of AI when all four NTPs are present shows bands with intensities much lower than in NTP starvation conditions (left panel in Fig. 10). This is expected because with all NTPs, eventually RNAP irreversibly goes into elongation, in which abortive products are not produced anymore. In the presence of GreA, the same lower intensity comparison to the NTP-starved state (right panel in Fig. 10). Overall, GreA in NTP starvation conditions leads to an AI pattern similar

to that in normal conditions. This may be explained by the fact abortive release occurs through the secondary channel, therefore GreA may catalyze their cleavage so the intensity of the bands of abortive products that escape from the endonucleolytic activity is as small as in normal conditions when all NTPs are present. In Fig. 10, abortive transcript production does not cease, therefore transcription pausing is temporary.

To the best of our knowledge, our findings are the first to suggest and support the existence of backtracking followed by pausing in the context of transcription initiation. Although the elongation factors GreA, GreB and TFIIS have already been reported to be associated with RNAP in initiation in different prokaryotic and eukaryotic systems[14,102–104], and to enhance the promoter escape[41,105,106], the exact mechanism responsible for this enhancement in the context of transcription initiation remains unclear. Our results point to an additional role in initiation for GreA, namely helping with the cleavage of backtracked parts of nascent RNA in the secondary channel, enhancing transcriptional competency under NTP starvation. Further examination of exit kinetics out of $RP_{ITC \leq 7}$ state in the absence and presence of GreA can shed light on the underlying mechanism of backtracking and pausing in $RP_{ITC \leq i}$ state.

Is the observed long-lived backtracked state an artifact solely of NTP starvation or can it also occur under normal conditions when all NTPs are present? Depending on the availability of the next NTP to be incorporated, the nascent RNA could either translocate or backtrack. When all NTPs are present, the flux towards the backtracked-paused state is likely to be small. When the next NTP to be incorporated is absent, this flux is maximized and can be identified through kinetic experiments. Another way to increase the flux towards the long-lived backtracked state is by maintaining the ratios of NTPs in such way that the subset of NTPs that lead to a given $RP_{ITC \leq i}$

will be abundant while the subset of NTPs that allow rescue out of this initiation state will be less abundant.

The effect of partial starvation of NTPs through non-equimolar concentrations (UTP and GTP at 100 μM, ATP and CTP at 2 μM) is shown in Fig. 11. Exiting out of $RP_{ITC=2}$ after incubation with a non-equimolar NTP bias is slower than the same kinetics when all NTP's are at a concentration of 100 μM, but faster than rescue kinetics after full starvation. The best-fit results for this experiment using the simplified transcription model (Fig. 17) are reported in Table 1. They quantitatively show that the initial population of the non-catalytic initiation state depends on the relative abundance of the 'missing' NTPs. Similar results (with non-equimolar NTP concentrations) were obtained for the T5N25 promoter (Fig. 12). The delay in exit kinetics occurs not only for full NTP starvation, but also for the more biologically relevant condition of NTPs concentration imbalance, typical for cells under metabolic stresses (such as nutrients deficiencies).

We conclude that the transition into the backtracked paused state depends on the concentration of the NTPs required for exiting from initiation into elongation. If this interpretation is correct, pausing is expected to occur in initiation with low probability even when all NTPs are present at high and equimolar concentrations. Our quenched kinetics assay is an ensemble assay (although the read-out is done by ALEX-FAMS) and as such, it is not suited for detecting rare events. By following the behavior of individual transcription complexes it is possible to capture such possible rare events.

For this reason, we performed magnetic tweezers transcription assays on dsDNA clones that contain the *lac*CONS or T5N25 promoter sequences. DNA molecules were immobilized to a coverslip on one end and attached to a magnetic bead on the other. The magnetic bead was rotated

in order to supercoil the DNA and form plectonemes. Once RNAP is introduced to the chamber, it binds the promoter and opens-up the transcription bubble. This, in turn, changes the number of plectonemes, resulting in a change in the bead height/DNA extension (See Methods and Materials). Once NTPs are added, additional changes to the transcription bubble size are reflected as additional changes in DNA extension. Due to the fact that different stages have own distinguishable transcription bubble sizes, monitoring the magnetic bead height allows us to identify these state in DNA extension time-trajectory (Fig. 13A & Fig. 14A). And indeed, using this assay's sensitivity to the bubble size, it was possible to show that transcription initiation occurs via a DNA scrunching mechanism[20,21].

The results of magnetic tweezers transcription assays for *lac*CONS promoter in the absence and presence of GreA (with equimolar NTPs at 100 μM) are shown in Fig. 13. A dsDNA that contains the *lac*CONS promoter sequence is tethered to a coverslip from one end and is attached to a magnetic bead from another end. The DNA molecule is then maintained in a supercoiled configuration with applying magnetic field on the bead. The height of the bead in this state is then recorded (Fig. 13A).

In the absence of GreA, we observed short- and long-lived $RP_{ITC}$ states (Figs. 13B, 13D, 13F & 13H). The lifetimes spent in $RP_{ITC}$ states are summarized in a histogram fitted with a double exponential in which 90% of events (n=216) were short-lived ($\tau$ = 300±40 s, Standard error of the mean (SEM)), and 10% were long-lived ($\tau$ = 2600±700 s SEM; Fig. 13H, blue). Correlating these data with DNA bubble sizes (representing distinct $RP_{ITC}$ states; Figs. 13A, 13D & 13F) revealed that 90% of events were characterized by a 15 ± 3 bp transcription bubble (Standard deviation (SD)). The remaining 10% of events were characterized by a more homogeneous 10 ± 1 bp

transcription bubble (SD). In agreement, correlative analyses (Fig. 13D) indicated $RP_{ITC}$ states with a shorter bubble (<12 bases) can be long-lived (~2200 ± 350 s SEM, n=37; Fig. 13D oval) whereas $RP_{ITC}$ states with a larger bubble (>12 bases) were generally shorter-lived (~1100 ± 110 s SEM, n=179).

These data suggested that, in addition to the well-characterized $RP_{ITC}$ state (Fig. 13A), a subset of RNAP complexes entered a distinct, long-lived state characterized by a smaller transcription bubble (denoted $RP^*_{ITC}$ in Fig. 13A, 13B & 13C). We hypothesized that this long-lived initiation intermediate represented backtracked RNAP previously characterized 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) became uniformly short-lived ($\tau$ = 350 ± 30 s, SEM) in the presence of GreA, with transcription bubbles of larger sizes (16 ± 2 bases, SEM); Figs. 13C, 13E, 13G & 13H).

Experiments completed on a different promoter (T5N25) template showed the same trend (Fig. 14). Without GreA, the bubble size in most events is small, while with GreA there is an increase in the number of events associated with larger bubble size as well as a net reduction in lifetime (Fig. 14D & 14E, quadrants). Collectively, these data further support the existence of a long-lived backtracked state during initiation and reveal a role for GreA in preventing RNAP from entering this state. In addition, the long-lived state (due to the backtracking) during initiation is further enhanced under conditions of non-equimolar NTP concentrations, similarly to the single run quenched kinetics assay observations (Fig. 15).

In summary, the results presented so far suggest the presence of a newly identified state in transcription initiation induced by backtracking and temporal pausing. Under normal conditions (high equimolar NTPs concentrations) this state is rarely populated. However, under partial or

full NTP starvation, the population of this off-pathway state can be significantly increased. This novel initiation pathway state can revert back to the productive RNA polymerization pathway, after replenishment of the missing NTPs, following a delay by minutes.

**Discussion and Conclusions**

Many of the NTP addition cycle steps in bacterial transcription elongation are biochemically similar to the equivalent steps in initiation, but the rates are different, owing to the structural differences between these states[9,23,77]. These structural differences are mainly caused by $\sigma^{70}$ factor[9,23,29]. In initiation, $\sigma^{70}$ is tightly bound to the -10 and -35 promoter hexamers through regions σR2 and σR4 respectively[5,6,107]. σR3 that connects between σR2 and σR4 blocks the RNA exit channel in the open complex $RP_O$ and in early $RP_{ITC}$ states[5,84]. In particular, the loop in region σR3.2 enters deep into the channel towards the active site, where its highly negatively charged tip is positioned against the highly negatively charged 5'-end of the nascent RNA[5,7]. This element is considered as a switch that has to be pushed out in order to relieve $\sigma^{70}$'s tight promoter binding leading to the opening of the RNA exit channel[7,26,84].

Transcription initiation has been considered to be rate limiting owing to the many abortive cycles RNAP undergoes until it is able to transit into elongation[22,81,85]. In abortive initiation, the nascent RNA is transcribed to a certain length, and then it dissociates and released through the secondary channel[13,108], the channel through which NTPs are provided to the catalytic site[109].

The nascent RNA is held through hybridization to the template DNA strand in the transcription bubble. At an early initiation stage only few RNA:DNA bp's stabilize the nascent RNA and it is therefore prone to spontaneous dissociation. At a late initiation stage, the nascent RNA includes additional hybridized bp's, but at the same time, the negatively charged 5'-end of the growing nascent RNA is pushed against the negatively charged tip of σR3.2[7,84]. In addition, because the -10 promoter hexamer is held tightly by σR2, the DNA bubble upstream to TSS, which is not involved in the hybridization to nascent RNA, gets compressed[5]. These two forces

counteract the lengthening of the nascent RNA and the system reaches a balance of forces. The result of this delicate force balance is either the release of the accumulated strain through pushing σR3.2, relieving $\sigma^{70}$'s tight promoter binding, and transitioning into elongation[26], or through back-translocation of the nascent RNA. When the nascent chain is long enough, backtracking leads to the insertion of its 3'-end into the secondary channel[13]. When the next NTP to be incorporated is present, the trigger loop in the β' RNAP subunit stabilizes the nascent RNA in the post-translocated state[86,87,101,110], thereby counter-stabilizing the nascent RNA against strain release in the backwards direction. Under NTP starvation, the translocational equilibrium is shifted towards the pre-translocated state[85,86].

In this work, we were able to modulate this balance of forces by temporarily (or permanently) removing the next NTP (to be incorporated) in transcription initiation. As a result of nascent RNA stabilization in the pre-translocated state, the balance shifts towards releasing the strain in the backwards direction resulting in the backtracking of the nascent RNA. When supplementing the starved complex with the missing NTP, this backtracked complex can be rescued and recovered back into active transcription initiation; however the response to the addition of all NTPs is quite slow compared to the overall transcriptional kinetics. We can therefore treat the backtracked state as a reversible kinetic trap.

Do NTPs actively invoke the excursion back to the initiation fast pathway? We can find an answer to this question in studies performed on the transcription elongation complex. In certain paused states, the 3' portion of the nascent RNA is backtracked into the secondary channel, and the bridge helix of the β' RNAP subunit adopts a bent conformation that prevent NTPs from entering into the active site[12,13]. We therefore suggest that the escape from the kinetic trap is not

dependent upon the added NTPs. If this is the case, what could be a possible mechanism for this escape? In elongation, there are three possible fates for the stable single-nucleotide backtracked-paused complex. First possibility is rescue of the backtracked RNAP-DNA complex by cleavage of RNA (via the intrinsic RNAP endonucleolytic activity[12]) making it transcriptionally competent again. Second possibility is the reversal of a single-nucleotide backtracking state through forward translocation. The last possibility is to slowly backtrack further and further until finally entering into a long-lived, backtracked, paused/arrested state [12].

In initiation, however, these three possibilities are somewhat different because the lengths of transcripts are much shorter than the ones in transcription elongation complex. One possibility (which is similar to the first transcription elongation complex case) is that a 3'-terminal portion of backtracked RNA in the secondary channel gets cleaved through the intrinsic endonucleolytic activity of RNAP. The complex then becomes capable of re-extending the 5'-terminal RNA portion that was left in the active site. This process is slow in neutral pH[111], but could be accelerated by GreA[95].

Indeed, our measurements show that the presence of GreA does accelerate the exit kinetics and therefore support this possibility. Another possibility is that a backtracked complex transitions back to the open complex through successive backtracking steps, leading to a slow abortive release.

Based on our findings and the above discussion, we propose a modified transcription initiation model (Fig. 16). The model suggests that initiation is rate limiting to transcription not only due to the multiple abortive cycles which occur before the transition into elongation, but also due to a slowed-down pathway triggered by a backtracked, temporarily paused, state. We hypothesize the presence of two AI pathways. After nascent RNA backtracking, RNAP can either swiftly release

it as an abortive product through the common AI pathway, or the backtracked complex can be stabilized through a change (possibly conformational[12,112]) in RNAP.

This possible change enhances the enzyme's intrinsic cleavage activity of the part of nascent RNA that extrudes into the secondary channel[113]. The backtracked-paused complex may also slowly release the nascent RNA through further backtracking. The model depicted in Fig. 16 adds an additional 'slow' pathway (red arrows and beige rectangle background) to the conventional AI model.

What is the biological relevance of our findings, if any? Is pausing in initiation an artifact of *in vitro* starvation? Or alternatively, is pausing in initiation relevant to gene expression in the live cell?

In our experiments, we have been able to identify pausing and backtracking in initiation through the utilization of ITS design and *in vitro* NTP-starvation. By complete starvation of 'missing' NTPs, the transition into this kinetic trap was maximized. In the biological realm, however, there is never a total lack of one (or more) of the NTPs. Rather, what is more common is a concentration reduction in one (or more) of NTPs due to metabolic stresses (such as, for example, nutritional shortage[73]). Imbalance in NTP concentrations may also occur in the absence of stress for some bacteria at specific growth phases[114].

When a non-equimolar mixture of all four NTPs is provided, the complex can either incorporate the next NTP, or enter into the (slow) non-catalytic state. This branching decision partially depends on the abundance of the next (low abundance) NTP. The lower the relative abundance of the next NTP (to be incorporated), the longer the complex will wait, and therefore the larger its probability to backtrack and pause. In addition, the longer the initiation complex

spends in a stalled position, the higher the probability that it will incorporate a wrong NTP (i.e. misincorporation). This in turn will drive the complex to backtrack and pause as part of proofreading. In this context, we performed experiments where all four NTPs are present, but at a concentration imbalance (non-equimolar). Indeed, even in this case a delay in exit kinetics was observed, but to a lesser extent. Magnetic tweezers experiments under equimolar NTPs concentrations show non-negligible number of backtracked-paused states (in the absence of GreA). These findings suggest that backtracking and pausing in initiation may be biologically relevant.

It has already been established that transcription initiation could be stalled *in vivo*[24], that a large portion of RNAP complexes reside most of the time at promoter sites[89,115] and that AI occurs also *in vivo*[95]. These observations could possibly be explained by backtracking and pausing in initiation. If indeed pausing in initiation also occurs *in viv*o, this pausing in initiation may be important factors in control of gene expression. In particular, if such a mechanism occurs, as in our measurements, when recovering from stressful conditions (supplying all NTPs after a period of starvation), the transition to elongation in some genes may be delayed while others may immediately start, depending on the ITS of the gene.

In summary, we identified a previously uncharacterized pathway in initiation that involves pausing and backtracking in which the nascent RNA chain is stabilized in the RNAP-Promoter complex with its 3' end extruding into the secondary channel. The transition into the backtracked-paused initiation state depends on the time the complex waits for the next NTP in a given ITC state. Based on our findings, we propose a modified model for *E. coli* transcription initiation in which a backtracked state is an additional regulatory intermediate.

## Methods and Materials

### 1. Preparation of a stable open complex, $RP_{ITC=2}$:

$RP_O$ solution is prepared with 3 μL *E. coli* RNAP holoenzyme (NEB, Ipswich, MA, USA, M0551S; 1.6 μM), 10 μL 2X 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 Bovine Serum Albumin (BSA), pH 7), 1 μL of lacCONS+20A promoter (sequence in Fig. 2) and 6 μL of water. $RP_O$ is then incubated in solution at $37^0$C for 30 minutes. To remove nonspecifically-bound RNAP, 1 μL of 100 mg/mL Heparin-Sepharose CL-6B beads (GE Healthcare, Little Chalfont, Buckinghamshire, UK) is added to $RP_O$ solution together with 10 μL of pre-warmed 1X transcription buffer. The mixture is incubated for 1 minute at $37^0$C and centrifuged for at least 45 seconds at 6000 rpm. 20 μL of the supernatant containing $RP_O$ formed on lacCONS or T5N25 promoters (sequences in Fig. 2) are transferred into a new tube for an extra incubation with 1.5 μL of 10 mM Adenylyl(3′-5′) adenosine or Adenylyl(3′-5′)uridine ($A_pA$ or $A_pU$; Ribomed, Carlsbad, CA, USA) at $37^0$C for 20 minutes, respectively, to form $RP_{ITC=2}$ solutions. These $RP_{ITC=2}$ solution s are used as stock for all transcription reactions. 2 μL of RNAse inhibitor (NEB, Ipswich, MA, USA, M0314S) are added into the $RP_{ITC=2}$ solution to prevent degradation of newly synthesized RNA molecules.

## 2. Transcription Quenched kinetics:

To produce run-off transcripts, high-purity ribonucleotide triphosphates (NTPs) (GE Healthcare, Little Chalfont, Buckinghamshire, UK) 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 the coding region of the nontemplate strand of the promoter used (*lac*CONS and T5N25, see in Fig. 2) and on nucleotide position at position +3 relative to the transcription start site (TSS). The presence of $A_pA$ (in lacCONS) or $A_pU$ (in T5N25) in $RP_{ITC=2}$ stabilized RNA up to position (+2), but also prevented transcription initiation from an unwanted site[116,117].

