The Kinetics and Mechanism of *E. coli* Transcription Initiation

By

Dylan Plaskon

A dissertation submitted in partial fulfillment of

The requirements for a degree of

Doctor of Philosophy

(Biochemistry)

At the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 04/13/2022

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

M. Thomas Record Jr., Professor, Biochemistry

Robert Landick, Professor, Biochemistry

Srivatsan Raman, Assistant Professor, Biochemistry

Katherine Henzler-Wildman, Professor, Biochemistry

Anthony Gitter, Professor, Biostatistics and Medical Informatics

Table of Contents

Acknowledgements

This thesis marks the final chapter of an educational process throughout which I have been truly blessed to be the recipient of so much encouragement and support. There is no question that without the family, mentors, educators, and peers that have gone above and beyond to support me during this process I would not have been able to reach this point. There are far too many people to be able to thank them all, but I would like to extend a special thanks to the following people.

First and foremost, I would like to thank my advisor Prof. Tom Record, who agreed to let me join his lab midway through my career as a graduate student and has been a wonderful and supportive mentor. Thank you also to my committee members, Profs. Robert Landick, Anthony Gitter, Katherine Henzler-Wildman, and Vatsan Raman, who also served as my advisor for the first two years of my graduate career, for their support and contributions through my graduate career.

I would also like to thank the past and current members of the Record lab, who were always willing to lend a hand or an ear. In particular, I owe so much to Dr. Kate Henderson for her friendship and mentorship.

I would like to thank the undergraduate and post-baccalaureate researchers who worked with me in the Record lab and contributed invaluable efforts to the projects in this thesis and in the lab: Taka Ishikuri, Sarah Doughty, Tristan Gunther, Will Langholz, Quinn McBride, Savannah Peterson, Claire Evensen, Jack Prazich, Ben Palatnik, Sanj Nair, Sumin Yang, and Maddie Boleyn.

Prof. Kevin Rice, Dr. Owen Murphy, and Joel Bates are among the many wonderful teachers and professors who fostered my love of learning and of science during my undergraduate and early education. Each of them went above and beyond to offer me opportunities and support without which I would not have gotten where I am today, for which I am very grateful.

I would like to thank my peers in the IPiB graduate program and in Madison, who have become wonderful friends and whose support means so much to me.

I would like to extend my special thanks to Kate Ryan, who has very patiently helped me navigate the administrative side of graduate school.

Finally, I would not be here if it were not for the unconditional support of my family, who never stopped supporting or believing in me. I would like to thank Sophia Sdao, who has been the most incredibly patient and understanding partner throughout this process, and whose support has meant the world to me.

Abstract

Transcription of relatively static genetic information stored in DNA into RNA used as a protein template, regulator, or even a catalyst is a central process in all known organisms. Precise regulation of transcription is therefore critical. Transcription initiation is highly regulated by promoter sequence, transcription factors, and ligands. This regulation is underpinned by the specific promoter sequence and the contacts it makes with RNA Polymerase (RNAP). To understand initiation, the basic mechanism of initial transcription, and the effects on that mechanism of open complex (OC) stability and promoter element sequences and lengths, must be understood. Here we use a kinetic-mechanistic approach to dissect the stepwise mechanism and rate constants for the model λP_R promoter. From the overall temperature dependence of the initiation rate we determine that an intermediate and not the stable form of the λP_R open complex initiates transcription. From the temperature dependence of the stepwise rate constants we determine that disruption of RNAP-promoter contacts occurs in three groups of initiation steps, and that collapse of the initiation bubble is also stepwise. We extend our studies of the λP_R promoter to determine the effects of the RNA synthesis byproduct pyrophosphate (PPi) on initiation kinetics. We determine that physiological levels of pyrophosphate play a significant role in initiation kinetics, and that the effect of PPi on individual RNA extension steps is sequence dependent, as in elongation. Through studies of promoters with the *rrnB* P1 discriminator we assess the effects of the discriminator on initiation. We determine and quantify the dependence of initiation kinetics on the second initiating NTP (iNTP), and find that priming initiation at these promoters with dinucleotides can overcome this dependence. We investigate the kinetics of initiation from hybrid

promoters containing the 7 bp T7A1 discriminator and determine from the stepwise initiation rate constants that disruption of RNA polymerase-promoter contacts occurs in two groups of initiation steps. We also determine the distribution of discriminator lengths in the *E. coli* genome. Finally, we discuss the significance of the research reported here and suggest future directions for determining the effects of promoter sequence and architecture on initiation kinetics and thermodynamics.

Chapter 1: Introduction

Motivation

Transcription of relatively static genetic information stored in DNA to RNA is a process fundamental to all known organisms [1, 2], and the RNA product plays a role in a diverse set of cellular process. Some examples of the vast array of RNA functions include: as the direct template for protein production (mRNA) [3], as all or part of a catalyst for a cellular process (ribozymes, the most famous of which is the ribosome) [4], and as a regulatory tool (e.g. sRNAs) [5]. New specific functions for RNA ranging in length from two to thousands of nucleotides are still being discovered [6]. The survival and success of an organism relies on the finely tuned interplay of all the processes within a cell, and appropriate regulation of those processes is of paramount importance. Transcription is therefore subject to layers of regulation that work together to control the RNA expression profile of the cell [7, 8].

There are three distinct phases of transcription: initiation, elongation, and termination. At an overview level, in initiation the transcription machinery recognizes and binds promoter DNA and begins transcribing RNA [9]. In elongation, the machinery moves processively along the DNA template without dissociating, extending the RNA [10]. In termination, transcription is halted by termination sequences or alternative means, and the RNA is released [11]. Each of these phases is subject to regulation evolved to finely control the transcript levels of each gene. Initiation in particular is critically positioned at the beginning of the gene expression cascade, in which the genetic information encoded in a gene is amplified into hundreds to tens of thousands of biomolecules [12].

Regulation of transcription initiation is layered and complex. Transcription factors, some of which are gene-specific and some of which are general, are present in all known

organisms [13]. All transcription factors act in some manner to control the activity of RNA polymerase (RNAP), which catalyzes the synthesis of RNA using DNA as the template [14]. This core function of RNAP is highly conserved across all domains of life, although the complexity and specificity of the enzyme increases with the complexity of the organism [14]. In phages, RNAP is a single-subunit enzyme with minimal regulation [15, 16]. Humans however have multiple, multi-subunit RNAPs, each of which is dedicated to transcribing a different type of RNA [17-19]. The multi-subunit organization of all bacterial, archeal, and eukaryotic RNAP is essential for regulation of all stages of transcription. The evolutionary divergence of RNAP in bacteria and more complex organisms like humans, alongside its importance, makes it an excellent target for antibiotics. Rifampicin and lipiarmycin, among others, are examples of clinical antibiotics that target bacterial RNAP [20, 21].

There is no inherent property of intragenic DNA that causes it to be transcribed instead of intergenic DNA. Rather, all genes have a promoter which at minimum acts as the RNAP recognition site upstream of the transcribed DNA [22]. Promoters within an organism have several common elements, but the sequence and architecture of the promoters vary widely [23]. This diversity results in a range of promoter efficacies even in the absence of other regulatory mechanisms such as transcription factors [24, 25]. This sequence-based regulation can be considered the basal level of initiation regulation, on top of which all other regulation is layered, allowing the cell very fine control of gene expression under difference conditions and life stages. Even in the simplest organisms, there are many questions remaining as to the mechanisms by which different promoter features regulate initiation efficacy. Without fully understanding and characterizing these

basic mechanisms, it is not possible to fully understand gene regulation at the organismal level. Bacterial RNAP is an excellent model for studying initiation because is has many fewer subunits than eukaryotic RNAP but is nevertheless highly regulated [26]. Additionally, important practical advances such as novel antibiotics have come from mechanistic studies of bacterial RNAP, the best studied of which is from the common bacterial model organism *E. coli* [27].

Promoter Sequence and Architecture

The basic function of an *E. coli* promoter is to specifically bind RNA polymerase, which then uses that binding energy to direct opening of the transcription initiation bubble [9, 28-32]. The interactions between RNAP and promoter DNA underlie most initiation regulation within *E. coli*, and so understanding these interactions is crucial. The elements of *E. coli* promoters have evolved to match the modular promoter recognition factor σ specific for that promoter class with high specificity [33], while allowing enough variation from promoter to promoter that initiation efficacy can vary widely. For σ^{70} promoters, many of these elements and their consensus sequences and functions are well known (Figure 1).

The core *E. coli* promoter elements are: the UP element (approximately 60 to 40 bp upstream of the transcription start site) [34] the -35 and -10 sigma recognition hexamers [35, 36], the spacer separating the -10 and -35 elements [37], and the discriminator separating the -10 region and transcription start site (+1, TSS) [38]. Most known σ^{70} promoters have some version of all these elements, while other more specific elements such as the core recognition element (CRE) [39] and extended -10 element [40] exist in a subset of promoters for more specific regulation.

The main RNAP recognition elements are the -35 and -10 hexamers, named based on their approximate distance from the TSS [35]. The -10 element is the most well-conserved promoter element, with a consensus non-template strand sequence 5'-TATAAT-3' [35, 36]. Despite the strong consensus, some deviations from this sequence are common [41]. These deviations are not equally likely to occur at each position [41]. High resolution open RNAP-promoter complex structures alongside biochemical studies have demonstrated that the nontemplate strand -11A likely nucleates strand separation to form the initiation bubble by unstacking and "flipping" into a binding pocket in the RNAP [32, 42, 43]. This melting is mandatory for initiation and therefore the -11T is highly conserved among σ^{70} promoters [41]. Other positions within the -10 hexamer are less well conserved, and these sequence differences are consequential. The -7T is the second most conserved base in the -10 element, and high-throughput studies on the sequence variations of the -10 element indicate that this position has an even larger effect than the -11A on initiation efficacy [25].

The -35 hexamer has a consensus sequence of 5'-TTGACA-3', although it is less well conserved than the -10 region [35]. Like the -10 element, the -35 sequence is specifically recognized by the σ^{70} RNAP holoenzyme [44]. Unlike the -10 element, the -35 element remains duplex DNA throughout initiation. The -35 element also acts as the downstream end of the site of bending of promoter DNA around RNAP during the initial stages of initiation bubble opening [28, 44, 45].

The -10 and -35 elements are separated by the spacer region. There is no strong consensus sequence for the spacer, although there is a consensus length of 17 bp [35, 37, 46]. Deviations from this length impact promoter efficacy [47-49] and TSS selection

[50]. In select promoters, the “extended -10” element is directly upstream of the -10 element within the spacer region [37, 40]. This element is not present on all σ^{70} promoters. It is most important at promoters with weak or missing -35 elements or longer spacers, supplementally stabilizing the promoter-RNAP complex alongside or in lieu of the important recognition hexamer [45].

The UP element is an AT-rich region upstream of the -35 element [51]. It is a binding site for the C-terminal domain of the α-subunits of RNAP (α-CTDs) [52]. Mutational studies indicate that presence of the consensus UP element can increase promoter efficacy more than 300-fold [52, 53].

The region of DNA between the transcription start site (TSS) and the -10 element is the discriminator [38]. The most common discriminator length is 6 bp, and the stability of the RNAP-promoter open complex (OC) tends to decrease with increasing discriminator length [50, 54-57]. Although there is no strong consensus sequence for the discriminator, some sequence-specific contacts are made between the discriminator and RNAP. In a 6 bp discriminator a nontemplate strand -6G will flip into a pocket formed by $\sigma_{1.2}$ [42]. At position -5, a nontemplate strand C is generally destabilizing and a nontemplate strand G is generally stabilizing [38]. The promoters for ribosomal RNA, which generally form very unstable initiation complexes, contain a nontemplate strand C two bases downstream of the -10 element, which contributes to this instability [38].

Although not canonically classified as a promoter element the first bases to be transcribed comprise the initial transcribed region (ITR), and also play an important role in the kinetics and mechanism of promoter escape [24, 58-61]. For a large majority of promoters, the template strand sequence at the TSS is a pyrimidine followed by a purine,

so the initiating NTP (iNTP) are the complementary purine and pyrimidine. Sequence elements in the ITR can cause pausing during initial transcription [62-64] as can steric impediments to initial transcription [65]. It is the interplay of all the elements mentioned here that define the basal activity of a promoter before promoter-specific transcription factors layer their regulation on top of this baseline. Given the huge variety of unique promoters within the cell, it is evident that promoter sequence is a key regulatory tool used by the cell to consistently maintain appropriate transcript levels.

The Structure of *E. coli* RNAP

In *E. coli*, all transcription is carried out by a multi-subunit RNAP comprised of a core of five subunits: two α-subunits, β and β' subunits, and an ω subunit [66]. In assembly of core RNAP, an α_2 dimer serves as the platform for addition of the large β and β' subunits. The 3-dimensional structure of a core bacterial RNAP enzyme was first determined for the enzyme from the thermophilic bacterium *T. aquaticus* in 1999 via x-ray crystallography [67]. Since then, crystallographic studies and the advent of powerful structural imaging techniques such as Cryo-EM have produced highly detailed structures of bacterial RNAP-promoter complexes bound to different promoters and at many different stages of initiation [42, 44, 65, 68-72]. The RNAP core forms a characteristic structure resembling a crab claw (Figure 2), where the template strand of the single stranded DNA of the transcription bubble is fit into the cleft formed primarily by β and β', and the base to be transcribed is situated in the active site, which contains a bound catalytic Mg^{2+} [73]. This core is sufficient to catalyze RNA extension in the presence of NTPs [10, 73].

Although the core RNAP enzyme is sufficient for elongation, initiation must be facilitated by an additional, modular RNAP subunit σ [74]. *E. coli* produces several different σ factors, each corresponding to a different subset of promoters. In this way, by regulating the production of various σ factors, the cell can respond transcriptionally to a variety of conditions including nutrient starvation, heat or chemical stress, and more [75]. The most abundant (and best studied) version during normal growth is σ^{70}, which upon association with RNAP core enzyme the RNAP holoenzyme [9]. All studies reported in this thesis were performed with the σ^{70} holoenzyme. Unless otherwise stated, all references made to RNAP refer to the RNAP σ^{70} holoenzyme.

The Initiation Mechanism

Overview

At the broadest level, *E. coli* transcription initiation can be described by a simple mechanism whereby RNAP recognizes and binds to promoter DNA and forms a closed complex (CC). Then, the initiation bubble is opened to form an open complex (OC), which binds the initiating NTP and begins RNA synthesis. Translocation of the RNA-DNA hybrid causes the buildup of translocation stress and breaks RNAP-promoter contacts, allowing RNAP to escape from the promoter and transition into the elongation phase [9, 76, 77]. This simplistic model is useful but belies the true complexity of the initiation mechanism, which is still being determined using traditional biochemical as well as single molecule and structural techniques [28, 62-65, 78-84]. The complexity of the initiation mechanism provides unparalleled opportunity for precise regulation of initiation within the cell.

The Closed Complex Ensemble

Initial recognition and binding of a promoter by RNAP is directed by interactions between the -10 and -35 elements with $\sigma_{2.4}$ and $\sigma_{4.2}$ respectively [32, 42, 45, 85]. Additional contacts between the RNAP α-CTDs, which are attached to the polymerase core by flexible tethers, and binding tracts on the promoter UP element facilitate formation of the first stable intermediate of the closed complex ensemble, RP_c [28, 51]. Initial studies support promoter discovery by RNAP sliding (one-dimensional diffusion along promoter DNA) [86-90], although single-molecule studies support a direct-binding mechanism [91, 92]. Initial binding of RNAP may also involve transfer from concentrated RNAP condensates that introduce an additional level of organization in the vicinity of highly-transcribed promoters [93, 94].

Mechanisms of the conversion of the initial RP_C closed complex to an open complex are being determined for several different promoters, including λP_R, T7A1, lacCONS by combinations of kinetic-mechanistic and structural studies that identify and characterize intermediates on the pathway of promoter opening [28, 95-97]. Key differences in how the DNA is opened are observed for different promoters. At the λP_R promoter, three closed intermediates after RP_C are detected on the pathway to OC formation using Fluorescence Resonance Energy Transfer (FRET) and Protein Induced Fluorescence Enhancement (PIFE), designated RP_c, I_{1E}, I_{1M}, and I_{1L} for early, mid- and late intermediates (Figure 3) [9, 28, 98-100]. In this pathway, upstream DNA is bent and tightly wrapped around RNAP in I_{1E}. Bending and wrapping is proposed to start in the vicinity of the UP element and involve a conformational change to compact the flexible tethers between α-NTD and α-CTD domains [28, 101]. Subsequent opening of the clamp (via movement of the β' jaw, facilitated by the bending/wrapping of upstream DNA around

RNAP) forms I_{1M}. Formation of I_{1L} occurs when the DNA is loaded into the cleft and the clamp is closed again [28]. Conversion of I_{1L} to the initial unstable open complex I_2 at λP_R opens 13 bp of promoter DNA, including the TSS, in a single kinetic step and is accompanied by clamp opening, which is the cause of the instability of I_2. Subsequent conformational changes involving the clamp and both in-cleft and downstream mobile elements of RNAP increase the stability of the open complex greatly, forming at least two more stable OC (I_3, RP_O) as described below [9, 102].

A series of RNAP-promoter complex structures at the *rpsT* P2 promoter were captured with the transcription factor TraR, which increases the population of these intermediates [84, 103-105]. In this system, only one fully closed complex was reported, although several partially open intermediates were seen. At the *rrnB* P1 promoter, structures of the RP_C with and without regulators DksA and ppGpp were interpreted as intermediates in a branched mechanism [68]. In the pathway without DksA and ppGpp regulation, duplex DNA was loaded into the cleft prior to initiation bubble opening. With DksA and ppGpp bound, the bubble opening was predicted to occur prior to loading [68]. Closed complex structures corroborating the DNA wrapping detected via footprinting and FRET/PIFE [28, 101] have not yet been obtained.