The NTP starvation schemes depend on the different initially transcribed sequences (ITS) being used (Fig. 1 & Fig. 2). To exit from the initiation/elongation NTP-starved state the reaction mixture was complemented with all four NTPs.

For kinetics, the reaction mixture is incubated with the partial set of NTPs for a constant duration of 40 minutes at $37^0$C. All four NTPs are then added to the reaction mixture and incubated for different durations (which make up the samples for the different time points in the kinetics) at $37^0$C, at which 0.5M Guanidinium Chloride (GndCl) is added to quench the reaction. Subsequently, a ssDNA FRET probe is added and hybridizes with the target run-off transcript (see Fig. 2 for probe target sequence).

To confirm that reaction kinetics are not affected by changes in pH, we measured the pH of the solution before and after quenching and found that it did not deviate much from the pH 7 of the buffer used (6.8 – 7.0). For transcription kinetics experiments with GreA, 1 μM of protein factor is added to transcription complexes in NTP-starved initiation or elongation states. For Exit

Kinetics, the reaction mixture is incubated with the partial set of NTPs for a constant entrance time of 45 minutes at $37^0$C. The missing NTPs are then added to the reaction mixture and incubated for different exit times at $37^0$C, at which 0.5M Guanidinium Chloride (GndCl) is added to quench the reaction. For Entrance Kinetics, the reaction mixture is incubated with the partial set of NTPs for different entrance times at $37^0$C. The missing NTPs are then added to the reaction mixture and incubated for a constant exit time of 10 minutes at $37^0$C. Finally, the reaction is quenched by adding 0.5M GndCl.

After quenching a transcription reaction with 0.5 M GndCl, 100 pM of ssDNA FRET probe was added and incubated with the quenched reaction mixture for an additional 20 minutes at room temperature. The quenched-probed reaction mixtures were then used for μsALEX measurements.

After treatment with Heparin-Sepharose beads, the exact RNAP-Promoter complex concentration is unknown, because some of the complexes had RNAP bound non-specifically to the promoter DNA. In addition, the activity of RNAP, dictated by the fraction of RNAP-Promoter complexes that yield a full transcript, changes depending on various conditions. Therefore, the concentration of the DNA-RNAP complexes is calibrated beforehand to yield a dynamic range of low FRET population-fraction between 0 and 0.9 with each change in conditions. Because the concentration of promoter DNA remains unchanged after the treatment with Heparin Sepharose beads, based on the promoter DNA concentration (1 nM), we estimate that less than 1 nM of RNAP-DNA complexes are used for all *in vitro* single-round quenched kinetics assays.

The last 20 bp's of the template strand coding region of all promoter sequences contain a 20 consecutive A's (20dA, Fig. 2). The sequence of the complementary ssDNA FRET probe is 20 dT and it is doubly labeled with a pair of fluorophores suitable for smFRET, a donor,

Tetramethylrhodamine at its 5' end (5' TAMRA modification) and an acceptor, Alexa Fluor 647 at its 3' (3' Alexa Fluor 647 modification; ordered from IDT, Coralville, IA, USA)[97].

Each time point in the quenched kinetics assay is measured for a duration of 10-15 minutes using a setup described already in Panzeri et. al.[69] using the Perkin Elmer SPADs and 532 and 638 nm CW lasers operating at powers of 170 and 80 μW, respectively.

Each kinetic measurement was performed at least in duplicates using different preparations obtained on different days. For each batch, we made sure that:

1. The FRET probe in the presence of $RP_{ITC=2}$ without NTPs yielded only a high FRET population (negative control).
2. The kinetic trace reaches a hybridized fraction (low FRET sub-population) of 0.90±0.05.
3. After a long incubation of $RP_{ITC=2}$ with all 4 NTPs (20 minutes), the fraction of hybridized probe reaches 0.90±0.05 (positive control). This control is performed daily on the same batch used to prepare NTP-starved RNAP states.
4. After a very long incubation time (typically several hours) of a sample with a quenched reaction, the measurement yielded the same low FRET population-fraction (quenching does work)

The negative control measurement yields a single high FRET efficiency population and serves as the "t=0" time point. The positive control measurement results in a 90±% of the bursts belonging to the low FRET efficiency sub-population and serves as the asymptotic kinetic value at very long times, "t=∞".

The difference between exit kinetics from $RP_{ITC=2}$ and exit kinetics from an NTP-starved state, both prepared from the same batch, is solely the time at which all four NTPs were added. All other experimental conditions (concentrations, temperature, etc.) are identical per batch. Therefore, 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 two kinetics from $RP_{ITC=2}$ and from NTP-starved states (within 5% error). Therefore, the abovementioned positive control served as a proof that experimental conditions (e.g. NTP-starvation) did not alter the relative activity.

The results shown in Figs. 7 – 9 & Fig 11-12 are all averages of such repeated measurements. The error bars reported in these figures are the standard deviation of the repeated measurements. The values at the end of the kinetic trace of repeated measurements were very close to 0.9 in all repeats. In order to compare kinetics starting from different states, however, we had to normalize all kinetic traces so that all of them end exactly at 0.9. An example for quenched kinetics repeats (before averaging) is shown in Fig. 20.

All quenched kinetics data were globally fit to a simplified model that allows the run-off production kinetics to go either directly from an on-pathway initiation state (*On*) to elongation (*RunOff*) or to start at an off-pathway state (the backtracked and paused state; *Off*) and then go to elongation through slow recovery to the on-pathway initiation state as in the following schematics:

(1) $$\boldsymbol{Off} \underset{k_{On\rightarrow Off}}{\overset{k_{Off\rightarrow On}}{\rightleftharpoons}} \boldsymbol{On} \xrightarrow{k_{init\rightarrow elon}} \boldsymbol{RunOff}$$

The data was globally fitted to the model assuming the on-pathway state is $RP_{ITC=2}$, hence kinetics starting from $RP_{ITC=2}$ is fitted with the model assuming that at t=0 all molecules are occupying the on-pathway state, while kinetics starting from $RP_{ITC\leq4,6,7}$ is fitted assuming at t=0 all molecules are occupying either the on-pathway or the off-pathway initiation states.

The results of the global fit are shown on the figures as continuous lines and the best fit values (the rate constants and the on-pathway population at t=0) are reported in Table 1.

### 3. µsALEX analysis:

Dual-color fluorescence photon-timestamps from freely diffusing molecules are recorded using a ALEX-FAMS set-up[55,56]. Fluorescence bursts are identified in the recorded stream of photon-timestamps, and the number of photons in a burst and the burst start/stop times are tabulated. Each burst is identified using an sliding-window burst search that looks for consecutive m(=10) photons exhibiting a count rate higher than a given threshold parameter F(=6) times the background rate[71,72].

The background rate is estimated as a function of time (typically over time-durations of 30s) via maximum likelihood fitting of the inter-photon delays distribution. This assures that slow changes in the background rate are accounted for. In single-molecule µsALEX analysis, three streams of photons are analyzed: donor and acceptor fluorescence photons during green laser excitation (noted here as DD and DA, respectively), and acceptor photons during red laser excitation (noted here as AA). Burst photon counts in each of these photon streams, are background-corrected by subtracting the burst duration times the background rate. First, an all-photon (all streams) burst search is applied. After filtering for bursts with sizes larger than 25 photons, the proximity ratio and the stoichiometry are calculated for each burst to identify the sub-population of bursts where both donor and acceptor are active (FRET sub-population), sub-population of donor-only fluorescence bursts (DO), and sub-population of acceptor-only fluorescence bursts (AO)[118].

Next, correction factors for donor fluorescence leakage into the acceptor detection channel (lk) and the factor that accounts for acceptors directly excited by the green laser (dir) are calculated (lk=~0.07 and dir=~0.04 for measurements presented here).

Next, a dual channel burst search (DCBS; intersection of bursts from green excitation burst search and red excitation burst search)[70] is performed using m=10 and F=6, in order to isolate the FRET-only sub-population for further analysis.

Next, the acceptor fluorescence during green laser excitation stream is corrected for lk and dir[83]. The γ correction factor is calculated by the following procedure: a burst size filter is applied on the sum of the corrected streams coming from green and red excitation of 25 photons each.

Following, the FRET efficiency value (E) and the Stoichiometry value (S) for each burst are tabulated in a 2D scatter plot of E vs. S for all burst events. The populations associated with the free ssDNA probe (high FRET efficiency) and the hybridized probe (low FRET efficiency), are identified in the 2D E-S scatter plot and each sub-population is fitted to a 2D Gaussian.

The γ factor is then calculated using a linear fit to the 1/<S> vs. <E> data produced from the two populations[118]. In these measurements reported here, we extracted a γ factor value of 0.61. After all correction factors are applied, two burst size thresholds are used to filter smFRET data (on results of DCBS with m=10 and F=6):

1. $\left(n_{DA} - lk \cdot n_{DD} - dir \cdot n_{AA}\right) + \gamma \cdot n_{DD} \geq 25$
2. $n_{AA} \geq 25$

1D FRET histograms are extracted from the 2D E-S scatter plot (as 1D projection on the E axis).

Each transcription quenched kinetics time point consists of the same two FRET populations but with different fractions that designate the evolution of the run-off transcript production (Fig. 1).

All corrected FRET histograms of all time points in a given kinetic track are globally fitted to a sum of two Gaussians with the constraint that their means and standard deviations are the same

for all datasets. The amplitudes of the two Gaussians are kept as free fitting parameters independent of the time point.

The background-dependent burst search and selection in this work was performed using FRETBursts[73], an open source burst analysis program for smFRET data. All model fittings were performed using the Matlab software (MathWorks Matlab, Natick, MA, USA) through the *lsqcurvefit* nonlinear regression function.

**4. Transcription bubble closure kinetics assays:**

Transcriptional activity quenched by 0.5 M GndCl (Fig. 3B) and bubble closure kinetics due to promoter clearance (Fig. 5) were characterized by transcription bubble conformation experiments. These experiments were performed on $RP_{ITC=2}$ utilizing the *lac*CONS promoter (sequence in Fig. 2) labeled with an Atto 550 donor at register (-5) of the template strand and an Atto 647N acceptor at register (-8) of the nontemplate strand (IBA oligos, GmbH). Transcriptional activity quenching assays were performed with a single-spot µsALEX set-up[55,56] to test whether the transcription bubble is open (the dyes on the bubble get farther away from each other and FRET decreases) or closed (the annealing of the DNA upon bubble closure brings the two dyes to be as close as 3 bp's apart and hence a very high FRET value) upon addition of 0.5 M GndCl. The same burst analysis as described above was utilized with the exception of higher corrected burst size filters ( $\left(n_{DA} - lk \cdot n_{DD} - dir \cdot n_{AA}\right) + \gamma \cdot n_{DD} \geq 100$ and $n_{AA} \geq 100$ ). The quenched kinetics assay probes run-off production rate, which depends both on initiation and elongation rates. The bubble closure kinetics assay, however, probes the promoter clearance kinetics in real-time. A multi-spot smFRET set-up (diffusion format) equipped with a 8x1 SPAD array[98,99] and exhibiting a temporal resolution of 30 sec for data acquisition per kinetic point was used for measuring the bubble closure kinetics. Comparison of quenched kinetics and bubble closure kinetics shows that elongation is very rapid and contributes negligibly to the overall run-off transcript production rate (Fig. 5).

## 5. Transcription assays visualizing abortive product formation using urea-denaturing PAGE analysis of [$^{32}$P]-radiolabeled RNA products

Abortive transcription assays were run using the *lac*CONS promoter having its probe target 20A sequence replaced by the WT LacUV5 sequence at registers from +20 to +39 (see Fig. 2). Three units of RNAP holoenzyme (NEB, Ipswich, MA, USA, M0551S) were mixed with 50 nM promoter DNA in 1x transcription buffer in a final volume of 20 µL. The reaction was then incubated at 37 $^0$C for 20 minutes 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 minute, centrifuged and 20 µL of the supernatant was removed and added to 10 µL pre-warmed transcription buffer. After incubating an additional 10 minutes, $A_pA$ was added at a final concentration of 1.3 mM and incubated for 30 minutes 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, Waltham, MA, USA) 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 $^0$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, an additional 15-minute incubation at 37 $^0$C was performed

before the addition of NTPs, either in the presence or absence of 1 μM GreA. The stopped reaction aliquots were stored at -20$^0$C until running the urea-denaturing PAGE.

Samples were heated for 3 minutes at 90 $^0$C and loaded on a 23%, (19:1 acrylamide:bis-acrylamide) 0.4 mm thick urea-denaturing polyacrylamide gel. The gel usually ran for 5 to 6 hours at 1500 V in 1x TBE with an additional 0.3M sodium acetate in the bottom well. The gels were then removed, dried, and exposed on a phosphor-storage screen about 2 days. Screens were visualized using a Typhoon PhosporImager.

## 6. Magnetic trapping assay

### 6.1. DNA constructs

We designed and had custom-synthesized (Eurofins MWG) DNA fragments flanked by KpnI sites containing a modular region for insertion of a promoter and initial transcript, followed by a transcribed region and a terminator. The modular region is flanked by HindIII and SpeI sites:

5'GGTACCAAGCTTGCGAACTGCACTCGGAAC**ACTAGT**ATGCATCGAATAGCCATCC
CAATCGATATCGAGGAGTTTAAATATGGCTGATGCATGAATTCGTTAATAACAGGCC
TGCTGGTAATCGCAGGCCTTTTTATTTGGGAATTCGGTACC

where KpnI sites are indicated in red; HindIII and SpeI sites are underlined; and the tR2 terminator is in purple. This transcription backbone was cloned into the KpnI site of the *Th. aquaticus* RPOC gene, and a 2.2 kbp subfragment of this construct centered about the transcription unit was PCR amplified and subcloned into the XbaI and SbfI sites of pUC18 using HiFi thermostable polymerase (Roche) and PCR primers (XbaI and SbfI sites underlined):

5' GAGAGATCTAGAGACCTTCTGGATCTCGTCCACCAGG
and 5' GAGAGACCTGCAGGACATCAAGGACGAGGTGTGG

We then cloned the lacCONS promoter into the HindIII and SpeI sites underlined above using the oligo-based dsDNA fragment with the top strand:

5'AGCTAGGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCG
GATTAG

Similarly, we cloned the T5N25 promoter using the dsDNA fragment with the following top strand:

5’AGCTAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGTATAATAGATTCATAAATTT
GAGAGAGGAGTCC

DNA for single-molecule experiments was prepared from freshly grown DH5α by ion-exchange chromatography (Macherey-Nagel), digested with XbaI and SbfI, and the 2.2 Kb band isolated by gel purification and extraction using spin column (Macherey Nagel).

The 2.2 kbp DNA fragments containing the centrally-located transcription unit were ligated at the XbaI site to 1 kbp DNA multiply-labelled with biotin, and at the SbfI site to 1 kbp DNA multiply-labelled with digoxigenin. Labelled DNAs were synthesized via PCR carried out in the presence of dUTP-biotin and dUTP-digoxigenin, respectively (Roche)[21,119].

### 6.2. Single-Molecule experiments

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)[119]. 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.

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 a ~55 nm decrease in DNA extension.

Experiments were carried out in standard buffer at 34°C using 100 pM RNAP saturated with σ70 (prepared as in (*15*)) and 100 μM $A_pA$ (for experiments on *lac*CONS 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.