The Open Complex Ensemble

Conversion of the final closed complex intermediate (I_{1L} at λP_R [28]) to a stable open complex (RP_O) requires the melting of duplex DNA to form the initiation bubble. The upstream edge of the bubble is at position -12 (assuming a 6 bp discriminator), which remains as duplex DNA and contacts σ_2 [32, 44, 106, 107]. Structural [84] and biochemical [108] studies indicate the existence of a very unstable initial open complex

which may be only partially open. Melting nucleates at the -11A, which unstacks from the adjacent bases into the σ_2 pocket [32, 109]. The exact mechanism for DNA unwinding and melting after this point is not fully known, however the result is a fully open initiation bubble [42]. The rate of this DNA opening step does not vary greatly with salt concentration [110], in contrast to a similar length of DNA melting in solution [111, 112]. This supports a model of OC formation where DNA is melted after being loaded into the cleft, and so is not exposed to solution [9, 113].

Although all promoters must form some sort of open complex to initiate, the conformation and stability of that OC is highly variable [56, 102]. The OC lifetime at different promoters can vary over three to four orders of magnitude [9]. A variety of factors contribute to OC lifetime and stability. For example at the λP_R promoter contacts between downstream mobile elements (DMEs) of RNAP and downstream DNA stabilize RP_O, the most stable OC form at 37 °C [114]. The strength of RNAP contacts with individual promoter elements, such as the -10 element and CRE, can stabilize the OC as well [25, 39, 115]. The presence of negative supercoiling at a promoter is also predicted to make opening the DNA duplex less thermodynamically costly, and thus stabilize the OC [116].

A major determinant of OC lifetime is discriminator length [9, 50, 56, 117]. Due to the flipping of the -11A into a pocket in σ_2, the upstream edge of the transcription bubble is tightly held. However, the distance between this contact and the RNAP active site, where the TSS must be aligned for transcription to begin, is relatively fixed and is evolved to comfortably fit a 6 bp discriminator [44]. At longer discriminators, promoter DNA must be “pre-scrunched” into cleft, inducing energetically unfavorable distortions of about 1 kcal mol^{-1} bp^{-1} scrunched [56, 57]. This distortion is predicted to drastically destabilize the

open complex, which is correlated with faster and earlier promoter escape [56, 58, 59, 118].

Initial Transcription and Promoter Escape

After productive OC formation, the first initiation step is binding of the first two NTPs (iNTPs) and formation of the initiating dimer. The first nucleotide added is a purine (A or G) at the overwhelming majority of *E. coli* σ^{70} promoters [55, 119]. Formation of 2-mer is unique because it is the only step that does not require RNAP to translocate relative to the DNA strand prior to NTP binding and catalysis. Instead, structural and biochemical evidence at multiple promoters indicate the iNTP sites are positioned within the active site to catalyze 2-mer formation without the need to translocate [120, 121]. After 2-mer formation, each NTP addition step is a repeated cycle of translocation, NTP binding, and catalysis, with the next cycle starting with translocation [10, 122]. In elongation, this process can be continuously repeated for thousands of bases, although RNAP can move into “off-pathway” paused or backtracked states for the purposes of accuracy, regulation, or termination [10, 123, 124].

Each NTP addition cycle includes the unfolding and refolding of the α-helical trigger element located near the binding site of the incoming NTP into the trigger loop, and back [44, 72, 125-127]. The trigger helix is unfolded (forming the trigger-loop) in the post-translocated state, which allows binding of the i+1 NTP, folding of the trigger loop properly positions that NTP in the active site for RNA extension, acting as a positional catalyst [126, 128]. Trigger loop folding is not strictly mandatory for RNA extension, but increases the nucleotide addition rate in elongation by up to 10^5 [125, 127]. In *E. coli*, a 188 bp

insertion domain SI3 which contacts the β' jaw is located within the trigger loop, deletion of which reduces pausing in elongation [127].

In the presence of inorganic pyrophosphate (PPi), which is the byproduct of nucleotide addition, the nucleotide addition cycle (including translocation) is reversible indicating an equilibrium between pre- and post- translocated states of the RNAP-promoter complex [122, 129]. Elongation translocation models based on base-pairing energetics of the upstream and downstream edges of the bubble, as well as the RNA-DNA hybrid, predicted a pre-translocated state bias [130-132]. More recent models that incorporate favorable interactions between the folded trigger helix and the 3' RNA end predict a bias for the post-translocated state in elongation, which better explains the results of pyrophosphorolysis studies in elongation [122, 129].

In initiation, each repeat of the nucleotide addition cycle involves a translocation step. However the RNAP contacts made with the promoter to open the DNA and form the open complex remain in place, at least in the first step(s) of initiation. Therefore each translocation event in initiation, using the free energy of NTP binding and/or triphosphate cleavage, builds up translocation stress in RNAP-promoter complex [83, 133, 134]. Translocation stress can be stored as scrunching of the discriminator strands and/or can be used to break specific promoter-DNA contacts. Once enough energy is stored, the promoter contacts are broken and the complex moves into elongation [56, 83, 120].

The translocation stress introduced by scrunching biases initiation complexes to the pre-translocated state [120]. Scrunching translocation stress should be cumulative, in that each additional base scrunched should make subsequent steps more difficult, and thus slower. However, bulk kinetic studies at the λP_R promoter showed that rate constants

to the point of escape show a periodic increase and decrease with RNA length, interpreted as the stepwise, rather than concerted, breaking of RNAP-promoter contacts on the pathway to promoter escape [120]. Stepwise breaking of promoter contacts is also seen during promoter escape by T7 phage polymerase [135].

In addition to breaking promoter contacts, σ must be released and the initiation bubble must collapse to move from initiation to elongation. The mechanism of bubble collapse was not known, and bubble collapse was thought to be a significant determinant of escape efficacy under certain circumstances. There are 11 bp of ssDNA in the elongation bubble, and 13 bp of ssDNA in the initiation bubble at a promoter with a 6 bp discriminator [44, 125]. At the λP_R promoter, which escapes after production of a 10-mer RNA in initial transcription [56, 120], 8 additional bp of DNA must be opened and so a maximum of 10 bp must re-form duplex DNA between initiation and elongation. Although such an enthalpically favorable process might be expected to thermodynamically drive promoter escape, abolition of bubble collapse by initiating from a heteroduplex has no effect on initiation kinetics at *E. coli* promoters [118]. Release of σ is not thought to be a concerted event, but a stochastic one, and σ can be retained well into elongation and can induce pausing [82, 136-138].

Abortive RNA Synthesis

Productive RNA synthesis, in which RNAP successfully escapes the promoter, is not the whole story. At many promoters, small RNA species are reiteratively transcribed during initiation, referred to as abortive RNA [56, 59, 61, 82, 133, 139-147]. Two general mechanisms have been proposed to describe the origin of these complexes. The linear mechanism posits that RNAP has several “false starts” before escaping the promoter,

transcribing a finite amount of abortive RNA [133]. The branched mechanism posits that in a population of OC, there are at least two distinct classes: one class of OC that is capable of escaping the promoter and does so immediately without any reiterative transcription, and one that is incapable of escaping the promoter and produces abortive RNA indefinitely. Studies of multi-round transcription cannot distinguish between the two mechanisms. Studies of single-round initiation unambiguously support the branched mechanism of initiation [56, 120, 147].

Summary

In this thesis we determine the kinetics and mechanism of initiation from several model promoters. Our findings expand our understanding of the interplay between promoter elements and initiation kinetics by determining and examining stepwise rate constants at several temperatures and NTP conditions. We determine thermodynamic properties of the RNAP-promoter complex and extrapolate detailed mechanisms for initiation.

In chapter 2, we determine the temperature dependence of stepwise rate constants for RNA synthesis at the λP_R promoter. From this temperature dependence we determine activation enthalpies for each translocation-dependent extension step and infer from those values the nature of bubble collapse during promoter escape. Furthermore, we determine that the most stable OC form for λP_R at 37 °C, RP_O, is not able to productively initiate, and instead must convert to a less stable I_3-like OC conformation.

In chapter 3, we determine the effect of physiological concentrations of pyrophosphate on the initiation kinetic at the λP_R promoter. We determine the stepwise pyrophosphorolysis rate constants for each translocation-dependent step up to 16-mer,

and determine that these rate constants are sequence-dependent. We also discuss the effect of PPi on the fraction of complexes that do not produce a productive RNA.

In chapter 4, we investigate the kinetics of initiation from hybrid promoters containing the *rrnB* P1 discriminator to assess the discriminator effects on initiation. We determine and quantify the dependence of initiation kinetics from these promoters on the second iNTP, and determine that priming initiation at these promoters at the -1 position with dinucleotide CpA can overcome this dependence.

In chapter 5, we investigate the kinetics of initiation from hybrid promoters containing the T7A1 discriminator to assess the discriminator effects on initiation. We also investigate the distribution of discriminator lengths in the *E. coli* genome. We determine stepwise rate constants for RNA synthesis at one promoter construct, and from those values determine that there are two phases of contact breaking in promoter escape from the T7A1 promoter.

In chapter 6, I discuss the significance and limitations of the studies described and offer future directions for determining the effects of promoter sequence and architecture on initiation kinetics and thermodynamics.

Figures

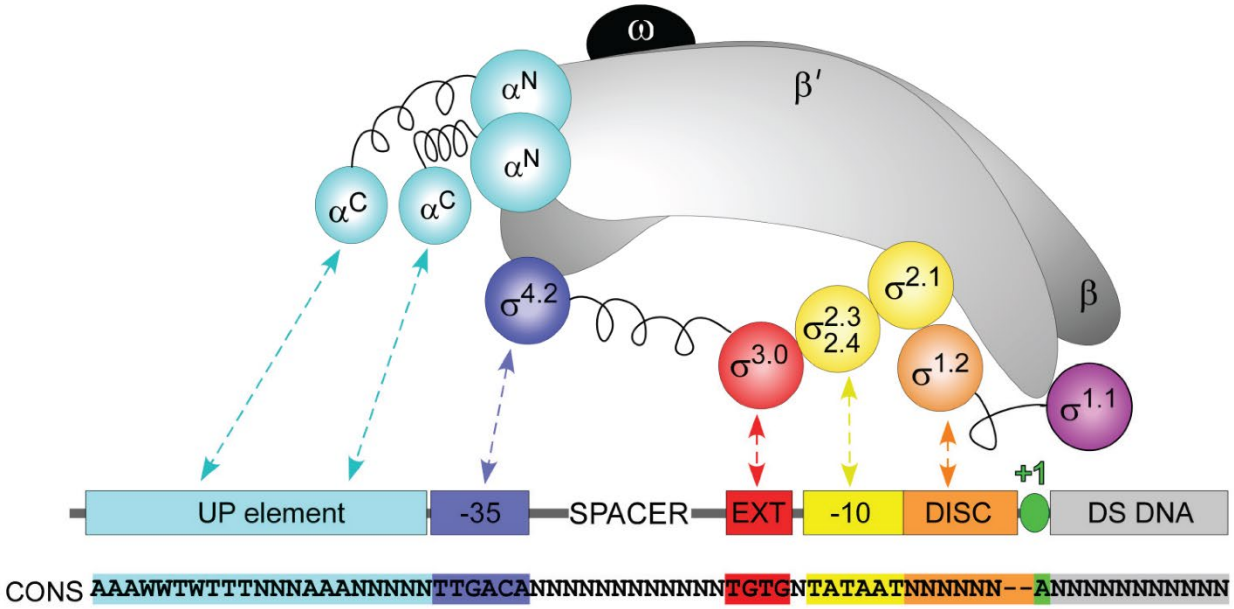


Figure 1: RNAP-Promoter Interactions: Figure adapted from Ruff *et. al.* [9]. Schematic representation of a consensus *E. coli* σ^{70} promoter and corresponding RNAP contacts are shown. RNAP α-C terminal domains (αCTDs) contact promoter UP elements (light blue). $\sigma_{4.2}$ contacts the promoter -35 element (dark blue). $\sigma_{3.0}$ contacts the promoter extended -10 element. $\sigma_{2.4}$ contacts the promoter -10 element (yellow). $\sigma_{1.2}$ contacts the discriminator region of the promoter. The TSS (+1) is indicated in green.

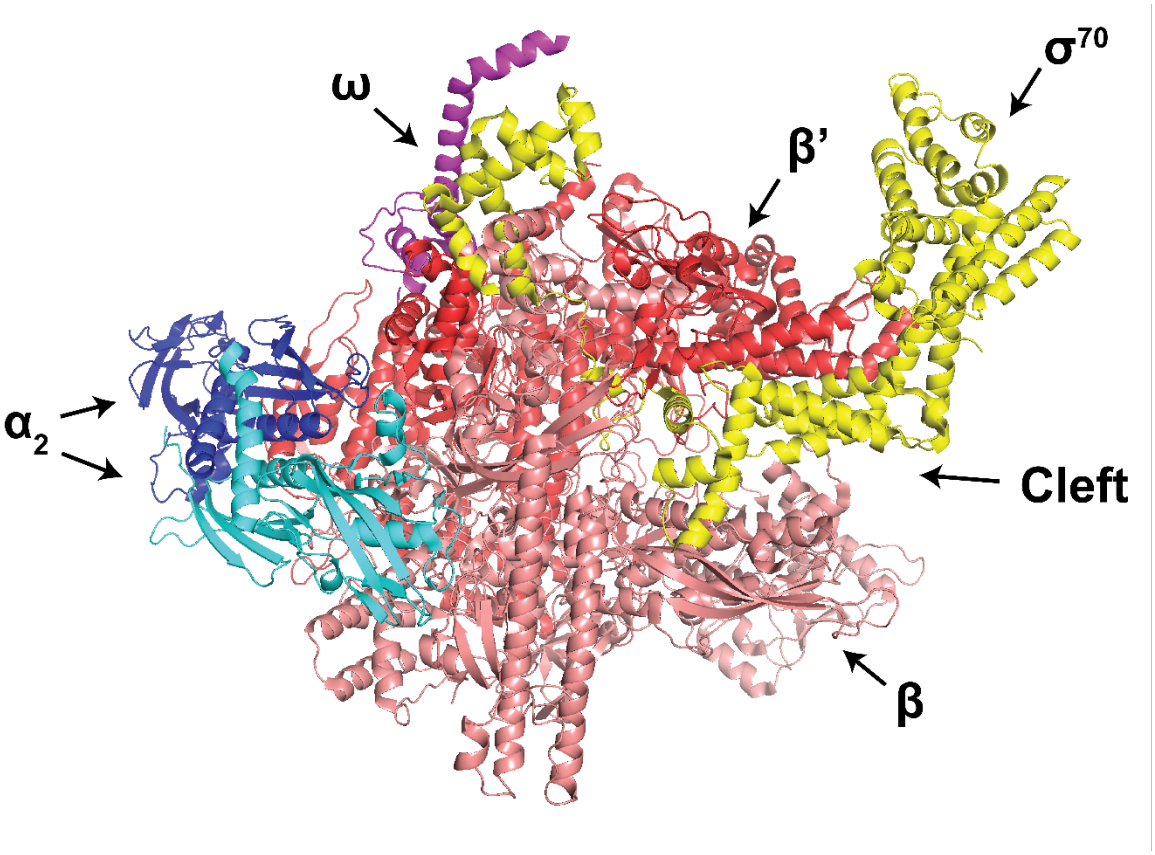


Figure 2: Structure of *E. coli* RNAP Holoenzyme: Crystal Structure of the RNAP holoenzyme from Zuo et. al. [44] (PDB: 4YLN). α subunits are in dark and light blue. ω subunit is purple. β subunit is light red. β' subunit is dark red. These two subunits form the cleft. The active site is positioned in the base of the cleft. Modular promoter recognition domain σ^{70} is yellow.

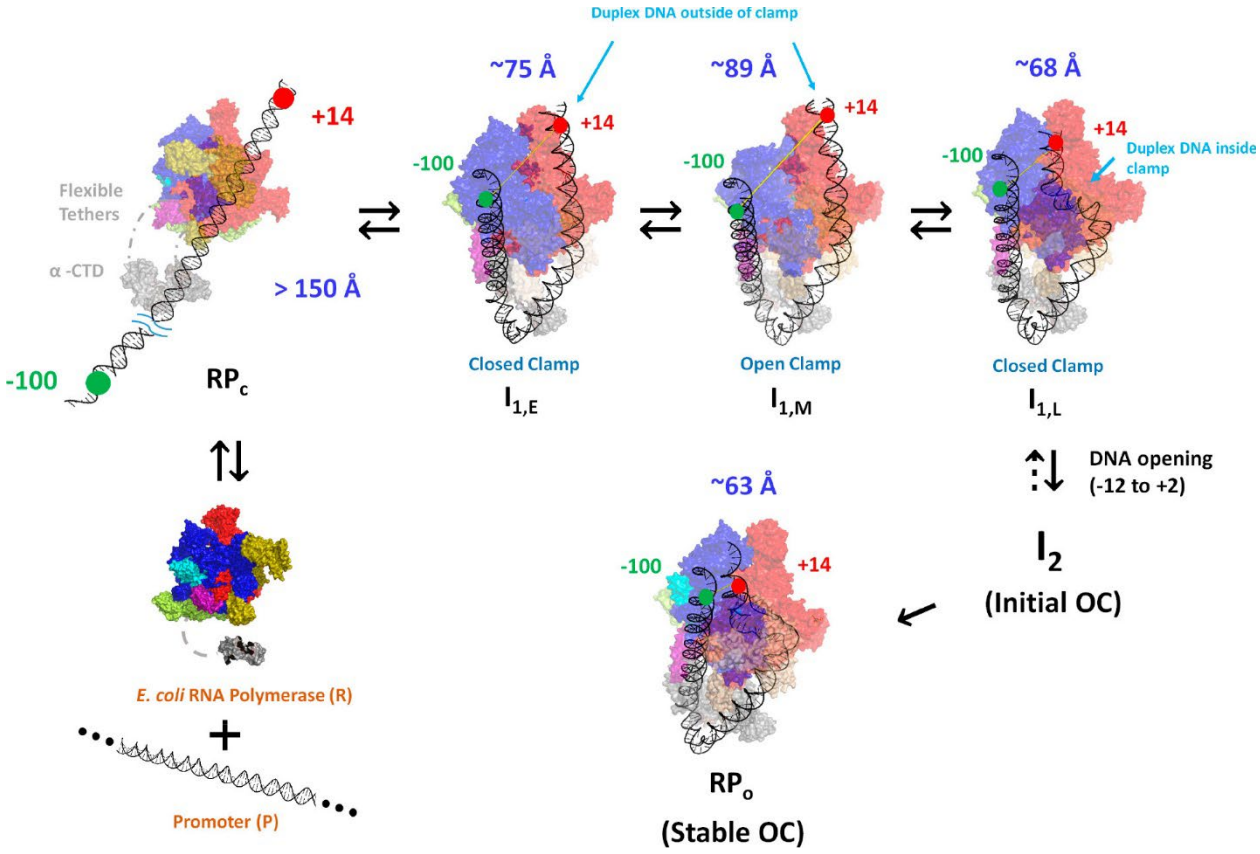


Figure 3: Intermediates on the Pathway to OC Formation at the λP$_R$ Promoter. Figure from Sreenivasan *et. al.* [28]. FRET and PIFE from probes located at -100 and +14 indicate that in RP$_C$ promoter DNA is not yet wrapped around the polymerase. In each subsequent intermediate (I$_{1E}$, I$_{1M}$, I$_{1L}$) DNA is highly bent and wrapped making RNAP contacts at both -100 and +14. Wrapping of DNA onto the β' clamp in I$_{1E}$ is proposed to stimulate clamp opening to form I$_{1M}$. Clamp opening allows entry of downstream DNA into the cleft and subsequent clamp closure to form I$_{1L}$. The initiation bubble is opened after I$_{1L}$ is formed, forming the initial unstable OC I$_2$ which undergoes further stabilization via DME contacts with downstream DNA to form the stable OC.

References

[1] Hurwitz J. The Discovery of RNA Polymerase. Journal of Biological Chemistry. 2005;280:42477-85.

[2] Crick F. Central Dogma of Molecular Biology. Nature. 1970;227:561-3.

[3] Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. Nature Reviews Genetics. 2020;21:630-44.

[4] Weinberg CE, Weinberg Z, Hammann C. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. Nucleic Acids Research. 2019;47:9480-94.

[5] Dutta T, Srivastava S. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. Gene. 2018;656:60-72.

[6] Druzhinin SY, Tran NT, Skalenko KS, Goldman SR, Knoblauch JG, Dove SL, et al. A Conserved Pattern of Primer-Dependent Transcription Initiation in Escherichia coli and Vibrio cholerae Revealed by 5′ RNA-seq. 2015;11:e1005348.

[7] Cramer P. Organization and regulation of gene transcription. Nature. 2019;573:45-54.

[8] Chen J, Boyaci H, Campbell EA. Diverse and unified mechanisms of transcription initiation in bacteria. Nature Reviews Microbiology. 2021;19:95-109.

[9] Ruff EF, Record MT, Jr. , Artsimovitch I. Initial events in bacterial transcription initiation. Biomolecules. 2015;5:1035-62.

[10] Belogurov GA, Artsimovitch I. The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. Journal of Molecular Biology. 2019;431:3975-4006.

[11] Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annual Review of Biochemistry. 2016;85:319-47.

[12] Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Molecular Systems Biology. 2011;7:548.

[13] Van Nimwegen E. Scaling laws in the functional content of genomes. Trends in Genetics. 2003;19:479-84.

[14] Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. Nature Reviews Microbiology. 2011;9:85-98.

[15] Cheetham GM, Steitz TA. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Current Opinion in Structural Biology. 2000;10:117-23.

[16] Steitz TA. The structural changes of T7 RNA polymerase from transcription initiation to elongation. Current Opinion in Structural Biology. 2009;19:683-90.

[17] Misiaszek AD, Girbig M, Grötsch H, Baudin F, Murciano B, Lafita A, et al. Cryo-EM structures of human RNA polymerase I. Nature Structural & Molecular Biology. 2021;28:997-1008.

[18] Fianu I, Dienemann C, Aibara S, Schilbach S, Cramer P. Cryo-EM structure of mammalian RNA polymerase II in complex with human RPAP2. Communications Biology. 2021;4.

[19] Ramsay EP, Abascal-Palacios G, Daiß JL, King H, Gouge J, Pilsl M, et al. Structure of human RNA polymerase III. Nature Communications. 2020;11.

[20] Marichaud Z, Chaloin L, Brodolin K. Regions 1.2 and 3.2 of the RNA polymerase s subunit promote DNA melting and attenuate action of the antibiotic lipiarmycin. Journal of Molecular Biology. 2016;428:463-76.

[21] Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, et al. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. Cell. 2001;104:901-12.

[22] Haugen SP, Ross W, Gourse RL. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nature Reviews Microbiology. 2008;6:507-19.

[23] Brázda V, Bartas M, Bowater RP. Evolution of Diverse Strategies for Promoter Regulation. Trends in Genetics. 2021;37:730-44.

[24] Heyduk E, Heyduk T. DNA template sequence control of bacterial RNA polymerase escape from the promoter. Nucleic Acids Res. 2018;46:4469-86.

[25] Heyduk E, Heyduk T. Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. Biochemistry. 2014;53:282-92.

[26] Sweetser D, Nonet M, Young RA. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proceedings of the National Academy of Sciences. 1987;84:1192-6.

[27] Sutherland C, Murakami KS. An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. EcoSal Plus. 2018;8.

[28] Sreenivasan R, Shkel IA, Chhabra M, Drennan A, Heitkamp S, Wang H-C, et al. Fluorescence-Detected Conformational Changes in Duplex DNA in Open Complex Formation byEscherichia coliRNA Polymerase: Upstream Wrapping and Downstream Bending Precede Clamp Opening and Insertion of the Downstream Duplex. Biochemistry. 2020;59:1565-81.

[29] Ohlsen KL, Gralla JD. DNA melting within stable closed complexes at the *Escherichia coli rrnB* P1 promoter. Journal of Biological Chemistry. 1992;267:19813-8.

[30] Craig ML, Tsodikov OV, McQuade KL, Schlax PE, Jr., Capp MW, Saecker RM, et al. DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. Journal of Molecular Biology. 1998;283:741-56.

[31] Davis CA, Capp MW, Record MT, Jr., Saecker RM. The effects of upstream DNA on open complex formation by *Escherichia coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2005;102:285-90.

[32] Feklistov A, Darst SA. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase s subunit. Cell. 2011;147:1257-69.

[33] Paget M. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules. 2015;5:1245-65.

[34] Gourse RL, Ross W, Gaal T. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Molecular Microbiology. 2000;37:687-95.

[35] Hawley DK, McClure WR. Compilation and analysis ofEscherichia colipromoter DNA sequences. Nucleic Acids Research. 1983;11:2237-55.

[36] Pribnow D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proceedings of the National Academy of Sciences. 1975;72:784-8.

[37] Mitchell JE. Identification and analysis of 'extended -10' promoters in Escherichia coli. Nucleic Acids Research. 2003;31:4689-95.

[38] Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA promoter regulation by nonoptimal binding of s region 1.2: an additional recognition element for RNA polymerase. Cell. 2006;125:1069-82.

[39] Vvedenskaya IO, Vahedian-Movahed H, Zhang Y, Taylor DM, Ebright RH, Nickels BE. Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. Proceedings of the National Academy of Sciences. 2016;113:E2899-E905.

[40] Keilty S, Rosenberg M. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. Journal of Biological Chemistry. 1987;262:6389-95.

[41] Shultzaberger RK, Chen Z, Lewis KA, Schneider TD. Anatomy of Escherichia coli σ 70 promoters. Nucleic Acids Research. 2007;35:771-88.

[42] Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural Basis of Transcription Initiation. Science. 2012;338:1076-80.

[43] Roberts JW, Roberts CW. Base-Specific Recognition of the Nontemplate Strand of Promoter DNA by E. coli RNA Polymerase. Cell. 1996;86:495-501.

[44] Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. Molecular Cell. 2015;58:534-40.

[45] Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, et al. Structure of the bacterial RNA polymerase promoter specificity s subunit. Molecular Cell. 2002;9:527-39.

[46] Shimada T, Yamazaki Y, Tanaka K, Ishihama A. The Whole Set of Constitutive Promoters Recognized by RNA Polymerase RpoD Holoenzyme of Escherichia coli. PLoS ONE. 2014;9:e90447.

[47] Aoyama T, Takanami M, Ohtsuka E, Taniyama Y, Marumoto R, Sato H, et al. Essential structure of *E. coli* promoter: effect of spacer length between two consensus sequences on promoter function. Nucleic Acids Research. 1983;11:5855-64.

[48] Mulligan ME, Brosius J, McClure WR. Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. Journal of Biological Chemistry. 1985;260:3529-38.

[49] Stefano JE, Gralla JD. Spacer mutations in the lac ps promoter. Proceedings of the National Academy of Sciences. 1982;79:1069-72.

[50] Winkelman JT, Chandrangsu P, Ross W, Gourse RL. Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. Proceedings of the National Academy of Sciences. 2016;113:1787-95.

[51] Estrem ST, Gaal T, Ross W, Gourse RL. Identification of an UP element consensus sequence for bacterial promoters. Proceedings of the National Academy of Sciences. 1998;95:9761-6.

[52] Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, et al. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science. 1993;262:1407-13.

[53] Rao L, Ross W, Appleman JA, Gaal T, Leirmo S, Schlax PJ, et al. Factor Independent Activation of rrnB P1. Journal of Molecular Biology. 1994;235:1421-35.

[54] Jeong W, Kang C. Start site selection atlacUV5 promoter affected by the sequence context around the initiation sites. Nucleic Acids Research. 1994;22:4667-72.

[55] Liu J, Turnbough CL. Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in Escherichia coli. Journal of Bacteriology. 1994;176:2938-45.

[56] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[57] Libing Y, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. Elife. 2017;6.

[58] Hsu LM, Cobb IM, Ozmore JR, Khoo M, Nahm G, Xia L, et al. Initial Transcribed Sequence Mutations Specifically Affect Promoter Escape Properties. Biochemistry. 2006;45:8841-54.

[59] Vo NV, Hsu LM, Kane CM, Chamberlin MJ. In Vitro Studies of Transcript Initiation by Escherichia coli RNA Polymerase. 3. Influences of Individual DNA Elements within the Promoter Recognition Region on Abortive Initiation and Promoter Escape. Biochemistry. 2003;42:3798-811.

[60] Kammerer W, Deuschle U, Gentz R, Bujard H. Functional dissection of Escherichia coli promoters: information in the transcribed region is involved in late steps of the overall process. The EMBO Journal. 1986;5:2995-3000.

[61] Hsu LM, Vo NV, Kane CM, Chamberlin MJ. In Vitro Studies of Transcript Initiation by Escherichia coli RNA Polymerase. 1. RNA Chain Initiation, Abortive Initiation, and Promoter Escape at Three Bacteriophage Promoters. Biochemistry. 2003;42:3777-86.

[62] Winkelman JT, Pukhrambam C, Vvedenskaya IO, Zhang Y, Taylor DM, Shah P, et al. XACT-Seq Comprehensively Defines the Promoter-Position and Promoter-Sequence Determinants for Initial-Transcription Pausing. Molecular Cell. 2020;79:797-811.e8.

[63] Duchi D, Bauer DLV, Fernandez L, Evans G, Robb N, Hwang LC, et al. RNA polymerase pausing during initial transcription. Molecular Cell. 2016;63:939-50.

[64] Bauer DLV, Duchi D, Kapanidis AN. E. Coli RNA Polymerase Pauses during Initial Transcription. Biophysical Journal. 2016;110:21a.

[65] Li L, Molodtsov V, Lin W, Ebright RH, Zhang Y. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. Proceedings of the National Academy of Sciences. 2020:201920747.

[66] Burgess RR. Separation and Characterization of the Subunits of Ribonucleic Acid Polymerase. Journal of Biological Chemistry. 1969;244:6168-76.

[67] Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. Cell. 1999;98:811-24.

[68] Shin Y, Qayyum MZ, Pupov D, Esyunina D, Kulbachinskiy A, Murakami KS. Structural basis of ribosomal RNA transcription regulation. Nature Communications. 2021;12.

[69] Saecker RM, Chen J, Chiu CE, Malone B, Sotiris J, Ebrahim M, et al. Structural origins of Escherichia coli RNA polymerase open promoter complex stability. Proceedings of the National Academy of Sciences. 2021;118:e2112877118.

[70] Narayanan A, Vago FS, Li K, Qayyum MZ, Yernool D, Jiang W, et al. Cryo-EM structure ofEscherichia coliσ70RNA polymerase and promoter DNA complex revealed a role of σ non-conserved region during the open complex formation. Journal of Biological Chemistry. 2018;293:7367-75.

[71] Liu B, Hong C, Huang RK, Yu Z, Steitz TA. Structural basis of bacterial transcription activation. Science. 2017;358:947-51.

[72] Liu B, Zuo Y, Steitz TA. Structures ofE.coliσS-transcription initiation complexes provide new insights into polymerase mechanism. 2016;113:4051-6.

[73] Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, et al. A structural model of transcription elongation. Science. 2000;289:619-25.

[74] Hansen UM, McClure WR. Role of the s subunit of *Escherichia coli* RNA polymerase in initiation. Journal of Biological Chemistry. 1980;255:9564-70.

[75] Gruber TM, Gross CA. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annual Review of Microbiology. 2003;57:441-66.

[76] McClure WR. Rate-limiting steps in RNA chain initiation. Proceedings of the National Academy of Sciences. 1980;77:5634-8.

[77] Jensen D, Galburt EA. The Context-Dependent Influence of Promoter Sequence Motifs on Transcription Initiation Kinetics and Regulation. Journal of Bacteriology. 2021;203.

[78] Plaskon DM, Henderson KL, Felth LC, Molzahn CM, Evensen C, Dyke S, et al. Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. Proceedings of the National Academy of Sciences. 2021;118:e2021941118.

[79] Imashimizu M, Tokunaga Y, Afek A, Takahashi H, Shimamoto N, Lukatsky DB. Control of Transcription Initiation by Biased Thermal Fluctuations on Repetitive Genomic Sequences. Biomolecules. 2020;10:1299.

[80] Jensen D, Manzano AR, Rammohan J, Stallings CL, Galburt EA. CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. Nucleic Acids Research. 2019;47:6685-98.

[81] Dulin D, Bauer DLV, Malinen AM, Bakermans JJW, Kaller M, Morichaud Z, et al. Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. Nature Communications. 2018;9.

[82] Margeat E, Kapanidis AN, Tinnefeld P, Wang Y, Mukhopadhyay J, Ebright RH, et al. Direct Observation of Abortive Initiation and Promoter Escape within Single Immobilized Transcription Complexes. Biophysical Journal. 2006;90:1419-31.

[83] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006;314:1144-7.

[84] Chen J, Chiu C, Gopalkrishnan S, Chen AY, Olinares PDB, Saecker RM, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. Molecular Cell. 2020;78:275-88.e6.

[85] Murakami KS, S M, Darst SA. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 A Resolution. Science. 2002;296:1280-4.

[86] Guthold M, Zhu X, Rivetti C, Yang G, Thomson NH, Kasas S, et al. Direct Observation of One-Dimensional Diffusion and Transcription by Escherichia coli RNA Polymerase. Biophysical Journal. 1999;77:2284-94.

[87] Harada Y, Funatsu T, Murakami K, Nonoyama Y, Ishihama A, Yanagida T. Single-Molecule Imaging of RNA Polymerase-DNA Interactions in Real Time. Biophysical Journal. 1999;76:709-15.

[88] Kabata H, Kurosawa O, Arai I, Washizu M, Margarson SA, Glass RE, et al. Visualization of single molecules of RNA polymerase sliding along DNA. Science. 1993;262:1561-3.

[89] Ricchetti M, Metzger W, Heumann H. One-dimensional diffusion of Escherichia coli DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. Proceedings of the National Academy of Sciences. 1988;85:4610-4.

[90] Singer P, Wu CW. Promoter search by Escherichia coli RNA polymerase on a circular DNA template. Journal of Biological Chemistry. 1987;262:14178-89.

[91] Friedman LJ, Mumm JP, Gelles J. RNA polymerase approaches its promoter without long-range sliding along DNA. Proceedings of the National Academy of Sciences. 2013;110:9740-5.

[92] Wang F, Redding S, Finkelstein IJ, Gorman J, Reichman DR, Greene EC. The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. Nature Structural & Molecular Biology. 2013;20:174-81.

[93] Ladouceur A-M, Parmar BS, Biedzinski S, Wall J, Tope SG, Cohn D, et al. Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid-liquid phase separation. Proceedings of the National Academy of Sciences. 2020;117:18540-9.

[94] Azaldegui CA, Vecchiarelli AG, Biteen JS. The emergence of phase separation as an organizing principle in bacteria. Biophysical Journal. 2021;120:1123-38.

[95] Feklistov A, Bae B, Hauver J, Lass-Napiorkowska A, Kalesse M, Glaus F, et al. RNA polymerase motions during promoter melting. Science. 2017;356:863-6.