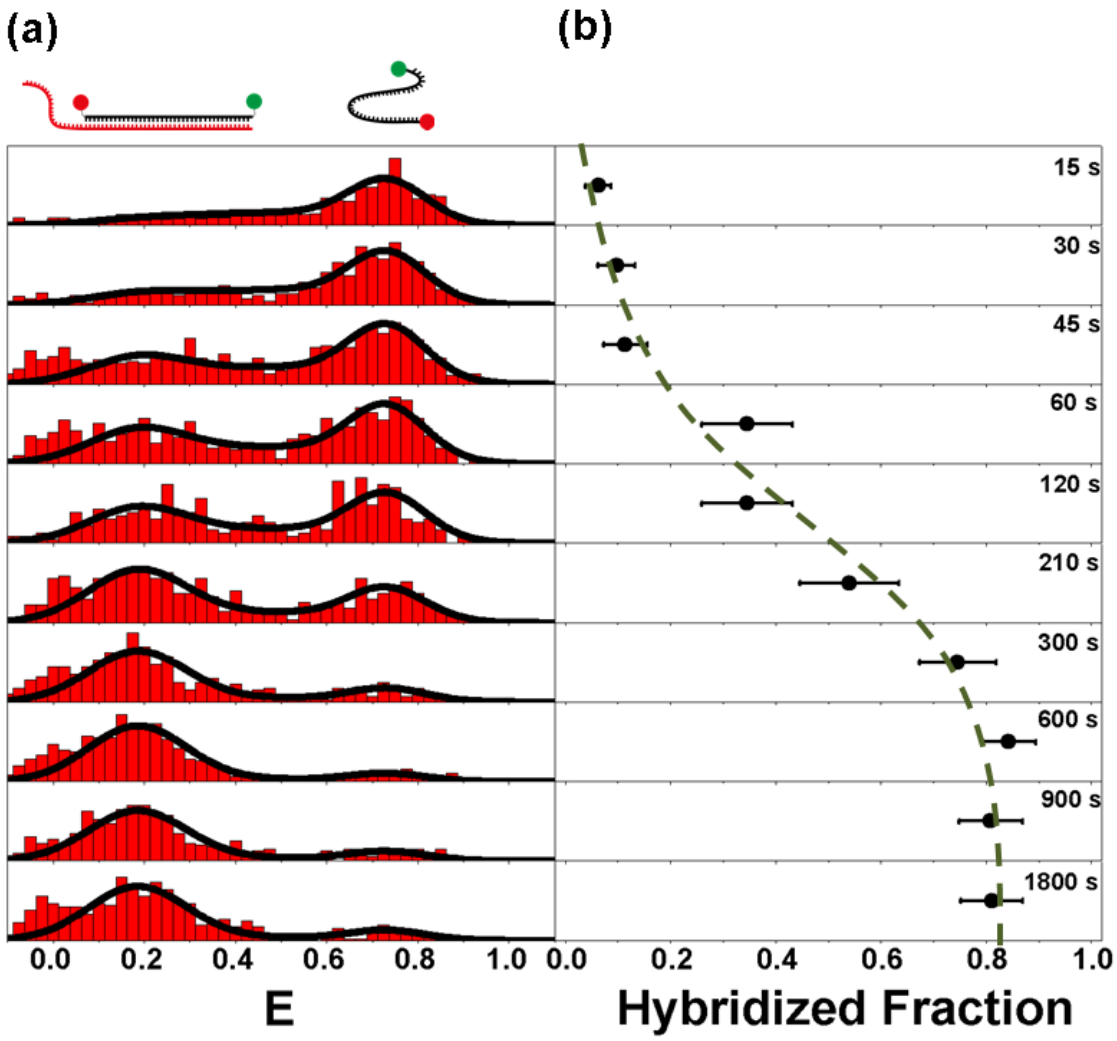



**Figure. 1: Quenched kinetics derived from DNA FRET probe hybridization to an RNA transcript**. **(a):** smFRET results of the quenched kinetic assay. 100 pM concentrated ssDNA FRET probes were added to transcription reaction mixtures and incubated for 20 mins. to allow hybridization. The transcription reaction mixtures were then quenched by 0.5 M GndCl after different incubation times with all four NTPs. Each smFRET acquisition lasted 10 mins. Double gaussian global fits with globally shared means and widths (for the whole data set)

yield the fraction of low FRET population (i.e. number of hybridized probes that report the percentage of synthesized run-off transcripts). **(b):** the fraction values (together with the associated fitting error at 95% confidence) are plotted as function of quenching times. The dashed curve serves as a guide to the eye. All kinetic data reported in this work are derived in the same way.

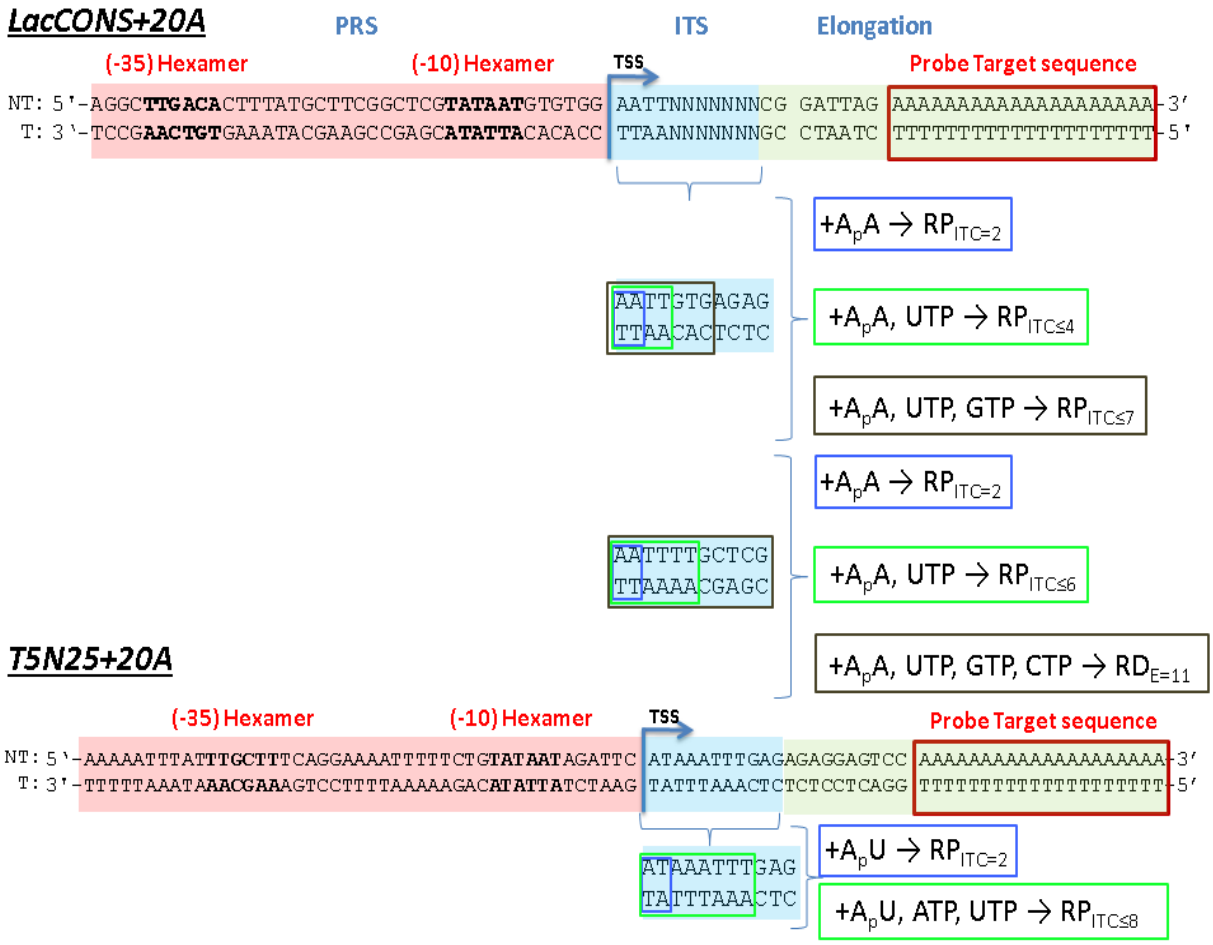



**Figure. 2: *lac*CONS and T5N25 promoter sequences used in Chapter 3.** Highlighted are the Promoter Recognition Sequence (PRS; in pink), the Transcription Start Site (TSS; in bent blue arrow), the Initially Transcribed Sequence (ITS; in cyan), and the elongation sequence (in yellow). The elongation sequence contains a 20 dT stretch that transcribes into a 20 dA stretch in the RNA, and detected by the doubly labeled ssDNA FRET probe. Bottom: ITS sequences that have been used to prepare the system in various $RP_{ITC}$ states, and the corresponding NTP starvation mixtures are also shown.

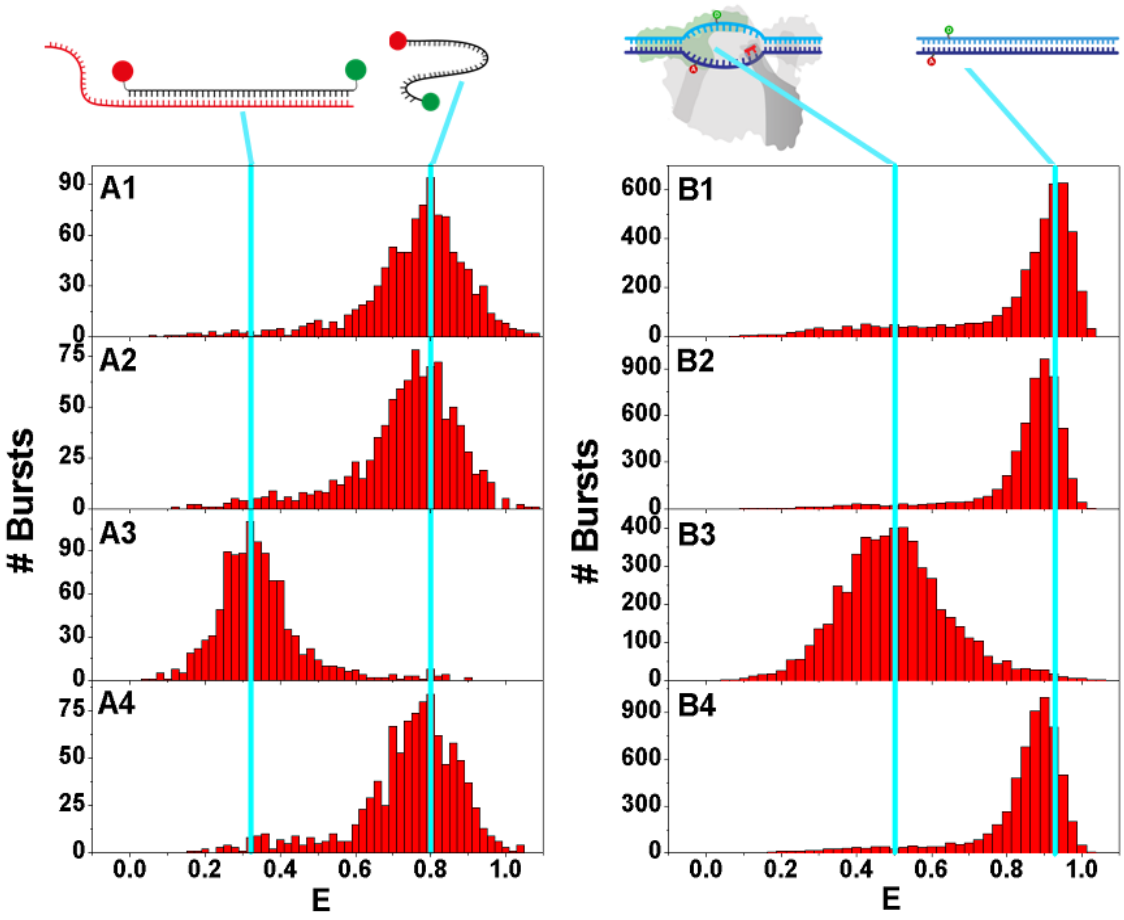



**Figure 3: The effectiveness of Guanidinium chloride as a transcription reaction quencher.**

(**A**) Quantification of RNA transcript production by hybridization of a FRET probe to the transcripts. The higher FRET peak represents probe only. The lower FRET peak represents probe hybridized to synthesized RNA transcripts. Panel **A1** (probe only) and panel **A2** (transcription complexes + probe + with no NTP) serve as controls. Panel **A3** shows transcript production (GndCl added only after the reaction is completed). Panel **A4** shows that if GndCl is added to the reaction mix before the addition of NTPs, the reaction is inhibited.

(**B**) Quantification of transcription bubble opening by smFRET. The higher FRET peak represents the free (doubly labeled) promoter. The lower FRET peak represents the promoter

with an open bubble ($RP_{ITC=2}$). Panel **B1** (promoter only) and panel **B2** (promoter + GndCl) serve as controls. Panel **B3** shows $RP_{ITC=2}$ (promoter + RNAP + $A_pA$). Panel **B4** shows that if GndCl is added, the bubble closes-up due to dissociation of the complex.

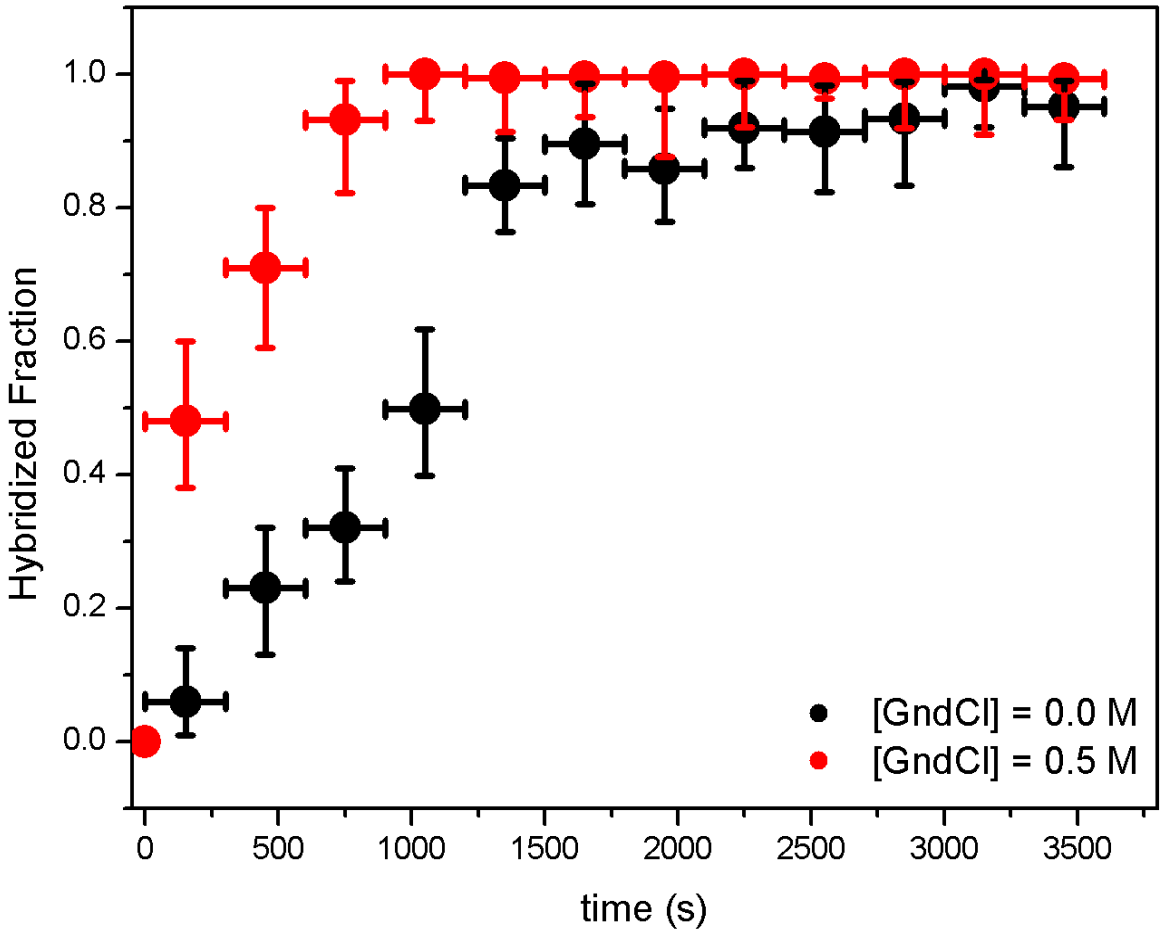
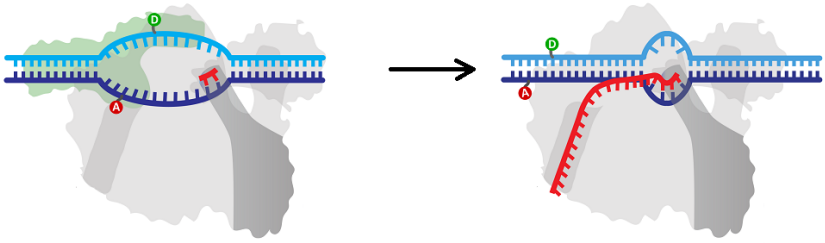
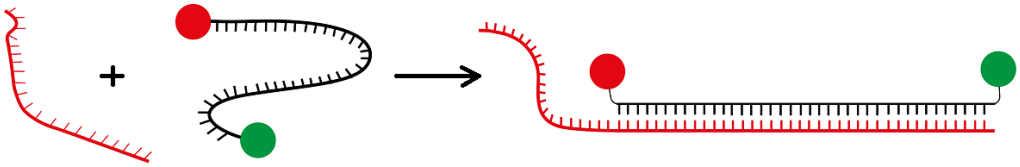



**Figure 4: Guanidinium chloride enhances and accelerates ssDNA probe hybridization to RNA transcripts**. Real-time hybridization kinetics of 20 dT ssDNA FRET probe to 20A RNA target. 100 pM of probe is added to 100 pM of target at t=0. smFRET histograms are accumulated at 5 minutes intervals (x-axis erros). Hybridization fraction is extracted from the ratio of the low FRET subpopulation to the sum of low FRET and high FRET subpopulations. The addition of 0.5 M GndCl to the hybridization mix accelerates the hybridization reaction (red) as compared to 0 M GndCl (black) and enhances the hybridization efficiency (red data points above black points at long times).