[96] Glyde R, Ye F, Jovanovic M, Kotta-Loizou I, Buck M, Zhang X. Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. Molecular cell. 2018;70:1111-20.e3.

[97] Chakraborty A, Wang D, Ebright YW, Korlann Y, Kortkhonjia E, Kim T, et al. Opening and Closing of the Bacterial RNA Polymerase Clamp. Science. 2012;337:591-5.

[98] Kontur WS, Capp MW, Gries TJ, Saecker RM, Record MTJ. Probing DNA binding, DNA opening and assembly of a downstream clamp/jaw in E. coli RNA polymerase-lP_R promoter complexes using salt and the physiological anion glutamate. Biochemistry. 2010;49.

[99] Saecker RM, Record MTJ, deHaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. Journal of Molecular Biology. 2011;412.

[100] Davis CA, Bingman CA, Landick R, Record MT, Jr., Saecker RM. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2007;104:7833-8.

[101] Sreenivasan R, Heitkamp S, Chhabra M, Saecker RM, Lingeman E, Poulos MA, et al. Fluorescence resonance energy transfer characterization of DNA wrapping in closed and open *Escherichia coli* RNA polyerase-lP_R promoter complexes. Biochemistry. 2016;55:2174-86.

[102] Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record MT, Jr. *E. coli* polymerase determinants of open complex lifetime and structure. Journal of Molecular Biology. 2015;427:2435-50.

[103] Blankschien MD, Potrykus K, Grace E, Choudhary A, Vinella D, Cashel M, et al. TraR, a Homolog of a RNAP Secondary Channel Interactor, Modulates Transcription. PLoS Genetics. 2009;5:e1000345.

[104] Gopalkrishnan S, Ross W, Chen AY, Gourse RL. TraR directly regulates transcription initiation by mimicking the combined effects of the global regulators DksA and ppGpp. Proceedings of the National Academy of Sciences. 2017;114:E5539-E48.

[105] Chen J, Gopalkrishnan S, Chiu C, Chen AY, Campbell EA, Gourse RL, et al. E. coli TraR allosterically regulates transcription initiation by altering RNA polymerase conformation. eLife. 2019;8.

[106] Guo Y. Promoter opening by sigma 54 and sigma 70 RNA polymerases: sigma factor-directed alterations in the mechanism and tightness of control. Genes & Development. 2000;14:2242-55.

[107] Waldburger C, Gardella T, Wong R, Susskind MM. Changes in conserved region 2 of Escherichia coli σ70 affecting promoter recognition. Journal of Molecular Biology. 1990;215:267-76.

[108] Gries TJ, Kontur WS, Capp MW, Saecker RM, Record MTJ. One-step DNA melting in the RNA polyerase cleft opens the initiation bubble to form an unstable open complex. Proceedings of the National Academy of Sciences. 2010;107:10418-23.

[109] Heyduk E, Kuznedelov K, Severinov K, Heyduk T. A Consensus Adenine at Position –11 of the Nontemplate Strand of Bacterial Promoter Is Important for Nucleation of Promoter Melting. Journal of Biological Chemistry. 2006;281:12362-9.

[110] Kontur WS, Saecker RM, Davis CA, Capp MW, Record MT. Solute Probes of Conformational Changes in Open Complex (RPo) Formation byEscherichia coliRNA Polymerase at the λPRPromoter: Evidence for Unmasking of the Active Site in the

Isomerization Step and for Large-Scale Coupled Folding in the Subsequent Convers. Biochemistry. 2006;45:2161-77.

[111] Saecker RM, Tsodikov OV, McQuade KL, Schlax PE, Jr., Capp MW, Record MTJ. Kinetic studies and structural models of the association of *E. coli* sigma 70 RNA polymerase with the LPR promoter: large scale conformational changes in forming the kinetically significant intermediates. Journal of Molecular Biology. 2002;319:649-71.

[112] Roe J-H, Burgess RR, Record Jr. MT. Temperature dependence of the rate constants of the *Escherichia coli* RNA polymerase-Lambda PR promoter interaction. Assignment of the kinetic steps corresponding to protein conformational change and DNA opening. Journal of Molecular Biology. 1985;184:441-53.

[113] Ruff EF, Kontur WS, Record MTJ. Using solutes and kinetics to probe large conformational changes in the steps of transcription initiation. In: Artsimovitch I, Stanangelo TJ, editors. Bacterial transcriptional control: methods and protocols. New York, NY: Springer; 2015. p. 241-61.

[114] Drennan A, Kraemer M, Capp MW, Gries T, Ruff EF, Sheppard C, et al. Key roles of the downstream mobile jaw of *Escherichia coli* RNA polymerase in transcription initiation. Biochemistry. 2012;51:9447-59.

[115] Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, et al. Interactions between RNA polymerase and the "core recognition element" counteract pausing. 2014;344:1285-9.

[116] Liermo S, Gourse RL. Factor-independent activation of *Esherichia coli* rRNA transcription I. Kinetic Analysis of the roles of the upstream activator region and supercoiling on transcription of the *rrnB* P1 promoter *in vitro*. Journal of Molecular Biology. 1991;220:555-68.

[117] Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, et al. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science. 2016;351:1090-3.

[118] Ko J, Heyduk T. Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochemistry Journal. 2014;463:135-44.

[119] Lewis DEA, Adhya S. Axiom of determining transcription start points by RNA polymerase in Escherichia coli. Molecular Microbiology. 2004;54:692-701.

[120] Henderson KL, Evensen CE, Molzahn CM, Felth LC, Dyke S, Liao G, et al. RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. Biochemistry. 2019;58:2339-52.

[121] Basu RS, Warner BA, Molodtsov V, Pupov D, Esyunina D, Fernandez-Tornero C, et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. Journal of Biological Chemistry. 2014;289:24549-59.

[122] Hein PP, Palangat M, Landick R. RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. Biochemistry. 2011;50:7002-14.

[123] Saba J, Chua XY, Mishanina TV, Nayak D, Windgassen TA, Mooney RA, et al. The elemental mechanism of transcriptional pausing. eLife. 2019;8.

[124] Turtola M, Belogurov GA. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. eLife. 2016;5.

[125] Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. 2007;448:157-62.

[126] Mishanina TV, Palo MZ, Nayak D, Mooney RA, Landick R. Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. Proceedings of the National Academy of Sciences. 2017:201702383.

[127] Windgassen TA, Mooney RA, Nayak D, Palangat M, Zhang J, Landick R. Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. Nucleic Acids Research. 2014;42:12707-21.

[128] Zhang J, Palangat M, Landick R. Role of the RNA polymerase trigger loop in catalysis and pausing. Nature Structural & Molecular Biology. 2010;17:99-104.

[129] Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS, Belogurov GA. Active site opening and closure control translocation of multisubunit RNA polymerase. 2012;40:7442-51.

[130] Yager TD, Von Hippel PH. A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. Biochemistry. 1991;30:1097-118.

[131] Bai L, Fulbright RM, Wang MD. Mechanochemical Kinetics of Transcription Elongation. Physical Review Letters. 2007;98.

[132] Bai L, Shundrovsky A, Wang MD. Sequence-dependent kinetic model for transcirption elongation by RNA polymerase. Journal of Molecular Biology. 2004;344:335-49.

[133] Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science. 2006;314:1139-43.

[134] Winkelman JT, Winkelman BT, Boyce J, Chen AY, Ross W, Gourse RL. Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. Molecular Cell. 2015;59:768-80.

[135] Bandwar RP, Tang G-Q, Patel SS. Sequential Release of Promoter Contacts during Transcription Initiation to Elongation Transition. 2006;360:466-83.

[136] Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH. Translocation of σ70 with RNA Polymerase during Transcription. Cell. 2001;106:453-63.

[137] Perdue SA, Roberts JW. σ70-dependent Transcription Pausing in Escherichia coli. Journal of Molecular Biology. 2011;412:782-92.

[138] Harden TT, Wells CD, Friedman LJ, Landick R, Hochschild A, Kondev J, et al. Bacterial RNA polymerase can retain σ⁷⁰ throughout transcription. Proceedings of the National Academy of Sciences. 2016;113:602-7.

[139] Chander M, Lee A, Vallery TK, Thandar M, Jiang Y, Hsu LM. Mechanisms of very long abortive transcript release during promoter escape. Biochemistry. 2015;54:7393-408.

[140] Vahia AV, Martin CT. Direct tests of the energetic basis of abortive cycling in transcription. Biochemistry. 2011;50:7015-22.

[141] Imashimizu M, Tanaka K, Shimamoto N. Comparative study of Cyanobacterial and *E. coli* RNA polymerases: misincorporation, abortive transcription and dependence on divalent cations. Genetics Research International. 2011;2011:1-11.

[142] Goldman SR, Ebright RH, Nickels BE. Direct Detection of Abortive RNA Transcripts in Vivo. Science. 2009;324:927-8.

[143] Sen R, Nagai H, Hernandez VJ, Shimamoto N. Reduction in abortive transcription from the LPR promoter by mutations in region 3 of the s70 subunit of *Escherichia coli* RNA polymerase. Journal of Biological Chemistry. 1998;273:9872-7.

[144] Munson LM, Reznikoff WS. Abortive initiation and long ribonucleic acid synthesis. Biochemistry. 1981;20:2081-5.

[145] Gralla JD, Carpousis AJ, Stefano JE. Productive and abortive initiation of transcription in vitro at the *lac* UV5 promoter. Biochemistry. 1980;19:5864-9.

[146] Johnston DE, McClure WR. Abortive initiation of in vitro RNA synthesis on bacteriophage l DNA. In: Losick R, Chamberlin MJ, editors. RNA Polymerase. Cold Spring Harbor Laboratory: Cold Spring Harbor; 1976. p. 413-28.

[147] Kubori T, Shimamoto N. A branched pathway in the early stages of transcription by *Escherichia coli* RNA polymerase. Journal of Molecular Biology. 1996;256:449-57.

Chapter 2: Temperature Effects on RNA Polymerase Initiation Kinetics Reveal Which Open Complex Initiates and the Bubble Collapse is Stepwise

This chapter was published as:

Plaskon DM*, Henderson KL*, Felth LC, Molzahn CM, Evensen C, Dyke S, Shkel I, Record MT Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. Proceedings of the National Academy of Sciences. 2021;118:e2021941118.

Experiments in this chapter were performed by Kate Henderson, Lindsey Felth, Cristen Molzahn, Claire Evensen, and Sarah Dyke. Data analysis and interpretation was performed by Dylan Plaskon, Kate Henderson, Irina Shkel, and Tom Record. The paper was written by Tom Record and Dylan Plaskon.

Abstract

Transcription initiation is highly regulated by promoter sequence, transcription factors and ligands. All known transcription inhibitors, an important class of antibiotics, act in initiation. To understand regulation and inhibition, the biophysical mechanisms of formation and stabilization of the "open" promoter complex (OC), of synthesis of a short RNA-DNA hybrid upon NMP addition, and of escape of RNA polymerase (RNAP) from the promoter must be understood. We previously found that RNAP forms three different OC with λP_R promoter DNA. The 37 °C RNAP-λP_R OC (RP_O) is very stable. At lower temperature RP_O is less stable and in equilibrium with an intermediate OC (I_3). Here we report step-by-step rapid quench-flow kinetic data for initiation and growth of the RNA-DNA hybrid at 25 °C and 37 °C that yield rate constants for each step of productive nucleotide addition. Analyzed together with previously-published data at 19 °C, our results reveal that I_3 and not RP_O is the productive initiation complex at all temperatures. From the strong variations of rate constants and activation energies and entropies for individual steps of hybrid extension, we deduce that contacts of RNAP with the bubble strands are disrupted stepwise as the hybrid grows and translocates. Stepwise disruption of RNAP-strand contacts is accompanied by stepwise bubble collapse, base stacking, and duplex formation as the hybrid extends to a 9-mer, prior to disruption of upstream DNA-RNAP contacts and escape of RNAP from the promoter.

Significance

To transcribe DNA sequences into RNA, RNA polymerase (RNAP) first binds to promoter DNA. Promoter sequence dictates both binding strength and rate of remodeling the DNA duplex by RNAP to open 13 base pairs in the start site region. This open "bubble" allows complementary nucleotides to pair with template-strand bases and be incorporated into

RNA. All sequence-specific RNAP-promoter contacts must break for RNAP to escape and elongate the RNA, but how this occurs is not well understood. Here we report rate constants for each initiation step at the λPR promoter at different temperatures. We analyze these data to obtain new insights into which open complex initiates and when RNAP-promoter contacts are disrupted, allowing bubble collapse, duplex formation, and promoter escape.

Introduction

Transcription of DNA information into RNA is fundamental to all life. In prokaryotes, all transcription is performed by a single multi-subunit RNA polymerase (RNAP). Gene expression is highly regulated, and much of this regulation occurs in the steps of transcription initiation. Promoter sequence, transcription factors, ligands, and conditions are key regulatory variables [1-3]. All known transcription inhibitors, an important class of antibiotics, act in initiation [4]. Fig. 1A shows the sequence and the key regions of λP_R promoter DNA. These are the upstream (UP) element, -35 and -10 hexamers, 6 bp discriminator region separating the -10 region from the transcription start site (TSS, +1), and the initial transcribed region (ITR).

Stages of productive initiation are summarized in Fig. 1B. Specific binding of RNAP to duplex (closed) promoter DNA forms an ensemble of closed complexes (CC), including an initial closed intermediate (RP_C) and a series of more advanced closed intermediates (collectively called I_1) in which the promoter DNA is remodeled. “Isomerization” of the ensemble of CC intermediates, including opening of the DNA from the -10 region to the TSS, forms a series of open complexes (OC), including an initial unstable open intermediate (I_2) which at λP_R promoter converts to more stable species (I_3, RP_O), all with

the same open region but with different interactions involving the discriminator strands and the ITR (for representations of these OC, see Fig. 1C). Nucleotide triphosphates (NTPs) complementary to the template DNA sequence are bound and the corresponding monophosphates (NMP) incorporated into a RNA-DNA hybrid in a series of initial transcription complexes (ITC). The hybrid translocates into the cleft with each step of RNA extension, stressing and disrupting RNAP-promoter contacts so that RNAP escapes from the promoter in the transition from initiation to elongation.

Initiation in all likelihood is regulated at all of these stages. To understand regulation, the detailed mechanisms of these stages must be understood. Recent structural [5-9], kinetic-mechanistic [10-19], and high-throughput sequencing studies [20, 21] have greatly advanced our understanding of initiation but much remains to be learned for the rational discovery of new drug targets and for the design of synthetic promoters optimized for specific applications in molecular biology and medical biotechnology [22].

Many examples exist of regulation of initiation at the CC level (i.e. rates and extents of CC formation and isomerization) by promoter sequence, factors, ligand and conditions [1-3]. Regulation of initiation at the OC level is potentially equally significant but not at all well understood. Correlations of OC stability or lifetime with transcription start site selection [23], promoter output [24], and RNA-DNA hybrid length for RNAP escape have been proposed [15]. The three λP_R OC in Fig. 1C differ in lifetime and stability by four orders of magnitude (from ~1 s (I_2) to ~13 hrs (RP_O)) at 37 ºC) [11, 15, 25]. The question of which OC is/are capable of productive initiation in the presence of NTPs has not been addressed previously for any promoter. As shown schematically in Fig. 1C, the λP_R OC differ in the strength of in-cleft interactions with the discriminator

strands and of interactions of the downstream duplex (ITR) with downstream mobile elements (DME) of RNAP [11, 26-28]. The difference in stability between RP_O and I_3 is known to decrease at temperatures below 37 °C, but otherwise I_3 is much less characterized than RP_O and I_2. Formation equilibrium constants and lifetimes of OC at the well-studied T7A1 and *rrnB* P1 promoters are approximately 1 and 3 orders of magnitude less than for λP_R. These OC are proposed to be similar in their downstream interactions to λP_R I_3 and I_2 respectively [11].

For the transition from initiation to elongation to occur, specific RNAP-promoter contacts [29] must be disrupted by translocation of the RNA-DNA hybrid into the upstream cleft [6, 16, 30-34]. Disruption of these RNAP-promoter contacts results in collapse of (and duplex formation by) the upstream portion of the initiation bubble (-1 to -11 for λP_R), escape of RNAP from the promoter, and dissociation of the σ^{70} subunit [12, 35]. Escape of RNAP from the λP_R promoter occurs after synthesis of an initial 10-mer RNA [15, 35].

To investigate these processes in escape of RNAP, we previously determined the step-by-step kinetics and mechanism of NMP incorporation into the growing RNA-DNA hybrid at the λP_R promoter at 19 °C up to the escape point [16]. This information, not available for any other promoter DNA or multi-subunit RNAP, parallels that obtained for the single-subunit T7 bacteriophage RNAP [36]. From a novel analysis of the repeated pattern of small and larger rate constants found for successive steps of RNA extension in initial transcription, we proposed that disruption of RNAP-promoter contacts occur in a stepwise manner prior to promoter escape [16].

Here we report the 25 °C and 37 °C kinetics of initial transcription at the λP_R promoter. Rapid-quench mixing is used to determine overall rates of full-length RNA

synthesis and rate constants for the individual steps of NMP incorporation into the RNA-DNA hybrid at two NTP conditions for comparison with 19 °C results. A change in the initial step of the mechanism at 37 °C as compared to 19 °C reveals that the stable 37 °C OC (RP_O) cannot initiate, and that an intermediate in RP_O formation (I_3) is the initiation complex. From analysis of the temperature-dependence of these rate constants, we deduce when specific contacts of RNAP with single-stranded and duplex regions of promoter DNA are disrupted as the hybrid translocates into the cleft, resulting in collapse of the initiation bubble, duplex formation, and escape of RNAP from the promoter.