○ Initiation-bubble closure



● Runoff transcript production



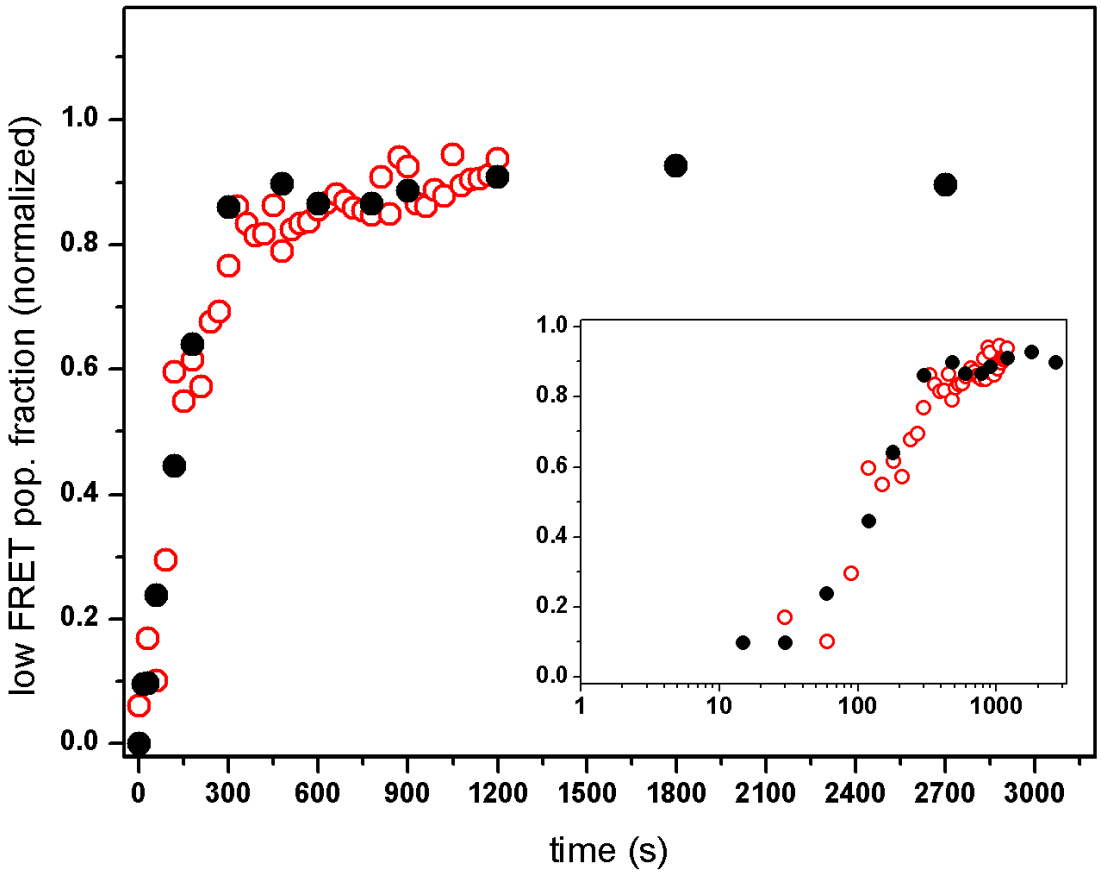


**Figure 5: Comparison of transcription run-off kinetics to transcription bubble closure kinetics.** Transcription run-off kinetics (filled black circles) follows closely the transcription bubble closure kinetics (measured using an 8x1 multi-spot excitation smFRET setup[98,99], open red circles). This comparison shows that the elongation part of the overall transcription kinetics is fast and initiation is rate-limiting. The inset shows the same kinetic curves on a semi-logarithmic scale.

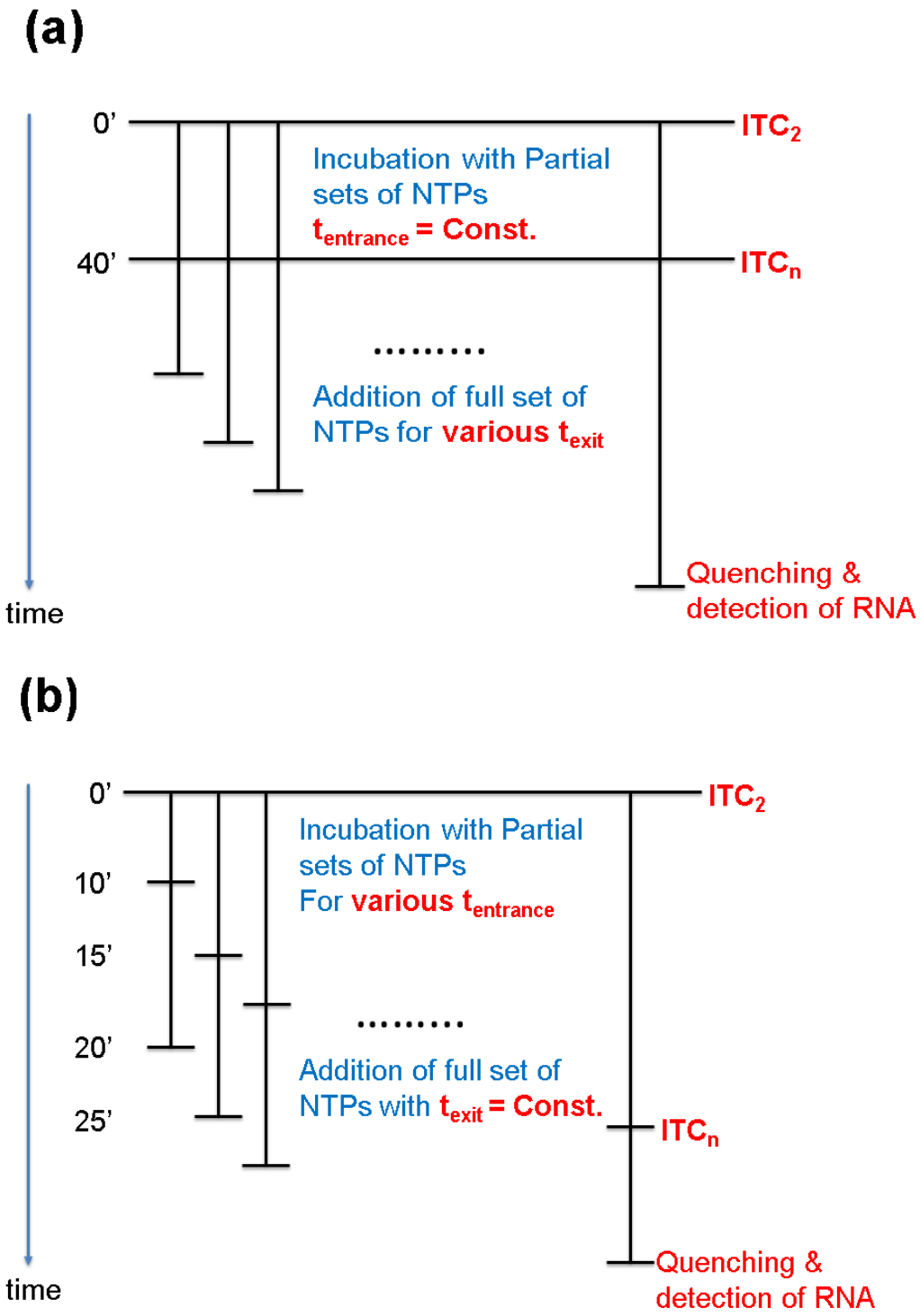



**Figure 6: Experimental schemes for measuring exit and entrance kinetics.** (**A**) Exit kinetics: $RP_{ITC=2}$ is incubated for 40 mins with a partial set of NTPs to drive the transcription complex

into a given $RP_{ITC \leq n}$ . Following, a full set of NTPs is added to multiple aliquots that are quenched by 0.5 M GndCl at different times, $t_{exit}$. ssDNA FRET probe is then added and the number of synthesized transcripts is analyzed by ALEX for each reaction. Kinetic curves are extracted as in Fig. 1. (**B**) Enterance kinetics: Multiple aliquots of $RP_{ITC=2}$ are incubated for various times, $t_{entrance}$ , after which a full set of NTPs is added and incubated for a fixed 10 minutes duration. Following, reactions are quenched by 0.5 M GndCl. ssDNA FRET probe is then added and the number of synthesized transcripts is analyzed by ALEX for each reaction. Kinetic curves are extracted as in Fig. 1.

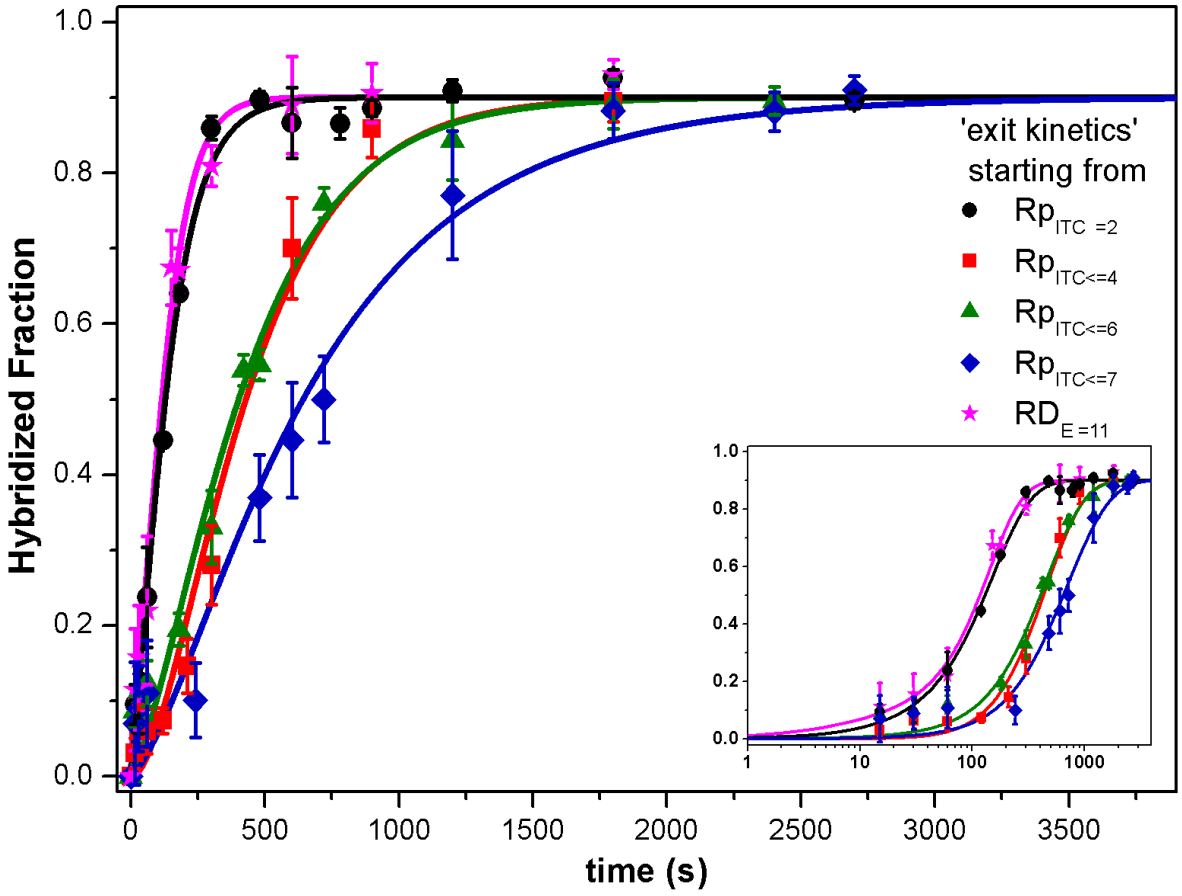



**Figure 7: Transcription runoff kinetics after rescue from NTP-starved initiation states (Exit kinetics).** The time course of run-off RNA production using a LacCONS promoter (Fig. 2) starting from different NTP-starved states in initiation and in elongation, stable open transcription bubble ($RP_{ITC=2}$, black), initially transcribing complexes ($RP_{ITC\leq i}$; i=4 – red; i=6 – green; i=7 – blue) and first stable RNA-DNA Elongation complex ($RD_{E=11}$ – magenta). The values designate arithmetic averages of duplicate or triplicate measurements (see Fig. 20) of the fraction of smFRET bursts of a FRET ssDNA probe that is hybridized to a run-off transcript when the transcription reaction was quenched after a given time. The error bars are the standard

deviations of smFRET population fractions. The curves are best-fit results of the kinetic data to a simplified model of branching into a non-catalytic state (see Fig. 17 for model and Table 1 for best-fit values). The inset shows the same kinetics on a semi-logarithmic scale.

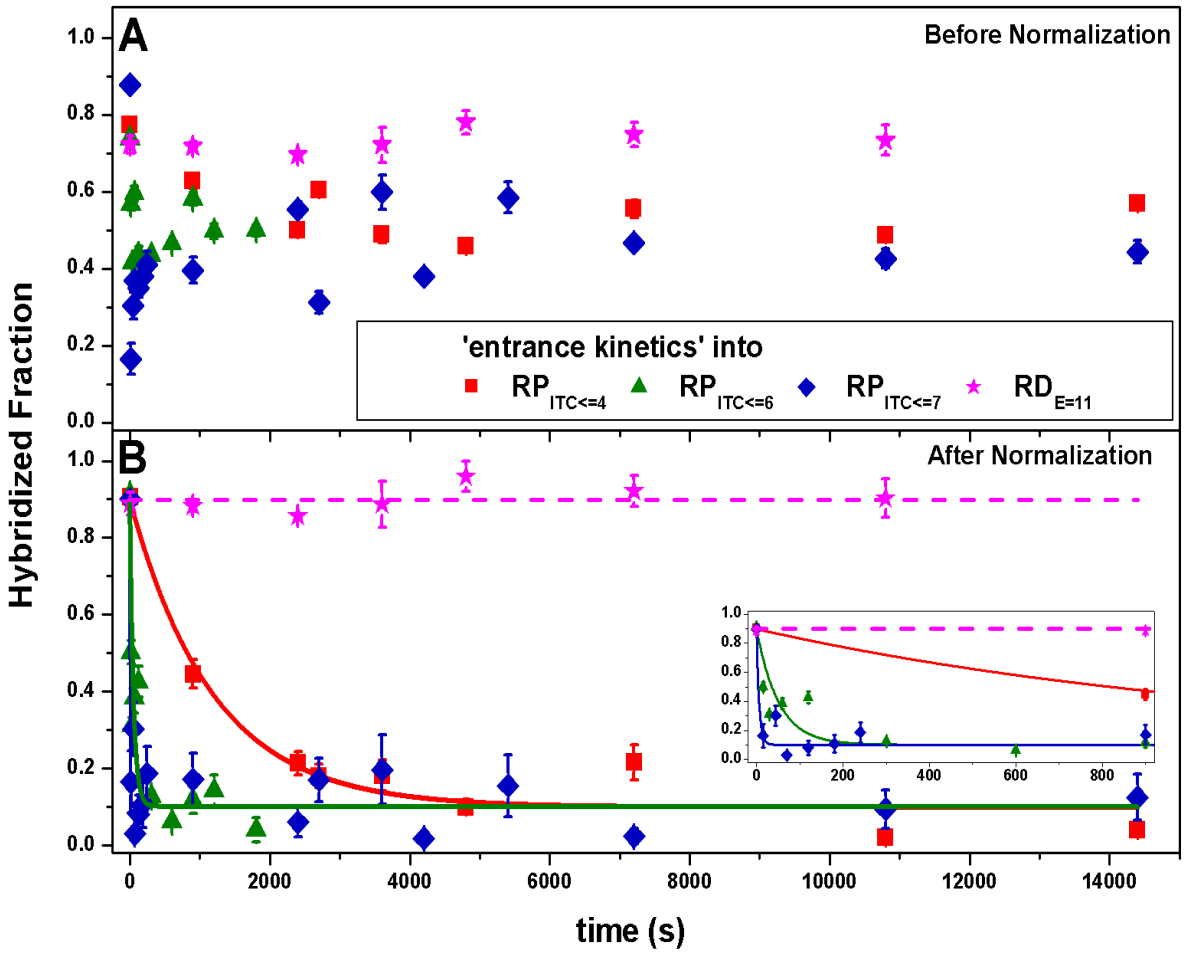



**Figure 8: Entrance kinetics into initiation off-pathway states.** Entrance kinetic were measured indirectly (through transcription run-off kinetics, see Fig. 6) as a function of incubation time under NTP-starvation conditions. Results of the time course of RNA run-off production a *lac*CONS promoter (See in Fig. 2) before (A) and after normalization (B). Entrance into NTP-starved states: $RP_{ITC\leq4}$ (red), $RP_{ITC\leq6}$ (green); $RP_{ITC\leq7}$ (blue) and $RD_{E=11}$ (magenta). These are rescued by the addition of all four NTPs for a fixed incubation time followed by reaction quenching. The values represent averages of duplicate or triplicate measurements of the fraction of low FRET efficiency bursts from ssDNA probe hybridized to a run-off transcript, after reaction quenching. The error

bars are the standard deviations of measurement repeats. The curves are best-fit results of the kinetic data to an exponential function (see Table 2 for best fit values). The inset shows the details of the first 900 s.

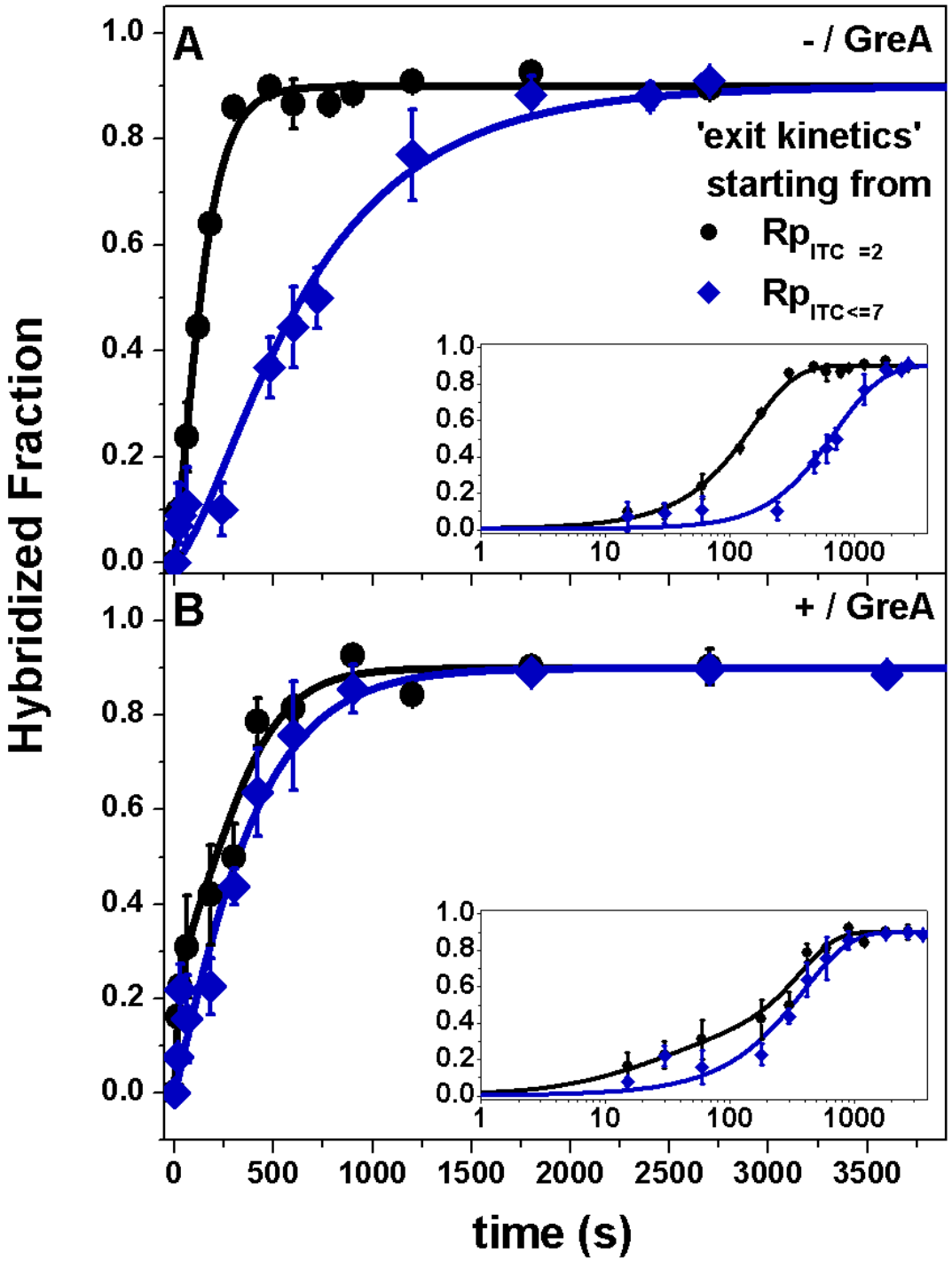

A
- / GreA
'exit kinetics'
starting from
Rp ITC =2
Rp ITC<=7
B
+ / GreA
Hybridized Fraction
time (s)
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
3500
1
10
100
1000

**Figure 9: The effect of GreA on transcription run-off kinetics after rescue from starved initiation states.** Shown are results of RNA run-off production kinetics using a *lac*CONS promoter starting from stable open transcription bubble ($RP_{ITC=2}$, black) and late initially transcribing complexes ($RP_{ITC\leq7}$, blue) in the absence (A) or presence (B) of 1 μM GreA. The values designate arithmetic averages of duplicate or triplicate measurements (see Fig. 20) of the fraction of smFRET bursts of a FRET ssDNA probe that is hybridized to a run-off transcript when the transcription reaction was quenched after a given time. The error bars are the standard deviations of smFRET population fractions. The curves are best fit results of the kinetic data to a simplified model of branching into an off-pathway state (see Fig. 17 for model and Table 1 for best fit values). The inset shows the same data on a semi-logarithmic scale.

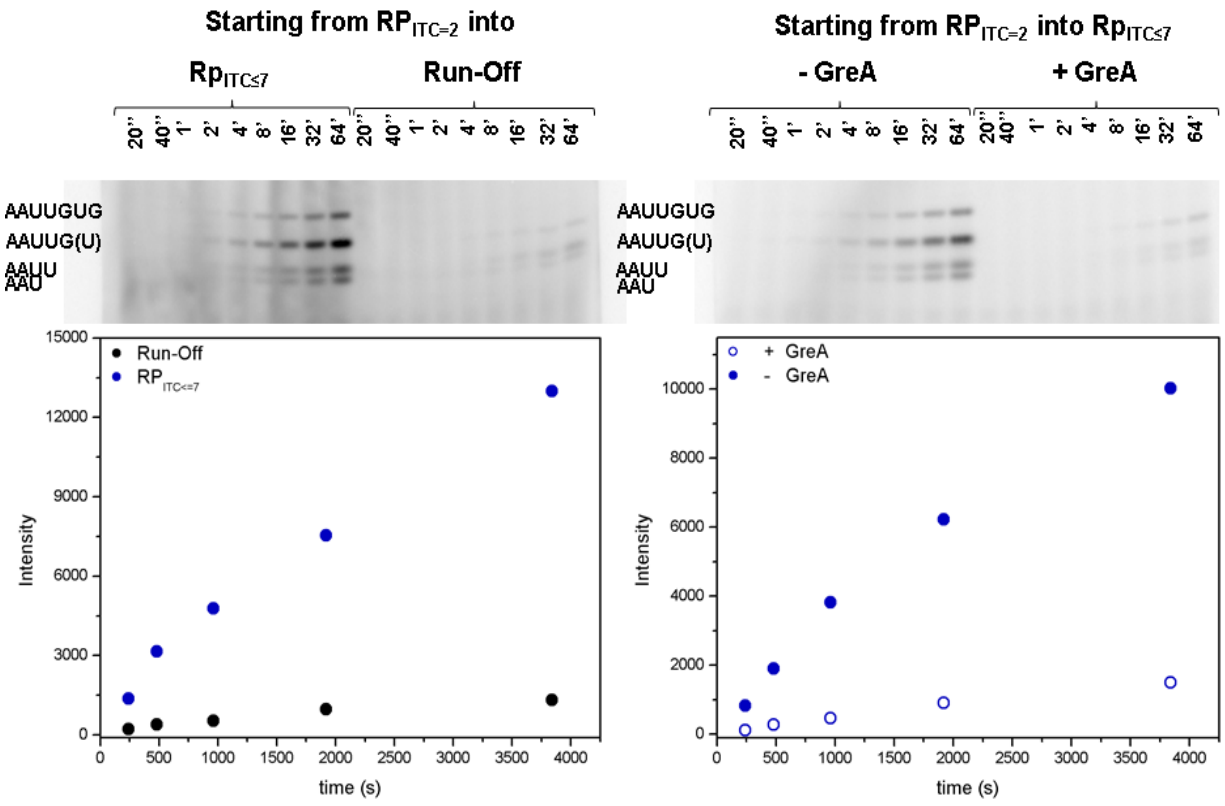



**Figure 10: Abortive Transcription Entrance Kinetics Gel Assays.** Top: Autoradiograms of the urea-denaturing PAGE results comparing the patterns of radiolabeled abortive RNAs produced a LacCONS LacUV5 promoter (LacCONS (Fig. 2) until position +19, LacUV5 from +20 to +39, replacing the target identification 20dA sequence, see Fig 2) under NTP starvation versus normal conditions (left panel) and under NTP starvation in the absence and presence of GreA (right panel). Band assignments follow the ITS sequence (see Fig. 2) up to a 7-mer abortive product ($RP_{ITC\leq7}$) and the marker-based band assignments (Fig. 18). Bottom Left: NTP starvation (blue) versus full set of NTPs present (black). Bottom Right: NTP starvation in the absence (filled) or presence (open) of GreA.

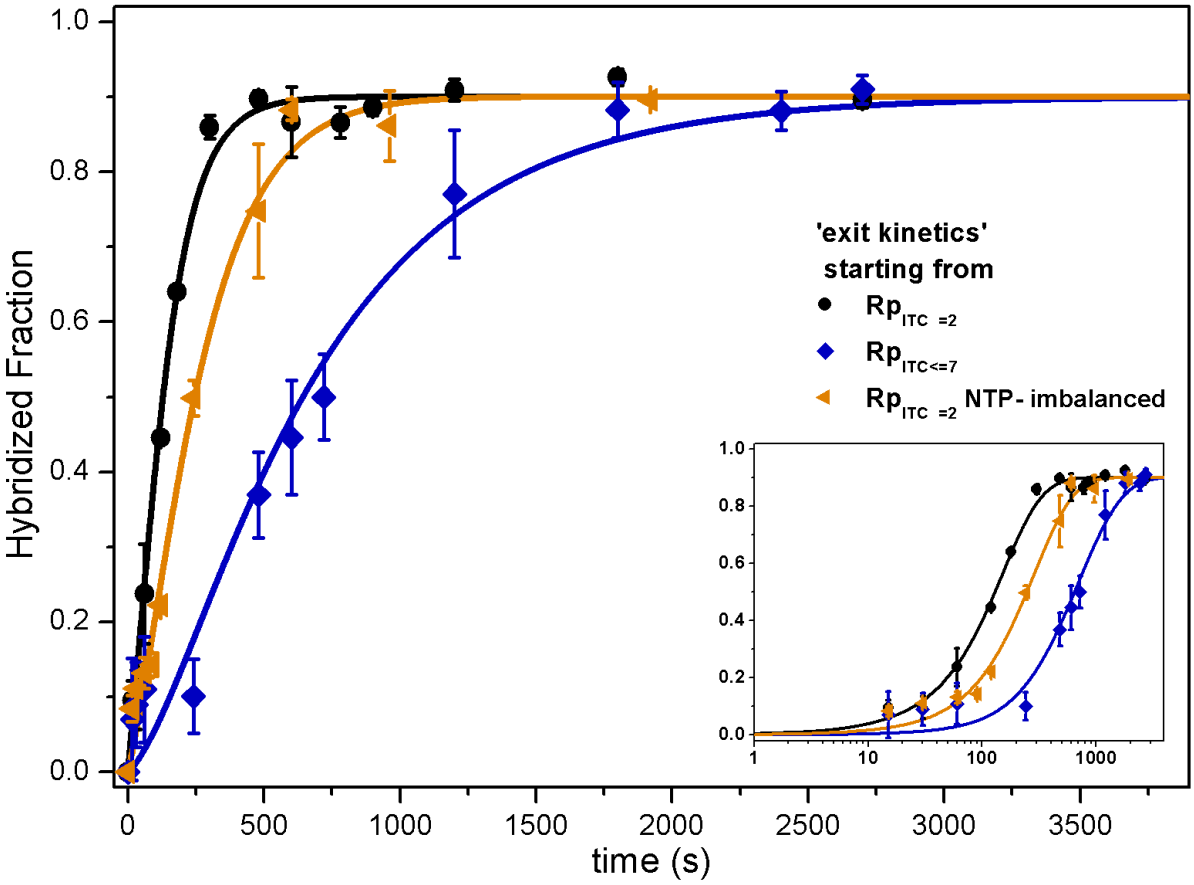



**Figure 11: Transcription run-off kinetics under partial NTP starvation.** RNA run-off production kinetics (when rescued with all four NTPs) for *lac*CONS promoter after incubation with no, partial, and full starvation NTP mixes. Shown are kinetic curves for exit out of stable open transcription bubble ($RP_{ITC=2}$) 1) after supplementing equimolar NTPs (all four NTPs at 100 μM; black), 2) after supplementing NTPs with a concentration imbalance biased towards $RP_{ITC\leq7}$ (UTP and GTP at 100 μM and ATP and CTP at 2 μM; orange), and 3) after supplementing equimolar NTPs to a fully starved $RP_{ITC\leq7}$ (first incubation with no ATP and CTP,then adding all four NTPs at 100 μM; blue). The values designate arithmetic averages of duplicate or triplicate

measurements (see Fig. 20) of the fraction of smFRET bursts of a FRET ssDNA probe that is hybridized to a runoff transcript when the transcription reaction was quenched after a given time. The error bars are the standard deviations of smFRET population fractions. The curves are best-fit results of the kinetic data to a simplified transcription model of branching into an off-pathway state (see Fig. 17 for model and Table 1 for best-fit values). The inset shows the same data on a semi-logarithmic scale.

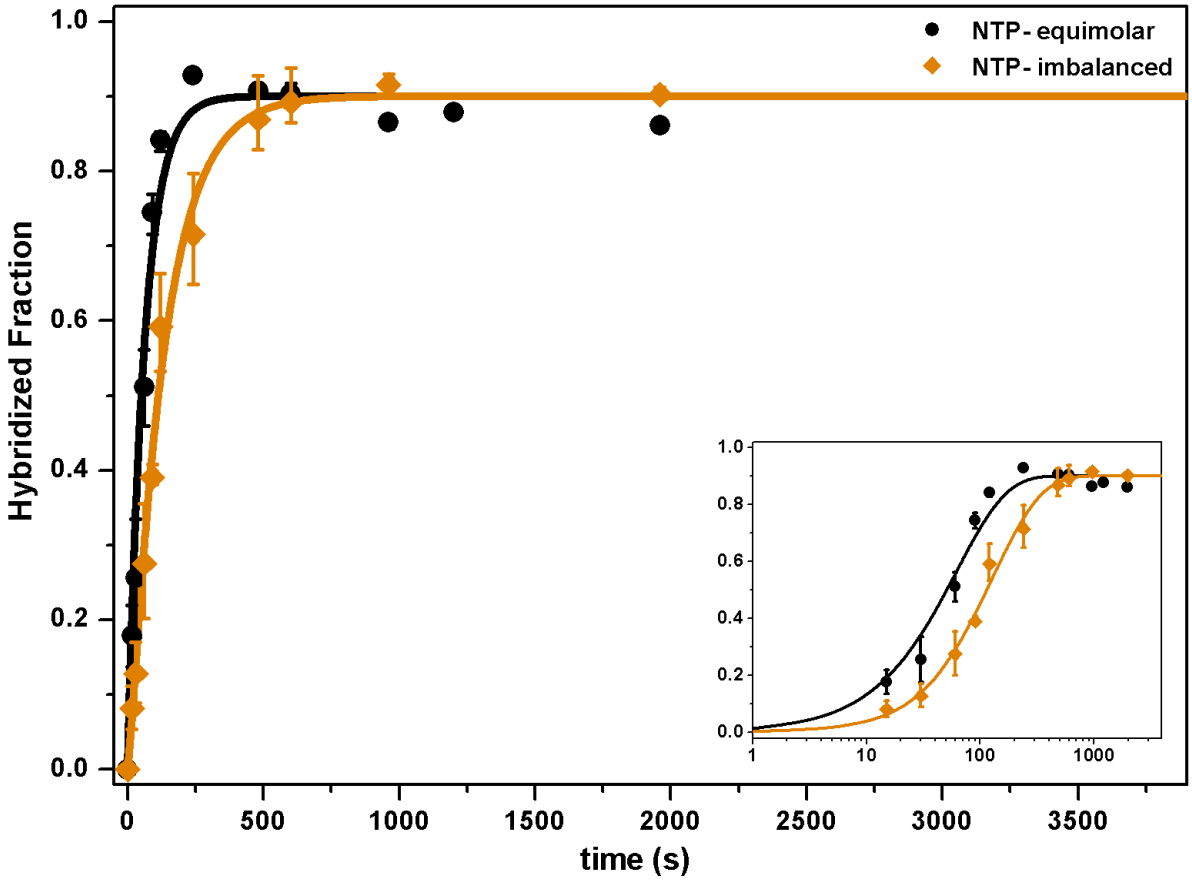



**Figure 12: Transcription run-off kinetics under no- and partial- NTP starvation for the T5N25 promoter.** Shown are RNA run-off production kinetics starting from $RP_{ITC=2}$ in the presence of equimolar NTPs ( all four NTPs at 100 Mm; black) and in the presence of imbalanced NTP concentrations (ATP and UTP at 100 μM and CTP and GTP at 2 Mm; orange). This NTPs imbalance biases complexes, on this promoter, towards $RP_{ITC\leq 8}$. The values designate arithmetic averages of duplicate or triplicate measurements (see Fig. 20) of the fraction of smFRET bursts of a FRET ssDNA probe that is hybridized to a run-off transcript when the transcription reaction was quenched after a given time. The error bars are the standard deviations of smFRET population fractions. The curves are best-fit results of the kinetic data to a simplified

model of branching into an off-pathway state (see Fig. 17 for model and Table 1 for best-fit values). The inset shows the same kinetics on a semi-logarithmic scale.