Results

Unexpected Effects of Temperature on the Rate of Transcription Initiation at the λP_R Promoter

Time courses spanning more than two orders of magnitude ($\leq$ 0.5 s to $\geq$ 90 s) of transcription initiation by *E. coli* RNA polymerase (RNAP) at the λP_R promoter at 37 °C and 25 °C were obtained by rapid mixing at two different sets of NTP concentrations (designated “low UTP” and “high UTP”) for comparison with previous results at 19 °C [16]. The initial transcribed sequence (Fig. 1A), specifying an RNA that starts with pppApU, is modified from that of λP_R to eliminate the need for CTP until position +17, causing RNA synthesis by productive complexes to pause after 16-mer production when CTP is omitted. The competitor heparin, added with the NTP mixture, ensures single-round productive initiation by preventing re-initiation by any dissociated RNAP.

Panels A and B of Fig. 2 show representative polyacrylamide gel electrophoresis separations of RNAs present in samples quenched at a series of time points during initiation and the transition to elongation at the “low UTP” condition (final concentrations

10 μM UTP, 200 μM ATP and GTP, no CTP, 17.5 nM α-^{32}P-UTP) at 25 °C and 37 °C. SI Appendix, Fig. S1 shows representative gels from initiation kinetics experiments at “high UTP” (final concentrations 200 μM UTP and ATP, 10 μM GTP, no CTP, labelling with 17.5 nM α-^{32}P-GTP). Amounts of each RNA length at each time are quantified by ^{32}P-UTP or ^{32}P-GTP incorporation. Efficient incorporation of α-^{32}P label into the transcript is achieved by use of a low concentration (10 μM) of the corresponding unlabeled NTP.

Because RNAP escapes from this λP_R promoter in the conversion of 10-mer to 11-mer, all RNAs greater than 10-mer in length are defined as full-length (FL) RNA [15]. The transient accumulation of 12-mer and 13-mer may result from the reduction in rate constants for the subsequent steps caused by coupling of translocation to disruption of σ^{70} – core RNAP contacts. Read-through by misincorporation results in extension of the 16-mer RNA to the position of the next C in the transcript at +32, near the fragment end. Transcription occurs slowly near a fragment end, so this second pause is effectively a stop point.

Full-Length (FL) RNA Synthesis

The kinetics of synthesis of FL RNA, determined by summing phosphorimager data for RNA species longer than 10-mer, are also shown in Fig. 2. Panels C and D compare time courses (log scale) of FL RNA synthesis for low UTP and high UTP conditions, respectively, at 25 °C and 37 °C with previous results at 19 °C. Results are averages of 2-4 independent experiments at each condition including those in Fig. 2A, B and SI Appendix, Fig. S1. In all cases, these kinetics are well described as a short lag phase followed by an first order (single exponential) approach to a plateau value with rate constant k_{FL}, as observed previously at 19 °C [16]. Values of the lag time and k_{FL} at high

UTP and low UTP conditions at each temperature are given in SI Appendix, Table S3. At the plateau, 0.5 ± 0.15 FL RNA are synthesized per OC, demonstrating that ~50% of the λP_R OC population at all three temperatures are productive and capable of promoter escape, as previously observed at 19 °C and 37 °C [15, 16].

Arrhenius plots of ln k_{FL} vs 1/T are shown in SI Appendix, Fig. S2. At the high UTP-low GTP condition, ln k_{FL} decreases monotonically but not linearly with 1/T, indicating a positive Arrhenius activation energy that decreases with increasing temperature. At the low UTP-high GTP condition, all k_{FL} values are much smaller, despite a similar number of U and G bases incorporated into the FL transcript (3 U, 4 G by 11-mer formation, Fig. 1A). In addition, k_{FL} is larger at 25 °C than at 37 °C, corresponding to a negative activation energy in this temperature range (SI Appendix, Fig. S2) and indicating a change in the mechanism with increasing temperature. These observations can all be explained by an additional step early in the mechanism of FL RNA synthesis that is very significant at 37 °C but not at 19 °C and is favored at high UTP concentration (see Analysis and Discussion).

Transient Short RNA Intermediates in FL RNA Synthesis

Panels E and F of Fig. 2 show the time evolution of amounts of four different short RNAs (3-mer, 5-mer, 6-mer, 10-mer) formed transiently in FL RNA synthesis by productive complexes at low UTP and high UTP conditions at 37 °C and 25 °C. Previously published results at 19 °C are shown for comparison. Results plotted are averages obtained from analysis of multiple gels such as those in Fig. 2A and 2B, and are normalized per open complex.

At 25 °C, at both low UTP and high UTP, Fig. 2E and 2F show that amounts of all four RNA species increase rapidly and then decrease in the first 10 s after NTP addition, consistent with previous observations at 19 °C and indicating significant transient intermediates on the pathway to productive synthesis [16]. Each transient occurs earlier at 25 °C than at 19 °C at both NTP conditions, as expected since most reaction rates increase with increasing temperature. Each transient occurs earlier at high UTP than at low UTP at both temperatures, expected because UTP is a reactant in the first step of initial transcription (pppApU synthesis).

At 37 °C and low UTP, however, no significant transient population of any intermediate is observed in FL RNA synthesis. At 37 °C and high UTP, transient populations of some longer RNA intermediates (5-mer, 6-mer, 9-mer, 10-mer) are detected, but no transients for shorter RNAs (3-mer, 4-mer) are observed and transient peak amounts of longer RNAs are small by comparison to lower temperatures. These observations and the slower synthesis of FL RNA at low UTP at 37 °C than at 25 °C (Fig. 2C) indicate that the population of λP_R OC at 37 °C, unlike at lower temperatures, is unable to productively bind the initiating NTP and synthesize pppApU without undergoing a conformational change which is sufficiently unfavorable or slow to serve as a bottleneck and desynchronize subsequent steps of short RNA synthesis so fewer transients are observed.

The stability of the λP_R OC is more than 30-fold greater at 37 ºC than at 19 ºC [27, 37]. We previously proposed that the population distribution of λP_R OC also changed with temperature, shifting from the very-stable 37 ºC RP_O, with strong downstream interactions between RNAP DME and duplex DNA extending to +20, to a mixed population of RP_O

and the I_3 intermediate OC (Fig. 1C) at lower temperature [27]. If only I_3 and not RP_O can productively bind the two initiating NTP, then the differences in rates of FL RNA synthesis between 37 ^oC and lower temperatures are readily explained. In Analysis and Discussion this proposal is incorporated into the initiation mechanism and used to analyze the kinetics of transient and FL RNA synthesis by productive complexes.

Discussion

Evidence for an OC Conformational Change Prior to NTP Binding at 37 ^oC but not at 19 ^oC

The mechanism previously used to interpret the kinetics of productive initiation at the λP_R promoter at 19 ^oC, including the appearance and disappearance of short RNA intermediates and the synthesis of FL RNA [16], is shown as Mechanism 1 in Fig. 3. This mechanism begins with ordered, reversible binding of the substrates (ATP (+1), UTP (+2)) to the binary RNAP-promoter OC, followed by irreversible catalysis to synthesize the dinucleotide pppApU. No evidence was obtained for a conformational change in the OC prior to NTP binding at 19 °C [16].

Each subsequent step of initiation begins with reversible translocation. We previously deduced that most translocation steps in initiation are very unfavorable thermodynamically and rapidly reversible [16]. Each translocation step in initiation requires disruption of one downstream DNA-DNA base pair. In addition, translocation steps cause steric [6] and scrunching [31] stress as the RNA-DNA hybrid moves into the cleft, leading to disruption of RNAP-promoter contacts in many of these steps. Because translocation equilibrium constants for initiation steps are small, each step of RNA-DNA hybrid extension up to the predicted point of escape is accurately quantified using a

composite second order rate constant k_i (the analog of an enzymatic k_{cat}/K_m; see [16]), as shown in Mechanism 1 (Fig. 3). Much higher NTP concentrations would be necessary to approach a maximum velocity condition and thereby separate contributions of the overall equilibrium constant for the reversible translocation and NTP binding steps (analog of $1/K_m$) from the rate constant of the irreversible catalytic step (k_{cat}). Fig. 3 shows the good fit of Mechanism 1 to the short RNA transients and FL RNA synthesis kinetics at 19 °C at both low and high UTP conditions.

Also shown in Fig. 3 are fits of productive initiation kinetic data to Mechanism 1 at 25°C and 37°C at low and high UTP conditions. At 25°C, rate constants k_i obtained from these fits accurately reproduce the short RNA transients and FL RNA synthesis kinetics. However, it is clear from Fig. 3 that Mechanism 1 is not sufficient to describe the kinetics of productive initiation at 37 °C. To obtain an accurate fit to the 37 °C kinetic data requires the addition of an unfavorable reversible step at the beginning of the mechanism, prior to initial ATP binding (Fig. 3, Mechanism 2, step 1a). This step represents a conformational change which we propose is the conversion of the very stable 37 °C OC (RP_O) to another OC conformation that we designate the initiating complex (IC). This step appears to be rapidly reversible and is characterized by the equilibrium constant K_{1a} (for $RP_O \rightarrow$ IC). All other steps of Mechanism 2 are the same as Mechanism 1. Fig. 3 also shows that use of Mechanism 2 to fit 19 °C and 25 °C data sets does not affect the quality of these fits and yields estimates of K_{1a} at these temperatures.

These fits predict that K_{1a} is extremely temperature dependent, greatly favoring RP_O at 37 °C ($K_{1a} \approx 0.01$; 99% RP) but favoring IC at 19 °C ($K_{1a} \approx 5.3$; more than 80% IC). A near-equimolar ratio of RP_O and IC is predicted at 25 °C ($K_{1a} \approx 0.70$; ~40% IC;

~60% RP_O). Good fits of 19 °C and 25 °C kinetic data to Mechanism 1 are obtained because a significant fraction of the OC population is initially in the IC conformation at these temperatures. The strong decrease in K_{1a} with increasing temperature indicates that the enthalpy change for the conversion of RP_O to IC is large in magnitude and negative; van't Hoff analysis yields ΔH_{1a}^o = -60 $\pm$ 20 kcal/mol. The standard free energy change ΔG_{1a}^o for this conversion ranges from ~ 2.7 kcal at 37 °C to ~ -1 kcal at 19 °C, and the corresponding entropy change ΔS_{1a}^o is -200 $\pm$ 60 eu. Conversion of RP_O to IC therefore shows near-complete enthalpy-entropy compensation, like many other protein processes.

Evidence that Only the I_3 Intermediate OC and not RP_O or I_2 Initiates Transcription from λP_R Promoter Upon NTP Addition

Evidence exists for two open intermediates (I_2, I_3) on the pathway to formation of the RP_O complex at the λP_R promoter (Fig. 1C). The thermodynamic, kinetic, and footprinting information available for these intermediates support the proposal that IC is I_3. I_2 is unstable with respect to I_3 and/or RP_O at higher temperatures and unstable with respect to closed complexes at lower temperatures [11, 17, 26, 27]. Conversion of I_2 to I_3 involves folding of 100-150 amino acid residues of RNAP DME [27], and is thought to strengthen contacts between the proximal downstream duplex and the β lobe and the β' clamp [28]. Conversion of I_3 to RP_O is thought to involve primarily an interaction of the downstream jaw and associated DME with the distal downstream duplex (+10 to +20), which serves to tighten the entire RNAP-promoter interface in the OC. The OC formed by the jaw deletion variant of RNAP and by downstream truncation variants of the promoter are thought to be models of I_3; equilibrium constants for forming these variant OC are one

to two orders of magnitude smaller than binding of WT RNAP to full-length promoter DNA at 37 ºC. Hydroxyl radical footprinting of the open complex with the jaw deletion RNAP variant reveals that the entire RNAP-promoter interface in the OC is less protected and hence "looser" and more hydrated than in the WT RNAP OC [28].

From this body of previous research, the stable OC population was proposed to be an equilibrium mixture of I_3 and RP_O, with RP_O highly favored at 37 °C and I_3 increasing in significance at lower temperature [10, 28], but the details of this were not known. Here we find that the stable OC population is an equilibrium mixture of RP_O and the IC initiation complex, with RP_O favored at 37 C and IC favored below 25 °C, indicating that IC is I_3. In support of this, extrapolation of K_{1a} (Mechanism 2) to lower temperature assuming a temperature-independent enthalpy change for RP_O → IC ($\Delta H^o_{1a} \approx$ -60 kcal/mol) predicts that the IC:RP_O population distribution for λP_R at 10 ºC, before NTP addition, is ~ 99% IC and only 1% RP_O. At 10 ºC, Gries *et. al*. [38] determined MnO_4^- footprints of both strands of the open region in the stable λP_R OC, now identified as the IC initiation complex. In addition, salt-upshifts were used to rapidly destabilize the 10 ºC λP_R OC and obtain a burst of I_2, the least stable open intermediate, for MnO_4^- footprinting. Hence the 10 ºC λP_R OC population, identified in the current research as 99% IC, is more stable and hence more advanced than I_2 at 10 ºC and therefore must be I_3.

Kinetic-Mechanistic Evidence for Sequential Disruption of RNAP-Strand Contacts and Bubble Collapse in the Steps of Initiation

Fig. 4 (also SI Appendix, Table S4) compares rate constants k_i for each step of incorporation of NMP into the RNA-DNA hybrid at 25 ºC and 37 ºC with 19 ºC values. All are calculated using Mechanism 2 (Fig. 3); the 19 ºC values are not significantly different

from those obtained previously using Mechanism 1 [16]. Steps of initiation that involve translocation show the same pattern of large and small k_i values at all temperatures investigated. Three distinct regions of small k_i identified in Fig. 4. These correspond to **a**) synthesis of 4- and 5-mer, **b**) synthesis of 7-mer, 8-mer and 9-mer, and **c**) synthesis of 11-mer and are separated by single steps with larger k_i for synthesis of 3-mer, 6-mer and 10-mer. In initiation, where translocation is unfavorable because of steric and scrunching stress, k_i is interpretable (to a good approximation) as the product of the equilibrium constants for translocation and NTP binding and the catalytic rate constant k_{cat} [16].

To transition from initiation to elongation, the specific contacts between RNAP and promoter DNA that were essential to form and stabilize the OC must be broken. Previously we interpreted the pattern of 19 °C k_i values in terms of the serial disruption of these promoter contacts. In this interpretation, differences in k_i values arise from differences in the equilibrium constant for translocation at each nucleotide addition step, $K_{tr,i}$. The very similar patterns of small and large k_i values at all three temperatures investigated indicate that at all three temperatures, as shown in Fig. 4, interactions of RNAP with the discriminator strands (indicated as region **a** in Fig. 4) are broken in translocation preceding synthesis of 4- and 5-mer. Strong interactions of RNAP with the strands of the -10 region (**b** in Fig. 4) are broken in translocation preceding synthesis of 7-mer, 8-mer and 9-mer, and interactions of RNAP with the -35 and upstream regions (**c** in Fig. 4) are disrupted in translocation prior to synthesis of an 11-mer.

Increasing temperature from 19 °C to 37 °C affects k_i for each step very differently. With one exception, k_i values increase with increasing temperature, strongly for 3- and 4-mer, modestly for 5-, 6-, 7- and 8-mer, and more strongly for 10-mer and 11-mer. Notably,

rate constant k_9 for 9-mer decreases with increasing temperature. Analysis of these temperature dependences yields Arrhenius activation energies $E_{act,i}$ (Fig. 4; SI Appendix, Table S5) which vary by 30 kcal, from +26 to -4 kcal.

$E_{act,2}$ for incorporation of UTP into pppApU, a step that does not involve translocation stress, is ~ 9 kcal, similar to that reported previously for steps of elongation by *E. coli* RNAP (10 – 13 kcal; [39]). $E_{act,3}$ and $E_{act,4}$ for 3-mer and 4-mer are significantly larger (~ 26 kcal), while E_{act} values for the next 5 steps are quite small (~ 6 kcal for synthesis of 5-mer, 6-mer, 7-mer and 8-mer RNA and ~ -4 kcal for 9-mer). Contacts of RNAP with the strands of the upstream bubble are proposed to be disrupted in many of these steps [16], and an interpretation of k_i and $E_{act,i}$ values in terms of changes in base stacking in these steps is given below. $E_{act,10}$ and $E_{act,11}$ are larger than the previous E_{act} values, though not as large as $E_{act,3}$ and $E_{act,4}$.

From rate constants k_i and activation energies $E_{act,i}$, the quasi-thermodynamic quantities $\Delta G_i^{o\ddagger}$ $\Delta H_i^{o\ddagger}$, and $\Delta S_i^{o\ddagger}$ for conversion of reactants to the catalytic transition state of each step can be determined if the hypothetical maximum second-order rate constant k_{max} for this process (at $\Delta G_i^{o\ddagger} = 0$) and its temperature dependence are known. We assume that all steps have the same k_{max} and approximate it by the value for an orientation-corrected diffusion-limited rate constant ($k_{max} \approx 10^3$ $\mu M^{-1}s^{-1}$), neglecting the small temperature dependence of a diffusion-limited k_{max} in calculating $\Delta H_i^{o\ddagger}$ values [40]. Although the uncertainty in the appropriate value of k_{max} is probably one order of magnitude, this is of no consequence for the analysis in SI Appendix, Tables S5-S6 and below as long as k_{max} has the same value for each step.

For $k_{max} \approx 10^3\ \mu M^{-1}\ s^{-1}$, $\Delta G_2^{o\ddagger}$ for incorporation of the initiating UTP into pppApU is ~ 4.3 kcal, and values of $\Delta G_i^{o\ddagger}$ for subsequent steps of initiation ($3 \leq i \leq 11$) are in the range ~ 5.1 kcal to ~ 6.6 kcal (SI Appendix, Table S5). Values of $\Delta H_i^{o\ddagger}$ and $\Delta S_i^{o\ddagger}$ vary over much wider ranges (~ 30 kcal range in $\Delta H_i^{o\ddagger} = E_{act,i}$, ~ 100 eu range in $\Delta S_i^{o\ddagger}$) (SI Appendix, Table S5). For incorporation of the initiating UTP into pppApU (2-mer synthesis), $\Delta H_2^{o\ddagger}$ and $\Delta S_2^{o\ddagger}$ are modest (9 kcal, 16 eu). These values include contributions from the thermodynamics of UTP binding, including stacking of UTP on ATP, and from the intrinsic activation quantities for the catalytic step, but do not include any contributions from a translocation step.

Values of $\Delta H_i^{o\ddagger}$ and $\Delta S_i^{o\ddagger}$ for subsequent steps involving translocation are very different from those of 2-mer synthesis. Values of $\Delta H_i^{o\ddagger}$ and $\Delta S_i^{o\ddagger}$ for 3-mer and 4-mer are much larger than for 2-mer, while values of $\Delta H_i^{o\ddagger}$ and $\Delta S_i^{o\ddagger}$ for 5-, 6-, 7-, and 8-mer are much smaller (see SI Appendix, Table S5) and values of $\Delta H_9^{o\ddagger}$ and $\Delta S_9^{o\ddagger}$ for 9-mer are modestly negative. These steps involving translocation included those identified previously as ones in which contacts of the discriminator and -10 strands with RNAP are disrupted, freeing the initiation bubble strands. $\Delta H_i^{o\ddagger}$ and $\Delta S_i^{o\ddagger}$ for 10-mer and 11-mer are more comparable to 3-mer and 4-mer. Contacts that are disrupted in these steps are proposed to be with the duplex (-35 and upstream) and not with single-stranded DNA. Hence the unusual activation thermodynamics are confined to the steps that break RNAP contacts with the bubble strands.

Identifying Contributions of Downstream DNA Melting and of Base Stacking and Duplex Formation of the Upstream Bubble to the Kinetics of Initiation Steps

Disruption of base stacking interactions in the melting of double-helical DNA to two separated strands is a major determinant of the melting enthalpy [41]. Bases in single-stranded (ss) nucleic acids are highly stacked in solution at 0 °C and unstack noncooperatively with increasing temperature. The enthalpy change for base-unstacking in ss DNA is ~ 5 kcal mol^{-1}, and the enthalpy of duplex melting is predicted to vary from ~ 5 kcal $(mol\ base\ pair)^{-1}$ under conditions where the bases in the separated strands are fully stacked to ~15 kcal $(mol\ base\ pair)^{-1}$ under conditions where the bases in the separated strands are fully unstacked [41].

Steps synthesizing 3-mer and 4-mer, the first two steps of initiation that begin with translocation and melting of one bp at the downstream end of the initiation bubble, exhibit activation energies and enthalpies that are ~ 17 kcal larger than for the step synthesizing 2-mer, which at the λP_R promoter requires neither translocation nor base-pair melting. Melting of one base pair can account for much of the 17 kcal differences between $\Delta H_3^{o\ddagger}$ or $\Delta H_4^{o\ddagger}$ vs $\Delta H_2^{o\ddagger}$, depending on the extent to which the bases after melting are unstacked.

Since subsequent steps of initiation also involve translocation and melting of one downstream DNA base pair, how can the very small (and in one case negative) $\Delta H_i^{o\ddagger}$ values for the 5 subsequent steps of initiation (5- to 9-mer synthesis) be explained? SI Appendix, Table S5 shows that values of $\Delta H_i^{o\ddagger}$ for 5- to 8-mer synthesis are ~ 20 kcal less positive and $\Delta H_i^{o\ddagger}$ for 9-mer synthesis is ~ 30 kcal less positive than for 3- and 4-mer synthesis.

Previously, to explain the pattern of small and large rate constants of these steps, we proposed that contacts of RNAP with the discriminator and -10 regions of the bubble

strands are disrupted in a step-wise manner, beginning at the downstream end of the template and/or nontemplate discriminator strands, as the hybrid lengthens to a 9-mer. Here, to explain the unusual values of $\Delta H_i^{o\ddagger}$ for 5-mer to 9-mer synthesis, we propose that substantial stacking of bases in these strands accompanies the release of contacts with RNAP. A comparison of permanganate reactivities of thymines in the open region of the stable open complex I_3 and the unstable I_2 revealed that most thymines in I_3 are significantly more reactive and hence more exposed than in I_2 [38]. To explain this, we proposed that bases in the open region of I_3 are much more unstacked than in I_2 because of tighter binding of the strand backbones in the open region of I_3 [38]. Precedent for base unstacking in formation of a protein-ssDNA complex in which the DNA backbone is strongly bound is provided by the very stable SSB-ssDNA complex [42]. Hence, base stacking is expected to accompany disruption of contacts between RNAP and the open strands of the initiation bubble in steps of initiation involving translocation. The 20 - 30 kcal reductions in $\Delta H_i^{o\ddagger}$ for 5- to 9-mer synthesis, compared to 3-mer and 4-mer synthesis, are consistent with stacking of 4 bases in each step of synthesis of 5-mer, 6-mer, 7-mer and 8-mer, and stacking of 6 bases in synthesis of 9-mer (Table S6).

We expected to observe a large negative contribution to $\Delta H_i^{o\ddagger}$ from duplex formation by the bubble strands in one or more late steps of initiation. Formation of an 11 bp duplex is predicted to contribute at least -50 kcal mol^{-1} to $\Delta H_i^{o\ddagger}$ if those strands are already stacked and as much as -150 kcal mol^{-1} if they are initially unstacked. From the analysis in SI Appendix, Table S5, the only step that could include a negative enthalpy term of this magnitude is 9-mer formation, where we identify a -30 kcal contribution that could be from duplex formation from fully stacked strands. The smaller magnitude of this

contribution (-30 vs -50 kcal mol^{-1}) could result from the need to break enthalpically favorable interactions of the -7 and -11 bases on the nontemplate strand with aromatic residues in the base-binding pockets of sigma region 2 [29].

Base pair formation may occur incrementally as the hybrid extends from a 5-mer to a 9-mer. Base pair formation in these steps (including base stacking) would also explain their small activation enthalpies. Base pair formation would explain the finding (above and Table S6) that base stacking increases in increments of 2, 4, or 6 bases (i.e. 1, 2 or 3 base pairs) in these steps. Base pair formation can also explain the offset between the first step with a small rate constant (4-mer synthesis, accompanied by disruption of some RNAP-discriminator strand interactions) and the first step with a small activation energy (5-mer synthesis, accompanied by formation of 4 base-stacking interactions or two base pairs). This offset would not be expected for ss stacking interactions, but is explained if RNAP interactions with the downstream end of one discriminator strand are disrupted in 4-mer synthesis, while interactions with this region of the other strand are disrupted in 5-mer synthesis, allowing base pairing. A structural interpretation of activation enthalpies (SI Appendix, Table S5) in terms of step-by-step base pairing as RNAP-strand contacts are disrupted in initiation is given in SI Appendix, Table S6. Because structural studies of initiation complexes to date have used heteroduplexes with an open initiation bubble, they provide no information about the extent of pairing of the bases that accompanies the synthesis of 5-mer or longer RNA in a productive initiation complex.

Conclusions

We find that the λP_R initiation complex is the intermediate open complex I_3 (Fig. 1) and not the very stable RP_O. Initiation at 37 °C where the λP_R open complex is

predominantly RP_O requires an initial conformational change to form I_3 before productive binding of both initiating NTPs. We also find that the disruption of RNAP contacts with the strands of the initiation bubble and bubble collapse occur stepwise in initiation, prior to disruption of RNAP contacts with the upstream duplex and escape of RNAP from the promoter. At a minimum, step-wise bubble collapse involves step-wise stacking of ss bases, and may also involve step-wise base pairing. The activation thermodynamics provide no evidence that duplex formation by the bubble strands is delayed until the RNAP escape point, when interactions with the -35 region upstream duplex have been disrupted. Instead the activation thermodynamics are consistent with duplex formation in one or more earlier steps in which contacts of the bubble strands with the discriminator and -10 region are disrupted. Analogous quantitative studies of initiation kinetics and mechanism at other promoters and for λP_R with different discriminator lengths and ITR are needed to characterize the relationship between promoter identity and initiation kinetics and mechanism, and the regulatory possibilities therein.

Materials and Methods

Details about reagents (buffers, enzymes, DNA), initiation kinetic assays (single-round in synthesis of full-length RNA), and analysis of amounts of transient short RNAs from productive complexes and of stalled and released short RNA from nonproductive complexes are described in Supplemental, with references to previous publications. Briefly, the λP_R OC is preformed by incubation at the experimental temperature (25 °C or 37 °C) for one hour. Preformed OC and initiation solution containing NTPs and heparin are mixed 1:1 at time zero using a Kintek Corp. Rapid Quench Flow (RQF), and quenched with 8 M urea and 15 mM EDTA at the times indicated. RNA products are visualized and

quantified using PAGE followed by phosphorimager analysis of incorporated α-32P-NTPs, as described previously (15). Fits were obtained using Kintek Explorer.

Acknowledgements
DP was supported by NIH Biotechnology Traineeship NIH 5 T32 GM008349 and KH by NIH NRSA postdoctoral fellowship NIH GM 122303. We gratefully acknowledge the very helpful comments of the editor and reviewers and the support for this research from the above mentioned fellowships, UW-Madison, and NIH GM R35-118100 (MTR).

Figures

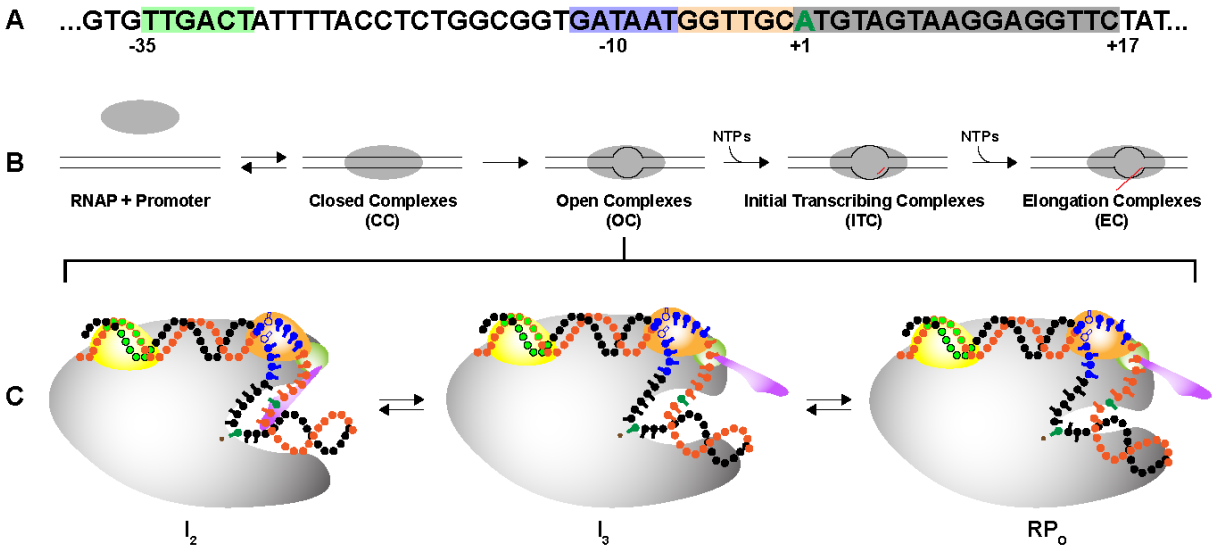


Figure 1: Transcription initiation at λP_R promoter. A) Nontemplate strand sequence of the λP_R promoter studied here, in which the initial transcribed region (ITR) is modified to eliminate incorporation of CTP before position +17, so transcription halts at a 16-mer RNA when CTP is withheld. The -35 (green), -10 (blue), and discriminator (tan) elements, and the start site (+1, green) and ITR (grey) are highlighted. **B)** Summary of stages in productive initiation by RNA polymerase (RNAP) holoenzyme ($\alpha_2\beta\beta'\omega\sigma^{70}$). Key aspects of the mechanisms of two of these stages (OC, ITC) are determined in this study. **C)** Schematic representation of OC species at the λP_R promoter at 37 °C: intermediates I_2 and I_3 and stable RP_O. As illustrated, interactions of in-cleft and downstream elements of RNAP with the discriminator and ITR, weak or absent in I_2, are stronger in I_3 and much stronger in RP_O at 37 °C [10]. RNAP core (grey) is shown with relevant σ^{70} subunits 1.1 (purple), 1.2 (green), 2 (light orange), and 4 (yellow). Linking domains between subunits are not shown for clarity. The template DNA strand is shown in black, and the nontemplate strand in orange, with the -35 (light green), -10 (blue), and start site (dark green) indicated.

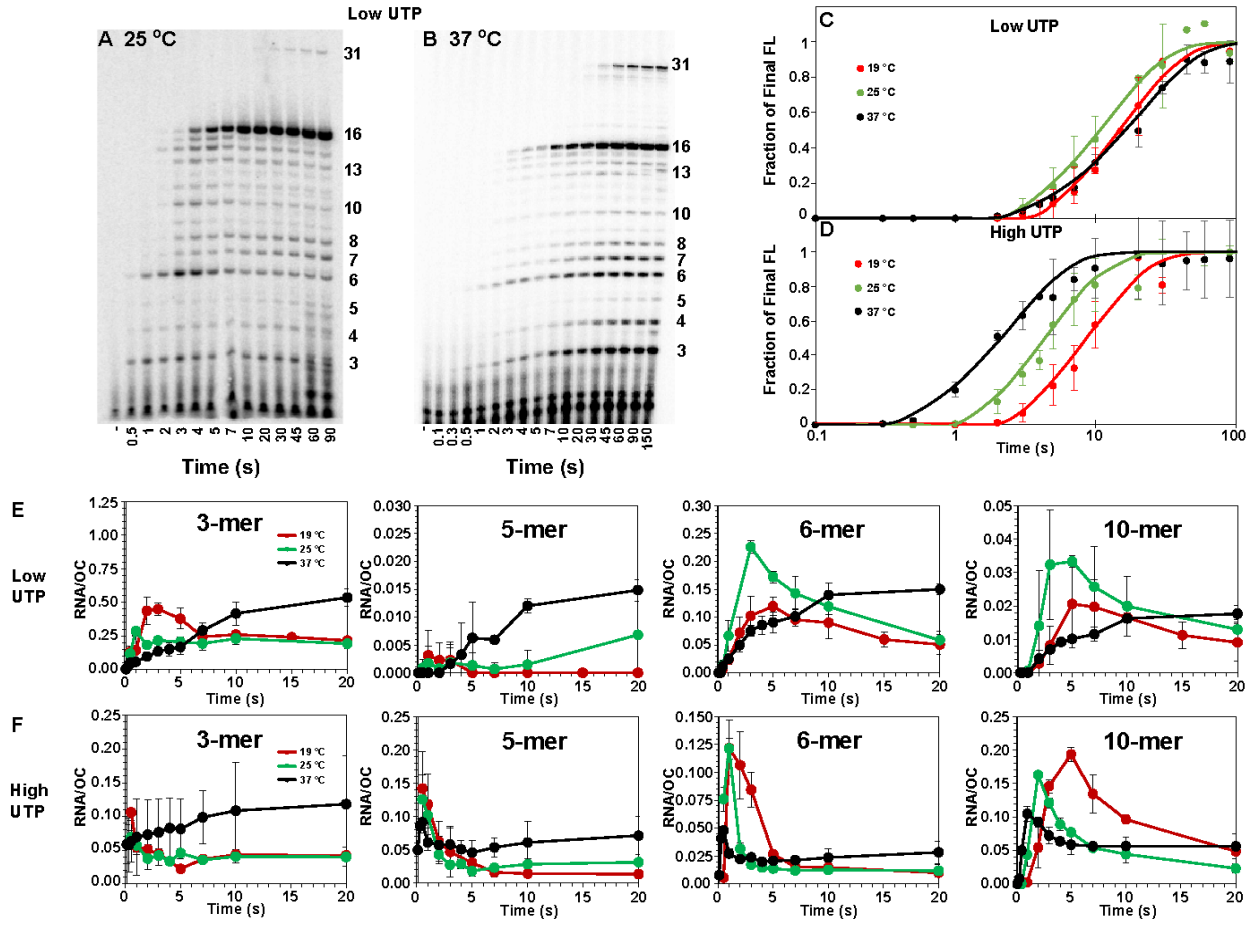


Figure 2: Step-by-Step Kinetics of Transcription Initiation from λP$_R$ Promoter at Different NTP Concentrations and Temperatures. Top Left: Polyacrylamide gel electrophoretic separations of individual ^{32}P-labelled RNA bands (from 3-mer to 16-mer and 31-mer) observed as a function of time during productive and nonproductive initiation at 25 °C (Panel A) and 37 °C (Panel B). Lanes span the time range from 0.5 to 90 s or 0.1 to 150 s after adding NTPs and heparin to OC formed by premixing RNAP and λP$_R$ promoter DNA (See SI Appendix, Methods). Gels shown are for the "low UTP" condition: 200 μM ATP and GTP, 10 μM UTP, 17.5 nM α-^{32}P-UTP. CTP is omitted causing transcription to pause at a 16-mer RNA, before read-through occurs to synthesize longer transcripts [15]. Representative "high UTP" (200 μM ATP and UTP, 10 μM GTP, 17.5 nM

α-^{32}P-GTP) gels at 25 °C and 37 °C are shown in SI Appendix, Fig. S1. **Top right:** Time courses (log scale) of synthesis of “full-length” (> 10-mer) RNA at 19 °C (red; [15]), 25 °C (green), and 37 °C (black) at low UTP (Panel C) and high UTP (Panel D). **Bottom:** Comparisons of linear time courses for representative short RNAs (3-mer, 5-mer, 6-mer, 10-mer) synthesized by productive and non-productive complexes at low UTP (Panel E) and high UTP (Panel F) at 19 °C (red; (15)), 25 °C (green), and 37 °C (black). At each time point, the average amount of each RNA present (per OC) is shown with the estimated uncertainty.