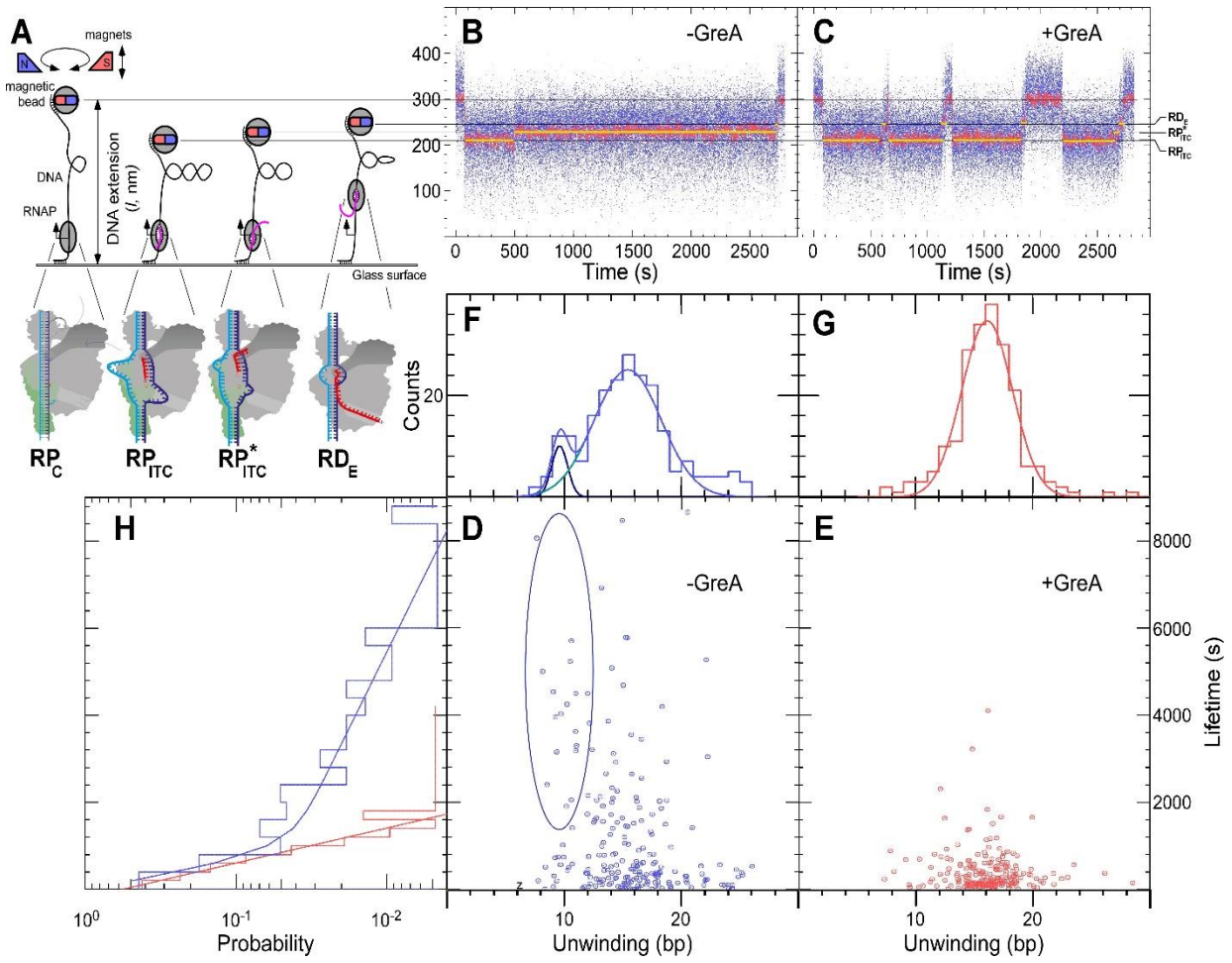



**Figure 13: *lac*CONS promoter - Backtracking in initiation correlates with RNAP pausing in the presence of equimolar NTPs.** (A): Schematics of the magnetic tweezer transcription assay: Upon RNAP binding, a transcription bubble is formed. Different transcription states are associated with different bubble sizes. According to the sequence of events, after RNAP-Promoter binding a transcription bubble is formed. Then, after adding NTPs, while in initiation, DNA downstream to the bubble is scrunched to enlarge the transcription bubble. After transition into elongation, the strain caused by the interaction with $\sigma^{70}$, is relieved, after promoter clearance and the size of the bubble decreases. After running off the promoter and RNAP dissociation from the promoter, the bubble closes. The larger the bubble size is the shorter the extension of the bead will be.

Representative bead extension trajectories are shown for experiments absence (B) or presence (C) of 1 μM GreA. One can observe extension levels (grey lines) associated to the different bubble sizes imposed by the different transcription states. Yellow lines highlight the typical dwell lifetimes in each state. The extension levels and lifetimes of different initiation state occupancies are summarized into extension-lifetime scatter plots when GreA was either absent (D) or present (E) and their 1D projections are shown in panels F, G and H. The ellipse in panel D highlights the existence of a low yet non-negligible number of events which show lifetimes of thousands of seconds and extensions shorter than the ones obtained for smaller lifetimes (see also the low extension subpopulation in panel F), that is absent in the presence of GreA (E, G).

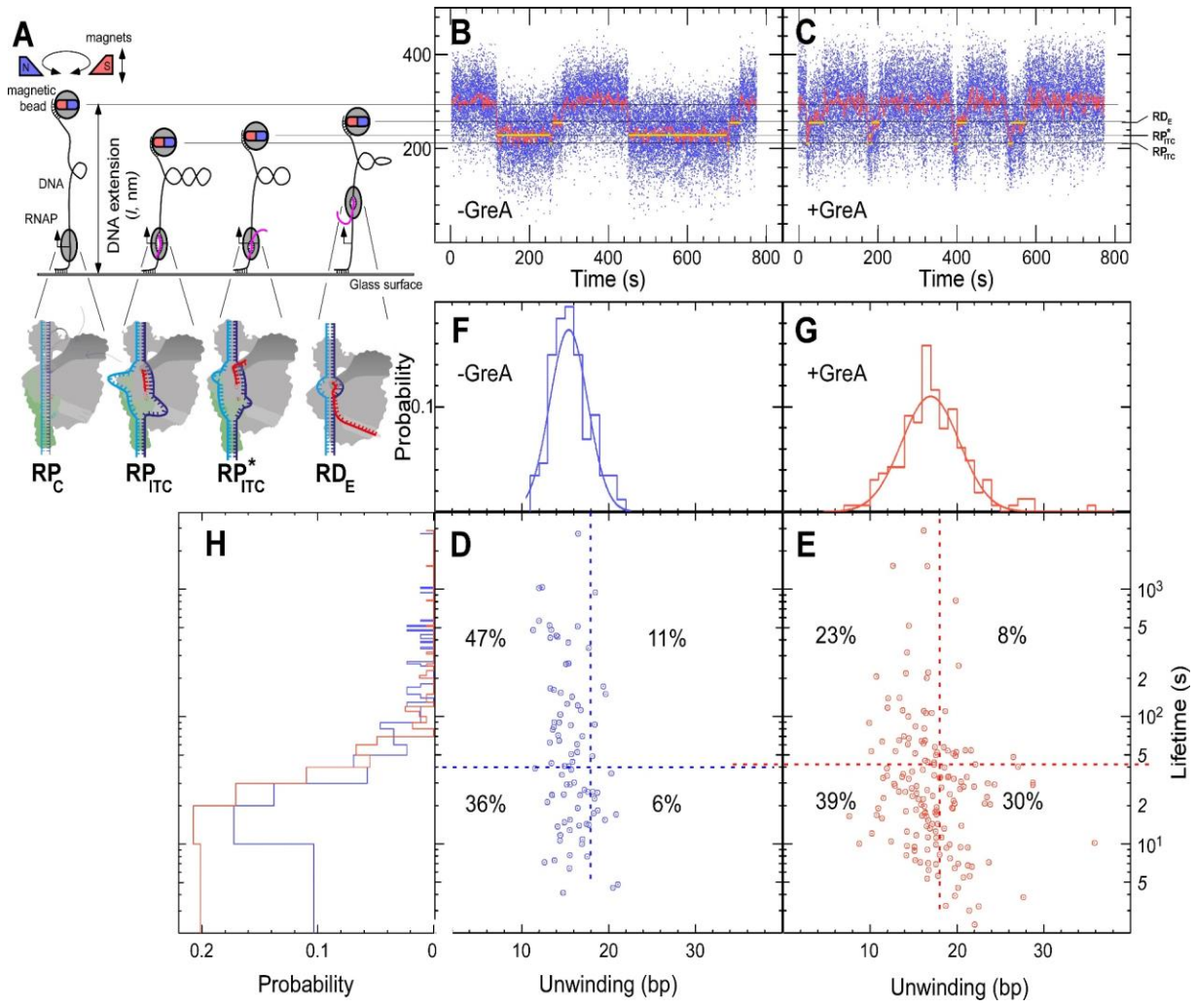



**Figure 14: T5N25 promoter - Backtracking in initiation is correlated with pausing in the presence of all NTPs.** (A): Schematics of the magnetic tweezer transcription assay. Representative DNA unwinding bead extension trajectories are shown for experiments in the absence (B) or presence (C) of 1 µM GreA. One can observe unwinding levels (grey lines) associated with the different bubble sizes imposed by the different transcription states. Yellow lines highlight typical lifetimes in $RP_{ITC}$ and $RP^*_{ITC}$ states. The unwinding levels and lifetimes of individual $RP_{ITC}$ and $RP^*_{ITC}$ events are summarized into unwinding-lifetime scatter plots when GreA was either absent (D, n=87) or present (E, n=164) and their 1D projections are shown in

panels F, G and H. The scatter plots (D & E) are divided into quadrants to highlight the change in the results upon the addition of GreA on the basis of lifetime and unwinding levels. Experiments were carried out under conditions essentially identical to those in Fig. 13, but for the use of $A_pU$ instead of $A_pA$ as initiating dinucleotide (100 μM, see Materials and Methods).

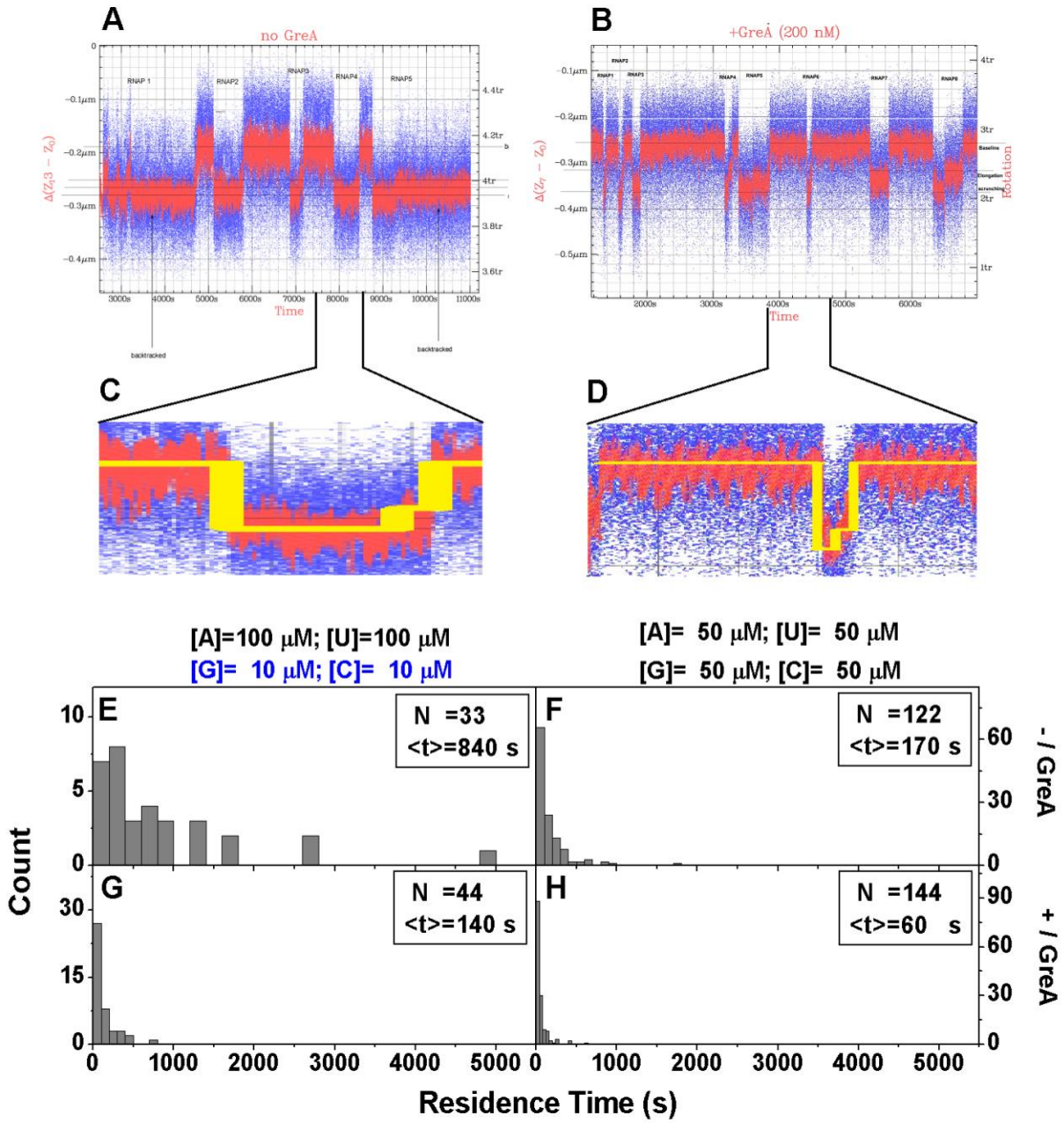



**Figure 15: T5N25 promoter - Transcription initiation under partial starvation of NTPs (concentration imbalance). (A)** and **(B)** Typical magnetic tweezer super coiling transcription bubble trajectories (steady state). (**C**) and (**D**) Zoom-ins on transcriptional events in **(A)** and **(B)**.

Four DNA extension states are identified: free DNA (baseline), on-pathway initiation (max-scrunch), off-pathway initiation (backtracked) and elongation. Initiation dwell time histograms (**E**) and (**G**) under NTPs imbalance and (**F**) and (**H**) under equimolar NTPs are shown. (**E**) and (**F**) – are for data taken without GreA, (**G**) and (**H**) – are for data taken in the presence of 1 μM GreA.

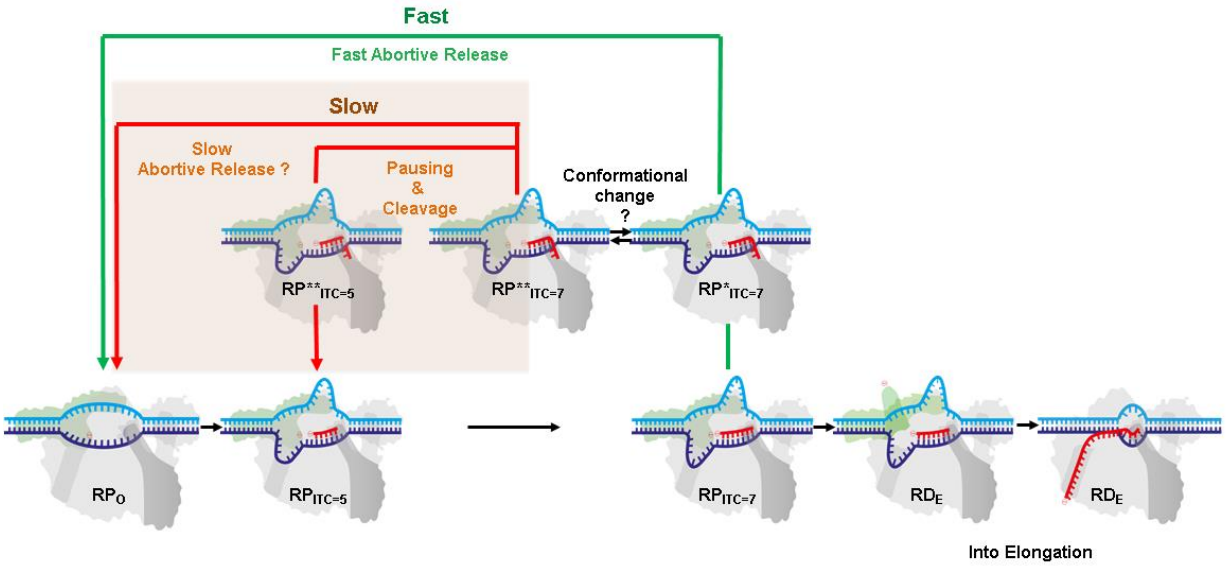
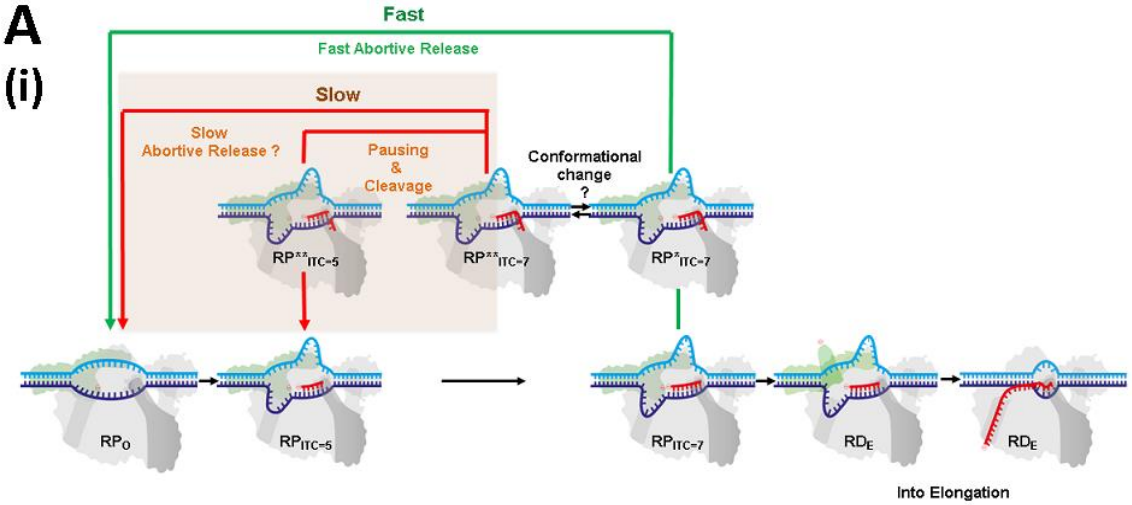


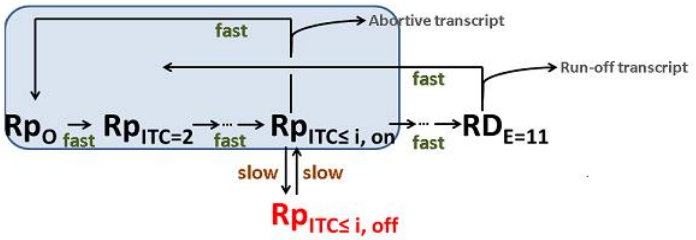

**Figure 16: A modified transcription initiation model.** RNAP transcription initiation branches to σ-subunit release, and transitions into elongation (black arrows) or into release of abortive transcripts (green and red arrows). After initial backtracking steps ($RP_{ITC=7}$ goes to $RP^*_{ITC=7}$), the complex can either continue with fast abortive transcripts release (classic model, green arrow), or transition to a non-catalytic backtracked state (via a hypothesized conformational change; $RP^*_{ITC=7}$ goes to $RP^{**}_{ITC=7}$). This transcriptionally incompetent, paused state ($RP^{**}_{ITC=7}$) is relatively stable (acting as a kinetic trap). Rescue from the trap can occur either by successive backtracking steps followed by transcript release (red arrow), or through cleavage of RNA bases that are in the secondary channel. At this point the complex can reinitiate (red arrow). The processes described by red arrows are rare and slow (they occur only if a backtracked state is stabilized enough to cause pausing).