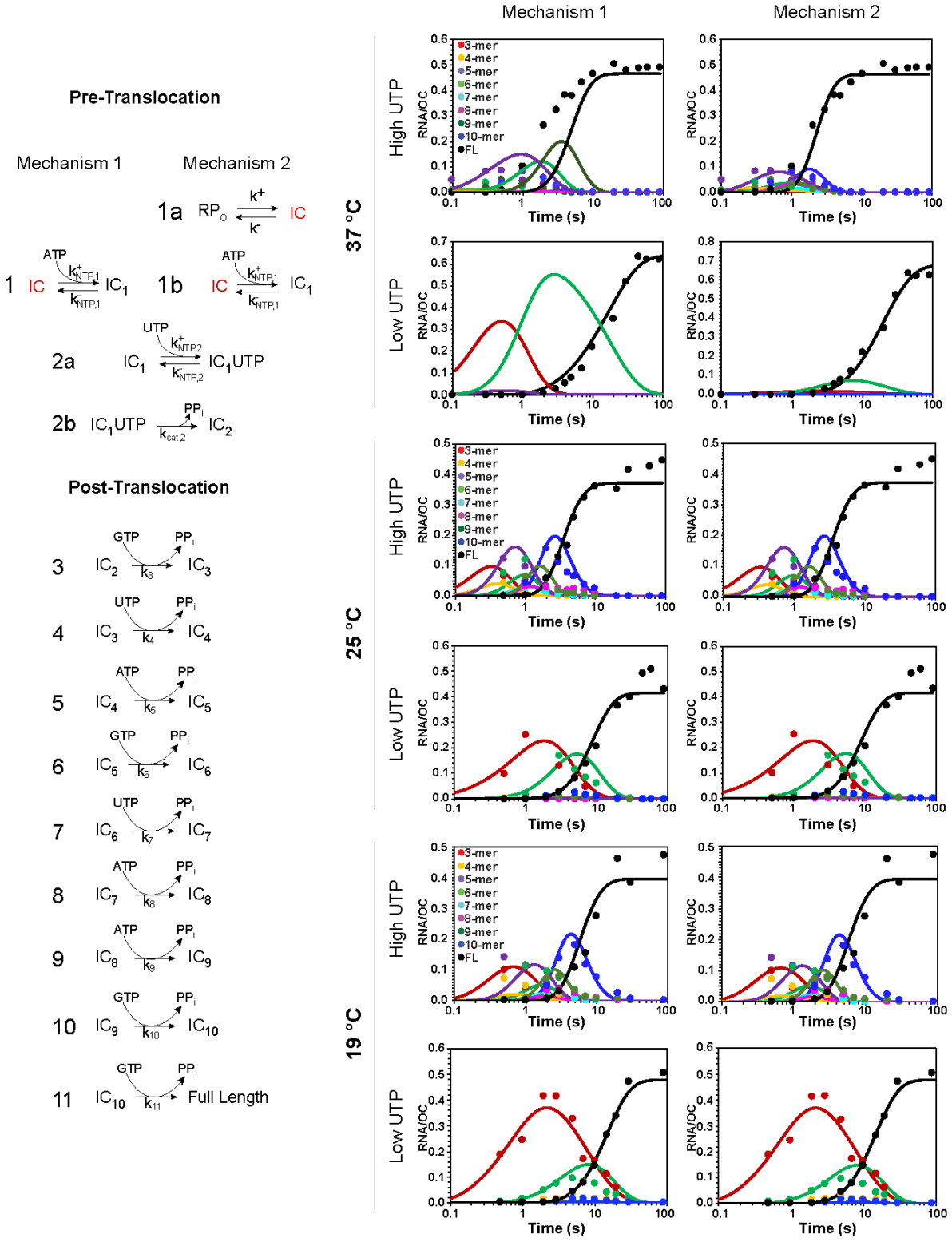


Figure 3: Fits of Initiation Kinetic Data to Step-by-Step Mechanisms. LEFT: Two mechanisms differing in initial (pre-translocation) steps are shown. **Mechanism 1,**

previously shown to describe λP_R initiation data at 19 °C [15], is a minimal initiation mechanism for situations where the stable OC is also the initiation complex (IC). **Mechanism 2**, proposed here to describe λP_R initiation data at 37 °C where the stable OC is RP_O, includes an initial conformational change that converts RP_O to the IC that binds NTP and productively initiates as in Mechanism 1. **Right:** Comparison of fits (solid curves) to Mechanisms 1 and 2 of all λP_R productive initiation data at 37 °C, 25 °C, and 19 °C for the two NTP conditions investigated (log time scale). Colors: 3-mer, red; 4-mer, yellow; 5-mer, purple; 6-mer, light green; 7-mer, light blue; 8-mer, pink; 9-mer, dark green; 10-mer, dark blue; FL, black. Analyses of short RNA time courses (e.g. Fig. 2E, F) to obtain the transients in full-length RNA synthesis plotted here (points) are shown in SI Appendix, Figs. S3 – S6. Rate constants k_i determined from these fits to Mechanism 2 at each temperature are given in SI Appendix, Table S4.

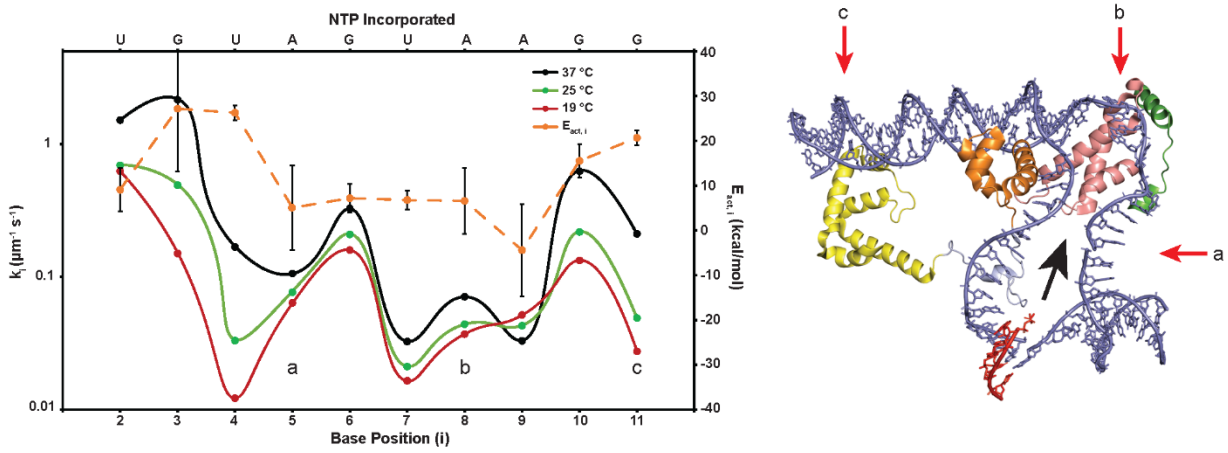


Figure 4. Left: Comparison of Rate Constants and Activation Energies for Individual Steps of Initiation: Composite rate constants $\mathbf{k_i}$ (left axis; μM^{-1} s^{-1}) for each step (i) of NMP incorporation and hybrid extension before escape of RNAP are plotted on a log scale for the λP_R promoter at 19 °C (red, [15]), 25 °C (green), and 37 °C (black). At each temperature differences in $\mathbf{k_i}$ values are interpreted as differences in the equilibrium constant $\mathbf{K_{tr,I}}$ for translocation that occurs before NTP binding and catalysis in steps 3-11. Three regions of small $\mathbf{k_i}$ (and therefore small $\mathbf{K_{tr,\,i}}$) values, observed at all temperatures, are labelled **a, b,** and **c** for comparison with the framework ITC structure (right). Arrhenius activation energies ($\mathbf{E_{act,\,i}}$, orange, right axis) of initial transcription steps determined from the temperature dependences of the individual $\mathbf{k_i}$ (SI Appendix, Fig. S7) are also shown.

Right: Structural Representation of an Initial Transcribing Complex (ITC). This ITC framework structure (adapted from 4YLP) employed a promoter heteroduplex in the region of the initiation bubble and a 5-mer RNA hybridized to the template strand (RNA in red; promoter DNA in blue). Relevant regions of σ^{70} include $\sigma_{1.2}$ (green), σ_2 (pink), σ_3 (orange), $\sigma_{3.2}$ (light blue), and σ_4 (yellow). Letters **a, b,** and **c** indicate the discriminator and -10 regions of the bubble strands and the -35 duplex respectively. The expected

direction of translocation of the hybrid into the RNAP cleft in initiation is indicated by the black arrow. Translocation is proposed to disrupt RNAP-promoter contacts with regions **a, b,** and **c** in the steps of initiation with small rate constants (indicated by the corresponding letters on the k_i plot (left)), resulting in stepwise bubble collapse.

References

1. Saecker RM, Record MTJ, & deHaseth PL (2011) Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of Molecular Biology* 412(754-771).

2. Haugen SP, Ross W, Manrique M, & Gourse RL (2008) Fine structure of the promoter-sigma region 1.2 interaction. *Proceedings of the National Academy of Sciences* 105(9):3292-3297.

3. Dorman C (2013) Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription. *Biochemical Society Transactions* 41(2):542-547.

4. Chen J, Boyaci H, & Campbell EA (2021) Diverse and unified mechanisms of transcription initiation in bacteria. *Nature Reviews Microbiology* 19(2):95-109.
5. Chen J*, et al.* (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Molecular Cell* 78(2):275-288.e276.

6. Zuo Y & Steitz TA (2015) Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Molecular Cell* 58:534-540.

7. Li L, Molodtsov V, Lin W, Ebright RH, & Zhang Y (2020) RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. *Proceedings of the National Academy of Sciences*:201920747.

8. Shin Y*, et al.* (2021) Structural basis of ribosomal RNA transcription regulation. *Nature Communications* 12(1).

9. Glyde R*, et al.* (2018) Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Molecular cell* 70(6):1111-1120.e1113.

10. Ruff EF*, et al.* (2015) *E. coli* polymerase determinants of open complex lifetime and structure. *Journal of Molecular Biology* 427:2435-2450.

11. Ruff EF, Record MT, Jr. , & Artsimovitch I (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5:1035-1062.

12. Ko J & Heyduk T (2014) Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. *Biochemistry Journal* 463:135-144.

13. Duchi D*, et al.* (2016) RNA polymerase pausing during initial transcription. *Molecular Cell* 63:939-950.

14. Hubin EA*, et al.* (2017) Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife* 6.

15. Henderson KL*, et al.* (2017) Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proceedings of the National Academy of Sciences* 114(15):E3032-E3040.

16. Henderson KL*, et al.* (2019) RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. *Biochemistry* 58:2339-2352.

17. Sreenivasan R*, et al.* (2020) Fluorescence-Detected Conformational Changes in Duplex DNA in Open Complex Formation by E. coli RNA Polymerase: Upstream Wrapping and Downstream Bending Precede Clamp Opening and Insertion of the Downstream Duplex. (Cold Spring Harbor Laboratory).

18. Jensen D, Manzano AR, Rammohan J, Stallings CL, & Galburt EA (2019) CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. *Nucleic Acids Research* 47(13):6685-6698.

19. Lerner E*, et al.* (2016) Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences* 113:E6562-E6571.

20. Heyduk E & Heyduk T (2018) DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Res* 46(9):4469-4486.

21. Winkelman JT*, et al.* (2019) XACT-seq comprehensively defines the promoter-position and promoter-sequence determinants for initial-transcription pausing. (Cold Spring Harbor Laboratory).

22. Han L*, et al.* (2019) Development of a novel strategy for robust synthetic bacterial promoters based on a stepwise evolution targeting the spacer region of the core promoter in Bacillus subtilis. *Microbial Cell Factories* 18(1).

23. Libing Y*, et al.* (2017) The mechanism of variability in transcription start site selection. *Elife* 6(e32038).

24. Heyduk E & Heyduk T (2014) Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. *Biochemistry* 53(2):282-292.

25. Kontur WS, Saecker RM, Davis CA, Capp MW, & Record MT (2006) Solute Probes of Conformational Changes in Open Complex (RPo) Formation byEscherichia coliRNA Polymerase at the λPRPromoter: Evidence for Unmasking of the Active

Site in the Isomerization Step and for Large-Scale Coupled Folding in the Subsequent Convers. *Biochemistry* 45(7):2161-2177.

26. Kontur WS, Capp MW, Gries TJ, Saecker RM, & Record MTJ (2010) Probing DNA binding, DNA opening and assembly of a downstream clamp/jaw in E. coli RNA polymerase-lP$_R$ promoter complexes using salt and the physiological anion glutamate. *Biochemistry* 49(20).

27. Kontur WS, Saecker RM, Capp MW, & Record MTJ (2008) Late steps in the formation of *E. coli* RNA polymerase-lP$_R$ promoter open complexes: characterization of conformational changes by rapid [perturbant] upshift experiments. *Journal of Molecular Biology* 376:1034-1047.

28. Drennan A*, et al.* (2012) Key roles of the downstream mobile jaw of *Escherichia coli* RNA polymerase in transcription initiation. *Biochemistry* 51:9447-9459.

29. Zhang Y*, et al.* (2012) Structural Basis of Transcription Initiation. *Science* 338(6110):1076-1080.

30. Revyakin A, Ebright RH, & Strick TR (2004) Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proceedings of the National Academy of Sciences* 101(14):4776-4780.

31. Kapanidis AN*, et al.* (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314:1144-1147.

32. Revyakin A, Liu C, Ebright RH, & Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314:1139-1143.

33. Winkelman JT*, et al.* (2015) Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. *Molecular Cell* 59:768-780.

34. Feklistov A & Darst SA (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase s subunit. *Cell* 147(6):1257-1269

.

35. Stackhouse TM, Telesnitsky AP, & Meares C, F. (1989) Release of the s subunit from *Escherichia coli* RNA polymerase transcription complexes is dependent on the promoter sequence. *Biochemistry* 28:7781-7788.

36. Koh HR*, et al.* (2018) Correlating Transcription Initiation and Conformational Changes by a Single-Subunit RNA Polymerase with Near Base-Pair Resolution. *Molecular Cell* 70(4):695-706.e695.

37. Saecker RM*, et al.* (2002) Kinetic studies and structural models of the association of *E. coli* sigma 70 RNA polymerase with the LPR promoter: large scale conformational changes in forming the kinetically significant intermediates. *Journal of Molecular Biology* 319:649-671.

38. Gries TJ, Kontur WS, Capp MW, Saecker RM, & Record MTJ (2010) One-step DNA melting in the RNA polyerase cleft opens the initiation bubble to form an unstable open complex. *Proceedings of the National Academy of Sciences* 107(23):10418-10423.

39. Mejia YX, Mao H, Forde NR, & Bustamante C (2008) Thermal Probing of E. coli RNA Polymerase Off-Pathway Mechanisms. *Journal of Molecular Biology* 382(3):628-637.

40. Berg OG & Von Hippel PH (1985) Diffusion-Controlled Macromolecular Interactions. *Annual Review of Biophysics and Biophysical Chemistry* 14(1):131-158.

41. Holbrook JA, Capp MW, Saecker RM, & Record MT (1999) Enthalpy and Heat Capacity Changes for Formation of an Oligomeric DNA Duplex: Interpretation in Terms of Coupled Processes of Formation and Association of Single-Stranded Helices †. 38(26):8409-8422.

42. Kozlov AG & Lohman TM (1999) Adenine base unstacking dominates the observed enthalpy and heat capacity changes for the *Escherichia coli* SSB tetramer binding to single-stranded oligoadenylates. *Biochemistry* 38:7388-7397.

Chapter 2 Appendix: Supplemental Methods, Tables, and Figures

Methods

Reagents, Buffers, and Gels

Reagents used for buffers and stock solutions were purchased in the highest available grade and used as received. All solutions were prepared using 18MΩ deionized water from a Barnstead EPure system. NTPs and dNTPs (Boston Bioproducts, Thermo Fisher, New England Biolabs) used in transcription assays and PCR reactions were 99% pure and used as received. Enzymes for PCR reactions were purchased from NEB and used according to the manufacturers protocols.

Storage Buffer for core RNAP, σ^{70} and RNAP holoenzyme is 50% v/v glycerol, 0.01M Tris, 0.1 M NaCl, 0.1 mM EDTA, and 0.1 mM DTT. Transcription buffer (TB) is 40 mM Tris, 5 mM $MgCl_2$, 60 mM KCL, 1 mM DTT, 0.05 mg/ml BSA, and is adjusted to pH 8.0 at experimental temperature (19°C, 25°C, or 37°C).

2X initiation solution (IS) for rapid quench flow (RQF) transcription assays at 25 °C and 37 °C with α-^{32}P-GTP is 0.1 mg/ml heparin, 400 μM ATP, 400 μM UTP, 20 μM unlabeled GTP, and 35 nM α-^{32}P-GTP. For 25 °C and 37 °C assays with α-^{32}P-UTP, IS is 0.1 mg/ml heparin, 400 μM ATP, 400 μM GTP, 20 μM unlabeled UTP, and 35 nM α-^{32}P-UTP. IS is mixed 1:1 with pre-formed OC to reach experimental concentrations of each NTP.