**A**

**(i)**




**(ii)**




**(iii)**


**B**

**Rate equations:**

$$\frac{d}{dt}\begin{pmatrix} on \\ off \\ RO \end{pmatrix} = \begin{pmatrix} -\left(k_{on\to RO} + k_{on\to off}\right) & k_{off\to on} & 0 \\ k_{on\to off} & -k_{off\to on} & 0 \\ k_{on\to RO} & 0 & 0 \end{pmatrix} \begin{pmatrix} on \\ off \\ RO \end{pmatrix}$$

**Initial conditions:**

| Starting from $Rp_{ITC=2}$ | Starting from $Rp_{ITC\le i}$ |
|---|---|
| $[on]_{t=0} = 1$ | $[on]_{t=0} = \alpha$ |
| $[off]_{t=0} = 0$ | $[off]_{t=0} = 1-\alpha$ |

**Figure 17: A simplified transcription model.** Panel (**A)**, shows how the modified transcription initiation model shown in Fig. 16 and here (**A-(i)**) is reduced to a chemical scheme (**A-(ii)**). Further simplification bunches all on-pathway states with fast transitions into a single 'On' state (**A-(iii)**). The model is therefore simplified into three main states, 'On' which represents all states in the on-pathway in initiation, 'Off' which represents the Off-pathway initiation state and 'RO' which represents transition into elongation and ultimately into run-off. Panel (**B)** shows the equations that describe this simplified model together with the corresponding initial conditions. These equations are used to globally fit all kinetic data in this study.

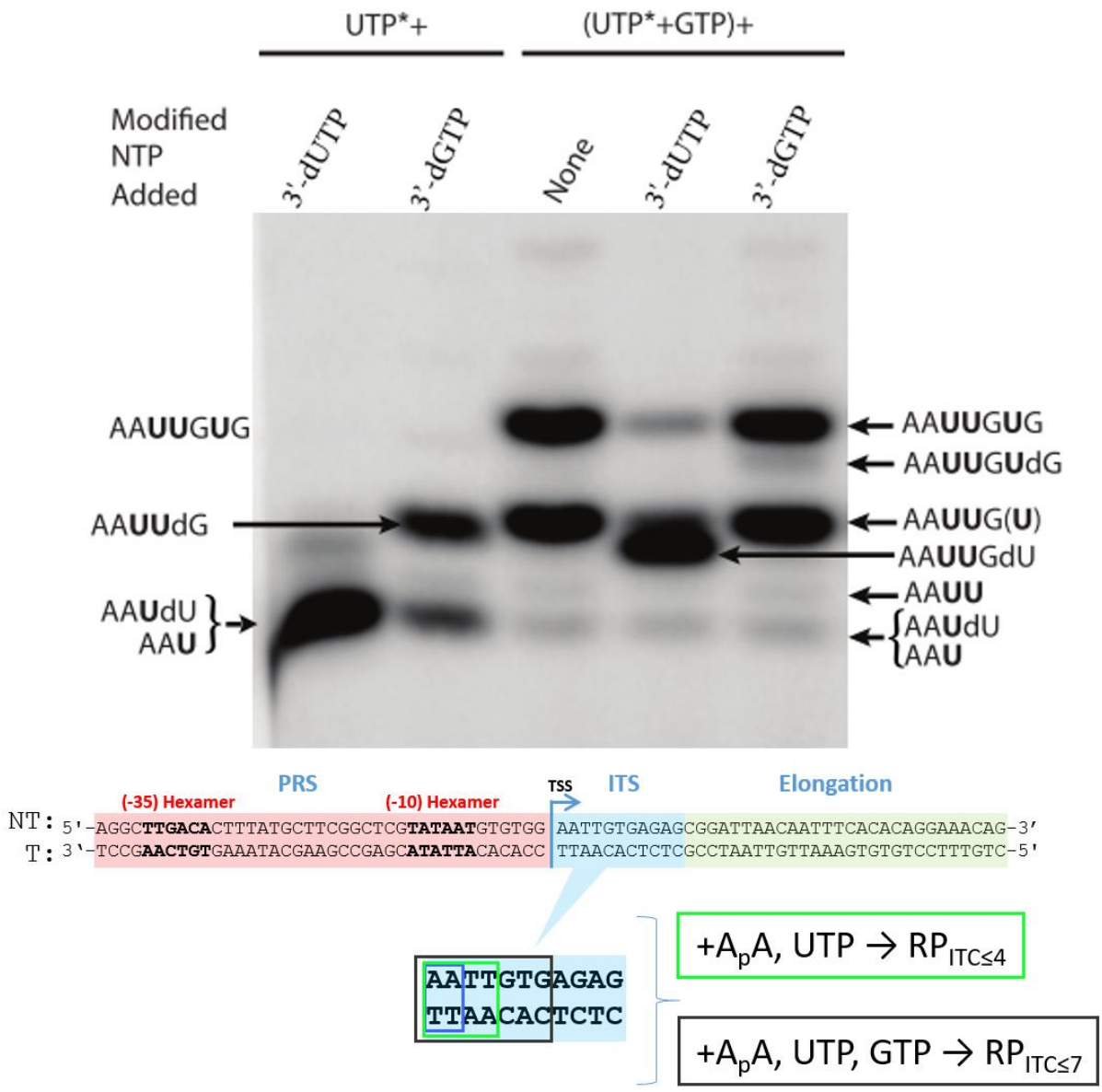



**Figure 18: Assignment of abortive products using NTP derivatives.** The left panel shows different stalled states with the addition of 3' deoxy nucleotides (3'-dNTPs). The 3'-dNTPs terminate the transcript upon incorporation and also migrate slightly faster, creating a doublet at

bands ending with that particular nucleotide. Therefore, the doublet in lane 5 of the left gel shows that the highest band ends in G, suggesting the highest band is indeed the 7-base product (AA**UU**G**U**G) that would arise from adding only UTP and GTP. Likewise, in lane 4, the top of the doublet in the middle must end in U, revealing that this band is the 6-base product (AA**UU**G**U**). In addition, in lane 2, the highest band is expected to correspond to the longest transcript terminated at the first G in the sequence (after supplementing the reaction only with UTP), suggesting the highest band is the 5-base product terminated by dGTP (AA**UU**dG). In this context, we conclude that the middle bands in lanes 3-5 of the left panel are a combination of the 5- (AA**UU**G) and 6-base products. It is not surprising that these abortive products migrate relatively close together because a single addition of U adds a relatively small amount of mass compared to the addition of a single G. In lane 2, in addition, the lowest two bands are the 3- and 4-base products (AA**U** and AA**UU**) expected to be produced upon adding only UTP.

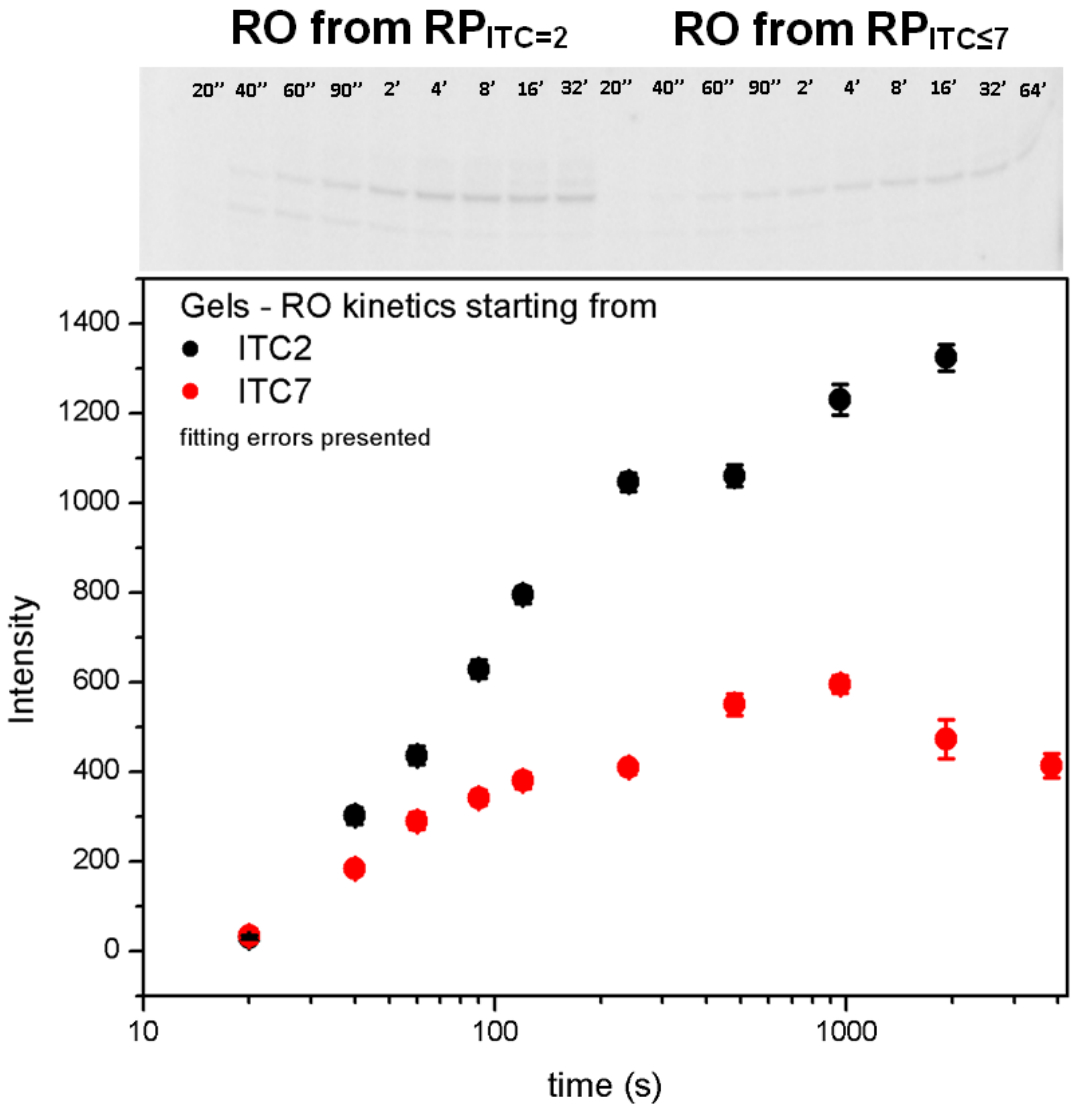



**Figure 19: Transcription run-off kinetics from RP$_{ITC=2}$ and RP$_{ITC\leq7}$ analyzed by $^{32}$P UREA-denaturing PAGE.** Exit (quenched) kinetics (bottom) from RP$_{ITC=2}$ (filled black circles) and RP$_{ITC\leq7}$ (filled red circles) are extracted from run-off gel band intensities (top). The assay does not reach single run condition (curves do not level off at long times) and therefore cannot quantitatively compare run-off kinetics from RP$_{ITC=2}$ and RP$_{ITC\leq7}$.

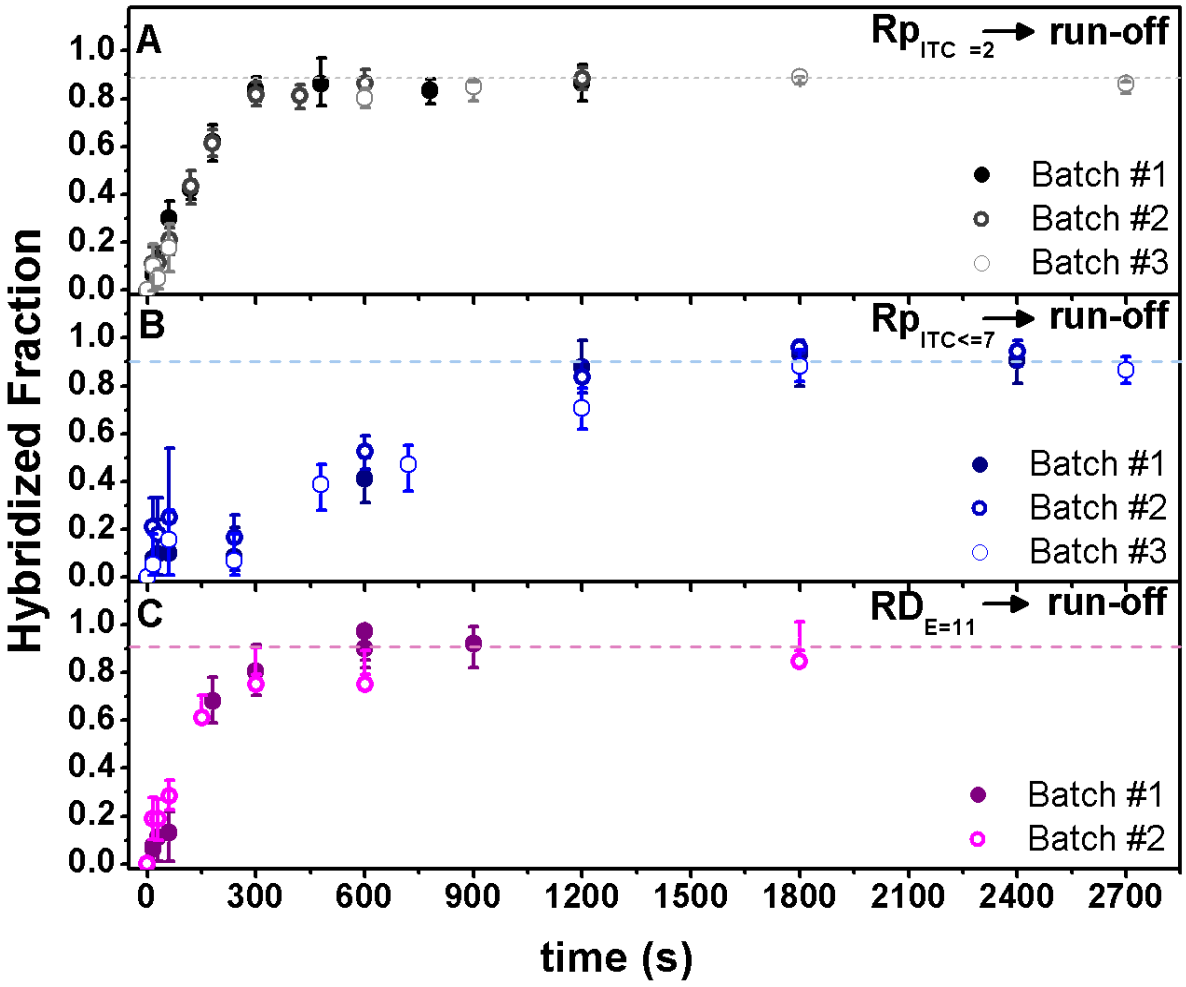



**Figure 20: Reproducibility of the quenched-kinetics assay.** Three representative examples of un-normalized and non-averaged kinetic data and their measurement repeats using the LacCONS promoter. All batches represent independent experiments performed on different days starting off from stock solutions. Error bars are derived from the fitting errors of the double Gaussian global fits of the FRET histograms (as in Fig. 1; with 95% confidence intervals). (**A**) Three experimental repeats of run-off kinetics from $RP_{ITC=2}$; (**B**) Three experimental repeats of run-off kinetics from $RP_{ITC\leq7}$ ; (**C**) Two experimental repeats of run-off kinetics from $RD_{E=11}$.

**Table 1: Best global fit values of transition rates and sub-population fractions using a simplified transcription model.**

| Promoter | GreA | Exit out of | $k_{on->off}$ ($10^{-3}$ $s^{-1}$) | $k_{off->on}$ ($10^{-3}$ $s^{-1}$) | $k_{init->elon}$ [a] ($10^{-3}$ $s^{-1}$) | on-pathway initial population | Figure/ panel/ color |
|---|---|---|---|---|---|---|---|
| *lac*CONS | - | $RP_{ITC=2}$ | - | - | 6±1 [c] | 1.00 (Const.) [d] | 7/black; 9/ top/black; 11/black |
| *lac*CONS | - | $RP_{ITC\leq4}$ | (0–1)[b] | 3.7±0.6 | | 0.20±0.06 | 7/red |
| *lac*CONS | - | $RP_{ITC\leq6}$ | (0–1) | 3.4±0.5 | | 0.15±0.05 | 7/green |
| *lac*CONS | - | $RP_{ITC\leq7}$ | (0–1) | 1.7±0.3 | | 0.07±0.03 | 7/blue; 9/ top/blue; 11/blue |
| *lac*CONS | - | $RD_{E=11}$ | - | - | 13±2 | 1.00 (Const.) | 7/magenta |
| *lac*CONS | + | $RP_{ITC=2}$ | - | - | 4±1 | 1.00 (Const.) | 9/ bottom/black |
| *lac*CONS | + | $RP_{ITC\leq7}$ | (0–1) | 4.1±0.9 | | 0.51±0.15 | 9/bottom/blue |
| T5N25 | - | $RP_{ITC=2}$ all NTPs eq. [e] | - | - | 16±3 | 1.00 (Const.) | 12/black |
| T5N25 | - | $RP_{ITC=2}$ NTP imbalance [f] | 18±5 | 26±8 | | 0.23±0.07 | 12/orange |
| *lac*CONS | - | $RP_{ITC=2}$ NTP imbalance | (0-1) | 6.5±0.8 | 6±1 | 0.28±0.11 | 11/orange |

[a] – initiation -> elongation

[b] – the best fit result was converging to a value of 0. The statistical upper bound value was to be the one reported in the parentheses

[c] – This parameter is globally shared in the model as one can assume the initiation -> elongation rate constant does not change for a given promoter in a given condition

[d] – the initial conditions are set to constant values in the case where by design there is no population of the off-pathway state at t=0, such as when starting from $RP_{ITC=2}$.

[e] – all NTPs at 100 μM

[f] – ATP and UTP at 100 μM and GTP and CTP at 2 μM

**Table 2: Best exponential fit values of entrance kinetics from $RP_{ITC=2}$ into several NTP starved states shown in Fig. 8**

| Entrance into[a] | $k^{b}$ ($s^{-1}$) | Figure/ panel/ color |
|---|---|---|
| $RP_{ITC\leq4}$ | $0.8\pm0.4 \times 10^{-3}$ | 8/red |
| $RP_{ITC\leq6}$ | $2.0\pm1.1 \times 10^{-2}$ | 8/green |
| $RP_{ITC\leq7}$ | $0.17\pm0.10$ | 8/blue |

[a] – from $RP_{ITC=2}$

[b] – this should represent the sum of $k_{on->off}$ and $k_{off->on}$

## References


1. Cooper, G. M. Eukaryotic RNA Polymerases and General Transcription Factors. (2000).
2. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).
3. Young, B. A., Gruber, T. M. & Gross, C. A. Views of transcription initiation. *Cell* **109,** 417–20 (2002).
4. Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* **13,** 31–9 (2003).
5. Zuo, Y. & Steitz, T. A. Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* **58,** 534–540 (2015).
6. Campbell, E. A. *et al.* Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Mol. Cell* **9,** 527–539 (2002).
7. Pupov, D., Kuzin, I., Bass, I. & Kulbachinskiy, A. Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* **42,** 4494–4504 (2014).
8. Strobel, E. J. & Roberts, J. W. Two transcription pause elements underlie a $\sigma^{70}$ -dependent pause cycle. *Proc. Natl. Acad. Sci.* **2015,** 201512986 (2015).
9. Kapanidis, A. N. *et al.* Retention of transcription initiation factor σ70 in transcription elongation: Single-molecule analysis. *Mol. Cell* **20,** 347–356 (2005).
10. Bae, B., Feklistov, A., Lass-napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 1–23 (2015). doi:10.7554/eLife.08504
11. Gezvain, K. . & Landick, R. in *The bacterial chromosome* 283–296 (2004).
12. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The Ratcheted and Ratchetable Structural States of RNA Polymerase Underlie Multiple Transcriptional Functions. *Mol. Cell* **57,** 1–14 (2015).
13. Epshtein, V. *et al.* Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol. Cell* **10,** 623–634 (2002).
14. Rutherford, S. T. *et al.* Effects of DksA, GreA, and GreB on Transcription Initiation: Insights into the Mechanisms of Factors that Bind in the Secondary Channel of RNA Polymerase. *J. Mol. Biol.* **366,** 1243–1257 (2007).
15. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7090–7095 (2000).
16. Mejia, Y. X., Nudler, E. & Bustamante, C. Trigger loop folding determines transcription rate of Escherichia coli's RNA polymerase. *Proc. Natl. Acad. Sci.* **112,** 743–748 (2015).

17. Benoff, B. *et al.* Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* **297,** 1562–6 (2002).

18. Vrentas, C. E., Gaal, T., Ross, W., Ebright, R. H. & Gourse, R. L. Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. *Genes Dev.* **19,** 2378–87 (2005).

19. Mathew, R. & Chatterji, D. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.* **14,** 450–5 (2006).

20. Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314,** 1144–1147 (2006).

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

22. Straney, D. C. & Crothers, D. M. Intermediates in transcription initiation from the E. coli lac UV5 promoter. *Cell* **43,** 449–59 (1985).

23. Hansen, U. M. & McClure, W. R. Role of the sigma subunit of Escherichia coli RNA polymerase in initiation. II. Release of sigma from ternary complexes. *J. Biol. Chem.* **255,** 9564–70 (1980).

24. Ellinger, T., Behnke, D., Bujard, H. & Gralla, J. D. 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,** 455–65 (1994).

25. Hsu, L. M. Monitoring Abortive Initiation. *Methods* **47,** 25–36 (2010).

26. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. *Science (80-. ).* **296,** 1285–1290 (2002).

27. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438,** 460–5 (2005).

28. Goldman, S. R., Nair, N. U., Wells, C. D., Nickels, B. E. & Hochschild, A. The primary σ factor in Escherichia coli can access the transcription elongation complex from solution in vivo. *Elife* **4,** e10514 (2015).

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

30. Feng, G. H., Lee, D. N., Wang, D., Chan, C. L. & Landick, R. 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,** 22282–94 (1994).

31. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22,** 6322–34 (2003).

32. Lodish, H. *et al.* Transcription Termination. (2000). at <http://www.ncbi.nlm.nih.gov/books/NBK21601/>

33. Fulton, A. B. How crowded is the cytoplasm? *Cell* **30,** 345–347 (1982).

34. Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **222,** 599–620 (1991).

35. Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37,** 375–97 (2008).

36. Hagen, S. J. Solvent viscosity and friction in protein folding dynamics. *Curr. Protein Pept. Sci.* **11,** 385–95 (2010).

37. Akabayov, B., Akabayov, S. R., Lee, S.-J., Wagner, G. & Richardson, C. C. Impact of macromolecular crowding on DNA replication. *Nat. Commun.* **4,** 1615 (2013).

38. Dupuis, N. F., Holmstrom, E. D. & Nesbitt, D. J. Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1316039111

39. Cheung, M. S., Klimov, D. & Thirumalai, D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 4753–8 (2005).

40. Vinella, D., Potrykus, K., Murphy, H. & Cashel, M. Effects on growth by changes of the balance between GreA, GreB, and DksA suggest mutual competition and functional redundancy in Escherichia coli. *J. Bacteriol.* **194,** 261–73 (2012).

41. Hsu, L. M., Vo, N. V & Chamberlin, M. J. Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11588–11592 (1995).

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

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

44. Chebotareva, N. a, Kurganov, B. I. & Livanova, N. B. Biochemical effects of molecular crowding. *Biochem. Biokhimiia* **69,** 1239–51 (2004).

45. Ellis, R. J. & Minton, A. P. Cell biology: join the crowd. *Nature* **425,** 27–8 (2003).

46. Borgia, A. *et al.* Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat. Commun.* **3,** 1195 (2012).

47. de Sancho, D., Sirur, A. & Best, R. B. Molecular origins of internal friction effects on protein-folding rates. *Nat. Commun.* **5,** 4307 (2014).

48. Siddiqui, K. S., Bokhari, S. A., Afzal, A. J. & Singh, S. A novel thermodynamic relationship based on Kramers Theory for studying enzyme kinetics under high viscosity. *IUBMB Life* **56,** 403–7 (2004).

49. Minton, A. P. How can biochemical reactions within cells differ from those in test tubes? *J. Cell Sci.* **119,** 2863–9 (2006).

50. Kuznetsova, I. M., Turoverov, K. K. & Uversky, V. N. *What macromolecular crowding can do to a protein*. *International Journal of Molecular Sciences* **15,** (2014).

51. Ge, X., Luo, D. & Xu, J. Cell-free protein expression under macromolecular crowding conditions. *PLoS One* **6,** (2011).

52. Sokolova, E. *et al.* Enhanced transcription rates in membrane-free protocells formed by coacervation of cell lysate. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11692–7 (2013).

53. Matsuda, H., Putzel, G. G., Backman, V. & Szleifer, I. Macromolecular crowding as a regulator of gene transcription. *Biophys. J.* **106,** 1801–10 (2014).

54. Demidenko, A. a, Lee, J., Powers, T. R. & Nibert, M. L. Effects of viscogens on RNA transcription inside reovirus particles. *J. Biol. Chem.* **286,** 29521–30 (2011).

55. Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38,** 523–33 (2005).

56. Kapanidis, A. N. *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,** 8936–41 (2004).

57. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437,** 55–75 (1948).

58. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 3670–5 (1999).

59. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **86,** 2530–7 (2004).

60. Brinkers, S., Dietrich, H. R. C., de Groote, F. H., Young, I. T. & Rieger, B. The persistence length of double stranded DNA determined using dark field tethered particle motion. *J. Chem. Phys.* **130,** 215105 (2009).

61. Abels, J. A., Moreno-Herrero, F., van der Heijden, T., Dekker, C. & Dekker, N. H. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* **88,** 2737–44 (2005).

62. Vogel, S. K. Binding affinity of Escherichia coli RNA polymerasemiddle dotsigma54 holoenzyme for the glnAp2, nifH and nifL promoters. *Nucleic Acids Res.* **30,** 4094–4101 (2002).

63. Tang, G.-Q., Roy, R., Bandwar, R. P., Ha, T. & Patel, S. S. Real-time observation of the

transition from transcription initiation to elongation of the RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 22175–22180 (2009).

64. Ko, J. & Heyduk, T. Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. *Biochem. J.* **463,** 135–144 (2014).

65. Dangkulwanich, M., Ishibashi, T., Bintu, L. & Bustamante, C. Molecular mechanisms of transcription through single-molecule experiments. *Chem. Rev.* **114,** 3203–3223 (2014).

66. Kuga, S. Pore size distribution analysis of gel substances by size exclusion chromatography. *J. Chromatogr. A* **206,** 449–461 (1981).

67. Devanand, K. & Selser, J. C. Asymptotic behavior and long-range interactions in aqueous solutions of poly(ethylene oxide). *Macromolecules* **24,** 5943–5947 (1991).

68. SCHULTZ, S. G. & SOLOMON, A. K. Determination of the effective hydrodynamic radii of small molecules by viscometry. *J. Gen. Physiol.* **44,** 1189–99 (1961).

69. Panzeri, F. *et al.* Single-molecule FRET experiments with a red-enhanced custom technology SPAD. *Proc. SPIE--the Int. Soc. Opt. Eng.* **8590,** (2013).

70. Nir, E. *et al.* Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B* **110,** 22103–24 (2006).

71. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86,** 163–80 (2001).

72. Michalet, X. *et al.* Development of new photon-counting detectors for single-molecule fluorescence microscopy. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120035 (2013).

73. Ingargiola, A., Lerner, E., Chung, S., Weiss, S. & Michalet, X. *FRETBursts: Open Source Burst Analysis Toolkit for Confocal Single-Molecule FRET. bioRxiv* (Cold Spring Harbor Labs Journals, 2016). doi:10.1101/039198

74. Basu, R. S. *et al.* Structural Basis of Transcription Initiation by Bacterial RNA Polymerase holoenzyme. *J. Biol. Chem.* **289,** M114.584037– (2014).

75. Nudler, E., Avetissova, E., Markovtsov, V. & Goldfarb, A. Transcription processivity: protein-DNA interactions holding together the elongation complex. *Science* **273,** 211–7 (1996).

76. Lee, D. N. & Landick, R. Structure of RNA and DNA chains in paused transcription complexes containing Escherichia coli RNA polymerase. *J. Mol. Biol.* **228,** 759–77 (1992).

77. Korzheva, N., Mustaev, A., Nudler, E., Nikiforov, V. & Goldfarb, A. Mechanistic model of the elongation complex of Escherichia coli RNA polymerase. *Cold Spring Harb. Symp. Quant. Biol.* **63,** 337–45 (1998).

78. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci.* **95,** 14699–14704 (1998).

79. TRAVERS, A. A. & BURGESS, R. R. Cyclic Re-use of the RNA Polymerase Sigma

Factor. *Nature* **222,** 537–540 (1969).

80. Krummel, B. & Chamberlin, M. J. RNA chain initiation by Escherichia coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry* **28,** 7829–7842 (1989).

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

82. Straney, D. C. & Crothers, D. M. A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *J. Mol. Biol.* **193,** 267–78 (1987).

83. Winkelman, J. T. *et al.* Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol. Cell* **59,** 768–80 (2015).

84. Samanta, S. & Martin, C. T. Insights into the mechanism of initial transcription in Escherichia coli RNA polymerase. *J. Biol. Chem.* **288,** 31993–2003 (2013).

85. Margeat, E. *et al.* Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys. J.* **90,** 1419–31 (2006).

86. Bar-Nahum, G. *et al.* A ratchet mechanism of transcription elongation and its control. *Cell* **120,** 183–93 (2005).

87. Holmes, S. F. & Erie, D. A. Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J. Biol. Chem.* **278,** 35597–608 (2003).

88. Foster, J. E., Holmes, S. F. & Erie, D. A. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106,** 243–52 (2001).

89. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Mol. Cell* **24,** 747–57 (2006).

90. Erie, D. A., Hajiseyedjavadi, O., Young, M. C. & von Hippel, P. H. Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **262,** 867–73 (1993).

91. Toulmé, F. *et al.* GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* **19,** 6853–9 (2000).

92. Sen, R., Nagai, H. & Shimamoto, N. Conformational switching of Escherichia coli RNA polymerase-promoter binary complex is facilitated by elongation factor GreA and GreB. *Genes Cells* **6,** 389–401 (2001).

93. Stepanova, E. *et al.* Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *J. Bacteriol.* **189,** 8772–8785 (2007).

94. Stepanova, E. V., Shevelev, A. B., Borukhov, S. I. & Severinov, K. V. Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. *Biophysics (Oxf).* **54,** 555–568 (2010).

95. Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts in vivo. *Science* **324,** 927–8 (2009).

96. Mukhopadhyay, J. *et al.* Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106,** 453–63 (2001).

97. Kim, S. *et al.* High-throughput single-molecule optofluidic analysis. *Nat. Methods* **8,** 242–5 (2011).

98. Ingargiola, A. *et al.* 8-spot smFRET analysis using two 8-pixel SPAD arrays. *Proc. SPIE--the Int. Soc. Opt. Eng.* **8590,** (2013).

99. Ingargiola, A. *et al.* Parallel multispot smFRET analysis using an 8-pixel SPAD array. in *Proceedings of SPIE* **8228,** 82280B–82280B–8 (2012).

100. Borukhov, S., Lee, J. & Laptenko, O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* **55,** 1315–24 (2005).

101. Sousa, R. A new level of regulation in transcription elongation? *Trends Biochem. Sci.* **26,** 695–697 (2001).

102. Guglielmi, B., Soutourina, J., Esnault, C. & Werner, M. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16062–16067 (2007).

103. Kim, B. *et al.* The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16068–16073 (2007).

104. Hatoum, A. & Roberts, J. Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. *Mol. Microbiol.* **68,** 17–28 (2008).

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

106. Stepanova, E. *et al.* Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *J. Bacteriol.* **189,** 8772–85 (2007).

107. Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A. & Allain, F. H.-T. Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.* **21,** 269–76 (2014).

108. Hsu, L. M. *et al.* Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry* **45,** 8841–8854 (2006).

109. Zhang, G. *et al.* Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98,** 811–24 (1999).

110. Foster, J. E., Holmes, S. F. & Erie, D. a. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106,** 243–52 (2001).

111. Orlova, M., Newlands, J., Das, a, Goldfarb, a & Borukhov, S. Intrinsic transcript

cleavage activity of RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 4596–4600 (1995).

112. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. Ratcheting of RNA polymerase toward structural principles of RNA polymerase operations. *Transcription* **6,** 56–60 (2015).

113. Zhang, N. *et al. Mutations in RNA polymerase bridge helix and switch regions affect active site networks and transcript-assisted hydrolysis*. *Journal of Molecular Biology* **427,** (Elsevier B.V., 2015).

114. Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *J. Bacteriol.* **190,** 718–26 (2008).

115. Kusuya, Y., Kurokawa, K., Ishikawa, S., Ogasawara, N. & Oshima, T. Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in Bacillus subtilis cells. *J. Bacteriol.* **193,** 3090–3099 (2011).

116. Susa, M., Sen, R. & Shimamoto, N. Generality of the branched pathway in transcription initiation by Escherichia coli RNA polymerase. *J. Biol. Chem.* **277,** 15407–12 (2002).

117. Kubori, T. & Shimamoto, N. A branched pathway in the early stage of transcription by Escherichia coli RNA polymerase. *J. Mol. Biol.* **256,** 449–57 (1996).

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

119. Revyakin, A., Ebright, R. H. & Strick, T. R. Single-molecule DNA nanomanipulation: improved resolution through use of shorter DNA fragments. *Nat. Methods* **2,** 127–38 (2005).