Quench Solution (QS) for RQF transcription assays is 8M Urea and 15 mM EDTA in TB. QS with added dyes (QSD) for polyacrylamide gel electrophoresis (PAGE) has

0.05% xylene cyanol and 0.05% bromphenol blue in QS. TBE buffer for PAGE is 90mM Tris-borate (pH 8.3) and 2mM Na_2EDTA. All transcription gels are 20% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics).

RNA Polymerase and λP_R Promoter DNA

RNAP core enzyme ($\alpha_2\beta\beta'\omega$) is overexpressed from pVS10 plasmid and purified via Ni affinity chromatography [1]. The σ^{70} subunit is overexpressed from pIA586 and purified using Ni affinity chromatography [1]. Holoenzyme is reconstituted using a 1:2 ratio of core to σ^{70}. Filter binding activity assays performed on preparations of RNAP holoenzyme used here show that 50% ± 10% of RNAP molecules form a stable open complex with the λP_R promoter at 37 °C. All RNAP concentrations reported here refer to this active fraction. λP_R promoter DNA is prepared as described previously [2]. Primer sequences used to assemble the λP_R promoter DNA fragment are reported in SI Appendix Table S1.

Initiation Kinetic Assays

Initiation kinetic assays are performed and imaged as described previously [1]. Assays are designed to obtain single-round synthesis of full length (FL) RNA, defined for this study as any RNA length greater than 10-mer (see main text). Briefly, 2X initiation solution (IS) is mixed 1:1 with preformed OC at time zero using a KinTek Corp. Rapid Quench Flow (RQF). Reactions are quenched with QS and RNA is separated via PAGE. Both the IS and preformed OC solutions are incubated at the experimental temperature (19 °C, 25 °C, 37 °C) for 1 h prior to mixing in the RQF at that temperature. Transcription gels are transferred to a phosphorimaging cassette and analyzed using a Typhoon 9000

phosphorimager (18 h exposure). Peak area is converted to moles of observed product as described previously [2]. To obtain good labelling efficiency without the use of high concentrations of labelled NTP that cause background problems and limit the ability to detect small RNAs on the gels, we use a low concentration of ^{32}P-labelled NTP (17.5 nM) and a reduced concentration (10 μM) of the corresponding unlabeled NTP. The probability of incorporation of the radiolabeled NTP at each position is determined by the ITR sequence, and was accounted for using incorporation probabilities (see SI Appendix, Table S2) calculated for different RNA lengths as described previously [2]. Quantitative results reported here are from the average of 2-4 RQF experiments at each set of NTP concentrations investigated, and the uncertainties reported are estimates of one standard deviation from the mean.

Analysis of RNA Synthesis as a Function of Time

The amount of FL RNA and of the eight detectable short RNA species (3-mer to 10-mer) are calculated as a function of time from phosphorimager line scans of transcription gels. The only undetected short RNA is 2-mer (pppApU), which is not labelled in α-^{32}P-GTP labeled experiments and does not separate sufficiently from the broad band of unincorporated monomer to quantify in α-^{32}P-UTP labelled experiments. Nevertheless, kinetic parameters (K_m, k_{cat}) of 2-mer synthesis are obtained from fitting the kinetic data for the longer RNAs to the initiation mechanism (Fig. 3).

Analysis of Short RNAs Synthesized by Nonproductive Complexes

Kinetics of synthesis of the first short RNA of each length by nonproductive (stalled) complexes are well-described as a first order (single exponential) approach to a plateau

value, as is the case for FL RNA [1]. For each short RNA length, at each time point during FL RNA synthesis, the amount of short RNA synthesized by nonproductive complexes is predicted using Origin 2018b and subtracted from the observed total amount of that RNA to obtain the amount of transient short RNA synthesized by productive complexes. This dissection is shown in Figures S3-S6 for the temperatures and NTP conditions investigated. Amounts of RNA transients determined by this procedure are plotted in Fig. 3.

Supplemental Information for Figures S3-S6

Amounts of each detectable short RNA species (3-mer to 10-mer) formed prior to promoter escape at the low-UTP and high-UTP conditions at 25 ^{o}C and 37 ^{o}C are plotted vs. time (linear scale) in Figs. S3-S6. At 25 ^{o}C, high UTP, significant transient peaks in amounts of all short RNAs except 7-mer are observed. This is consistent with previously-reported observations at 19 ^{o}C, high UTP [1]. Transient buildup of N-mer RNA is expected when the synthesis of N-mer is relatively rapid and the subsequent synthesis of (N+1)-mer is slow. At high UTP (and low GTP) the first step of initiation (pppApU synthesis) is rapid and not a bottleneck, and intermediates (5-mer, 9-mer, 10-mer) preceding steps where G is added are expected to accumulate because of the low GTP concentration. Accumulation of 3-mer, 4-mer, 6-mer and 8-mer must however indicate that the rate constants of the following steps of RNA synthesis (synthesis of 4-mer, 5-mer, 7-mer and 9-mer) are small. This was previously observed at this NTP condition at 19 ^{o}C, where rate constants for synthesis of these four RNAs as well as 8-mer and 11-mer are smaller than for other steps of initiation. By contrast, at 25 °C and low UTP, where synthesis of pppApU is slowed by the low UTP concentration, significant transient

amounts of only three RNAs (3-mer, 6-mer, 10-mer) are observed. By the same reasoning as above, the 10-mer transient indicates the rate constant for conversion to 11-mer is small, while 3-mer and 6-mer transients are interpreted in terms of slow rates for the following steps, both because the next base incorporated is U and because the rate constants of these steps are small. These results are consistent with previous findings at 19 °C (Henderson '19).

In contrast, at the high UTP condition at 37°C only four transient RNA intermediates (5-mer, 6-mer, 9-mer, 10-mer) from productively-initiating complexes are detected (Fig. S6), and no transient RNAs are detected at the low UTP condition (Fig. S5). These observations are consistent with the proposal that an additional step at the beginning of the mechanism slows the kinetics of FL RNA synthesis at 37 ^{o}C, especially at the low UTP condition.

Supplemental Information for Figure S7 and Table S5

Values of ln k_i for each step of initiation are plotted vs. 1/T in Fig. S7. Some 37 °C k_i values (for synthesis of 3-mer, 4-mer, 7-mer, and 8-mer) are lower bounds (as indicated in Table S4). When plotted with well-determined 19 °C and 25 °C k_i values, these 37 °C lower-bound k_i values give linear Arrhenius plots, and we accept them as the best estimates of these 37 °C k_i values. Arrhenius activation energies $E_{Act,\,i}$ calculated from slopes of plots in Fig. S7 (slope = $-E_{act,\,i}/R$) are plotted in Fig. 4 and listed in Table S5.

Activation free energies ($\Delta G_i^{o\ddagger}$) for each step of RNA-DNA hybrid extension are calculated from the equation

$$\ln k_i = \ln k_{max} - (\Delta G_i^{o\ddagger}/RT) \qquad \text{Eq. S1}$$

where $\Delta G_i^{o\ddagger}$ is the free energy difference between the high free energy transition state of the catalytic step and the reactants (incoming NTP, pre-translocated state of the initiation complex) and k_{max} is the maximum rate constant for the hypothetical situation where $\Delta G_i^{o\ddagger} = 0$. As an estimate of the maximum second order rate constant k_{max}, we use a plausible value of a diffusion-limited (orientation-corrected) rate constant for NTP binding ($k_{max} \approx 10^3\ \mu M^{-1}\ s^{-1}$). The uncertainty in this estimate, though large, is inconsequential for the analysis reported here. We assume that k_{max} (as a diffusion-limited rate constant) is the same for each step of initiation and neglect its expected small temperature dependence.

From k_i values as a function of temperature (Fig. S7) and k_{max}, activation free energies and their enthalpic and entropic components are obtained for each initiation step:

$$\Delta G_i^{o\ddagger} = -RT \ln (k_i/k_{max}) = \Delta H_i^{o\ddagger} - T\Delta S_i^{o\ddagger}$$

Where

$$\Delta H_i^{o\ddagger} \approx E_A = -R \left(\frac{\partial \ln k_i}{\partial\left(\frac{1}{T}\right)}\right) \text{ and } \Delta S_i^{o\ddagger} = (\Delta H_i^{o\ddagger} - \Delta G_i^{o\ddagger})/T \qquad \text{Eq. S2}$$

Values of $\Delta G_i^{o\ddagger}$, $\Delta H_i^{o\ddagger}$, $\Delta S_i^{o\ddagger}$ and their uncertainties for each step of extension of the RNA-DNA hybrid by NTP incorporation at 19 °C are listed in Table S5 and values of $\Delta H_i^{o\ddagger}$ are interpreted in Table S6.

Table S1: Primers used for λP_R template preparation

λP_R_wt_forward (-71 to -12)	CCACGAATTCGGATAAATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTG
λP_R_wt_reverse (-24 to 31)	ACAAAACCTTCATAGAACCTCCTTACTACATGCAACCATTATCACCGCCAGAGGT
HBOT	CACCTGCACCGACAAAACCTT
HTOP	CCAGCATTCCTCCACGAATTC

Table S2: Probabilities of Incorporation of α-^{32}P-UTP and α-^{32}P-GTP

Number of UTP or GTP in Transcript	Transcript Length (UTP)	Transcript Length (GTP)	α-^{32}P-NTP Incorporation Probability
0	-	2-mer	0
1	2-mer, 3-mer	3-mer to 5-mer	0.00175
2	4-mer to 6-mer	6-mer to 9-mer	0.0035
3	7-mer to 14-mer	10-mer	0.00525
4	15-mer	11-mer to 12-mer	0.007
5	16-mer	13-mer	0.00875
6	ND[a]	14-mer to 16-mer	0.0105
10	ND[a]	31-mer	0.0175
12	31-mer	-	0.021

[a] Not detected

Table S3: Lag Times and First Order Rate Constants k_{FL} for FL (> 10-mer) RNA Synthesis

	Low UTP Lag Time	High GTP Lag Time	Low UTP k_{FL}	High UTP k_{FL}
19 °C	3.6 s	2.3 s	0.062 s^{-1}	0.11 s^{-1}
25 °C	2.2 s	1.2 s	0.082 s^{-1}	0.22 s^{-1}
37 °C	2.0 s	0.35 s	0.048 s^{-1}	0.39 s^{-1}

Table S4: Composite Rate Constants k_i ($\mu M^{-1} s^{-1}$) for Each Initiation Step at 19, 25, and 37 °C [a]

Base Position i	Base Incorporated	19 °C [b] k_i ($\mu M^{-1} s^{-1}$)	25 °C k_i ($\mu M^{-1} s^{-1}$)	37 °C [c] k_i ($\mu M^{-1} s^{-1}$)
2	U	0.63 ± 0.22	0.69 ± 0.24	1.5 ± 0.5
3	G	0.15 ± 0.08	0.49 + 0.64	≥ 2
4	U	0.012 ± 0.001	0.033 ± 0.005	≥ 0.2
5	A	0.064 ± 0.044	0.077 ± 0.040	0.11 ± 0.07
6	G	0.16 ± 0.04	0.21 ± 0.04	0.24 ± 0.08
7	U	0.017 ± 0.002	0.021 ± 0.003	≥ 0.03
8	A	0.037 ± 0.020	0.044 ± 0.022	≥ 0.07
9	A	0.052 ± 0.035	0.043 ± 0.021	0.033 ± 0.024
10	G	0.13 ± 0.03	0.22 + 0.06	0.62 ± 0.17
11	G	0.028 ± 0.003	0.049 ± 0.006	0.21 ± 0.03

[a] From fitting kinetic data at each temperature to Mechanism 2 (Fig. 3).

[b] Previously reported 19 °C ki from fitting to Mechanism 1 [1] are not significantly different.

[c] Uncertainties in 37 °C k_i values could not be determined from the fit because of the insufficient number of detectable small RNA transient intermediates (Figs. S5-S6). Uncertainties in other 37 °C k_i are estimated using the larger of the % uncertainties for that step in the 19 °C and 25 °C data.

Table S5: Arrhenius Activation Energies $E_{A,i}$ and Transition State Barriers $\Delta G_i^{o\ddagger}$, $\Delta H_i^{o\ddagger}$, $\Delta S_i^{o\ddagger}$ for Each Step of Initiation[a]

Base Position i	k_i [b] ($\mu M^{-1} s^{-1}$)	$\Delta G_i^{o\ddagger}$ [c] (kcal/mol)	$E_{A,i} \approx \Delta H_i^{o\ddagger}$ (kcal/mol)	$\Delta S_i^{o\ddagger}$ (eu)
2	0.63 ± 0.22	4.3 ± 0.3	9 ± 5	~ 16
3	0.15 ± 0.08	5.1 ± 0.8	27 ± 14	~ 75
4	0.012 ± 0.001	6.6 ± 0.1	26 ± 2	~ 66
5	0.064 ± 0.044	5.6 ± 0.3	5.1 ± 10	~ -2
6	0.16 ± 0.04	5.1 ± 0.1	7.2 ± 3	~ 7
7	0.017 ± 0.002	6.4 ± 0.1	6.8 ± 2	~ 1
8	0.037 ± 0.020	5.9 ± 0.3	6.6 ± 7	~ 2
9	0.052 ± 0.039	5.7 ± 0.3	-4.4 ± 10	~ -35
10	0.13 ± 0.03	5.2 ± 0.2	16 ± 4	~ 37
11	0.028 ± 0.003	6.1 ± 0.1	21 ± 2	~ 51

[a] Activation free energy, enthalpy, and entropy changes $\Delta G_i^{o\ddagger}, \Delta H_i^{o\ddagger}, and\ \Delta S_i^{o\ddagger}$ are defined in Eq. S1-S2.

[b] Determined at 19 °C ([1]; see Table S4).

[c] Calculated from 19 °C k_i values using Eq. S1 with $k_{max} = 10^3\ \mu M^{-1} s^{-1}$

Table S6: Interpretation of Rate Constants k_i and Barriers $\Delta G_i^{o\ddagger}$, $\Delta H_i^{o\ddagger}$, $\Delta S_i^{o\ddagger}$ for Each Step of Initiation

Base Position i	k_i/k_3	Proposed Region where RNAP-DNA Contacts are Disrupted in step i [a]	Proposed Number of Bubble Bases that Stack in Step i [b]	Proposed Positions of Duplex Formation (relative to +1 TSS) [c]
2	~ 4.2	-	0	-
3	1	-	0	-
4	~ 0.1	Discriminator	0	-
5	~ 0.4	Discriminator	4	-1, -2
6	~ 1	-	4	-3, -4
7	~ 0.1	-10	4	-5, -6
8	~ 0.2	-10	4	-7, -8
9	~ 0.3	-10	6	-9, -10, -11
10	~ 0.9	-	2	-
11	~ 0.2	-35	0	-

[a] Based on analysis in [1]. Reductions in k_i values for steps 3, 6, and 10 as compared to step 2 (no translocation) cannot be dissected based on activation thermodynamics (Table S5) but may include disruption of some discriminator or -10 contacts.

[b] The proposed number of bases that convert from unstacked to stacked in each step is estimated from differences in activation enthalpies (Table S5): Number of bases that stack in step i = ($\Delta H_3^{o\ddagger}$- $\Delta H_i^{o\ddagger}$)/5, assuming -5 kcal/mol per base stacking interaction. Similar results are obtained from analysis of ($\Delta S_3^{o\ddagger}$- $\Delta S_i^{o\ddagger}$) assuming an entropy change of approximately -15 to -20 eu per stacking interaction.

[c] These predictions assume that disruption of RNAP-strand contacts and duplex formation occur first at the downstream edge of the discriminator and propagate upstream, and that formation of one base pair has the same enthalpy change (-10 kcal mol^{-1}) as stacking of two bases.

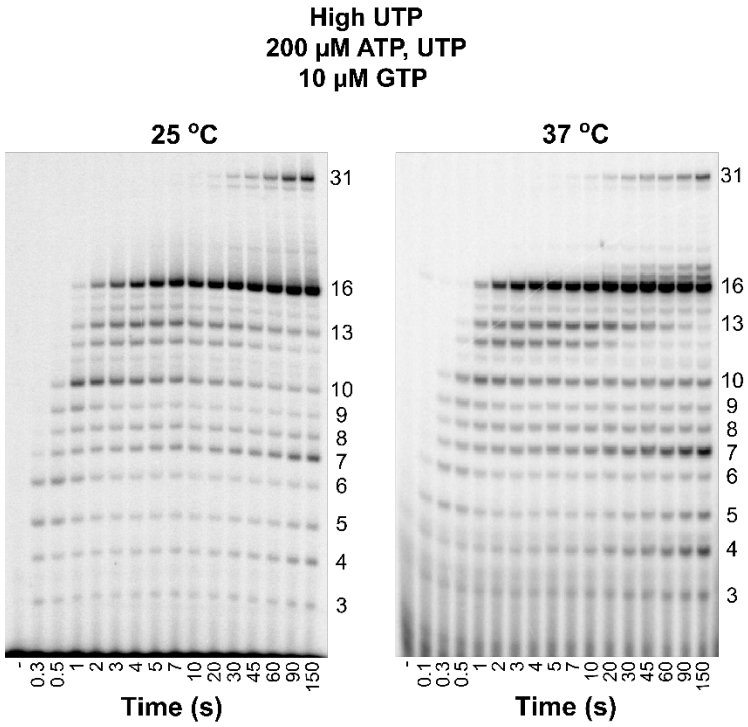


Figure S1: Step-by-Step Kinetics of Transcription Initiation from λP_R Promoter at the High-UTP Condition. Polyacrylamide gel electrophoretic separations of individual ^{32}P-labelled RNA bands (from 3-mer to 16-mer and 31-mer) observed as a function of time during productive and nonproductive initiation at 25 °C **(left)** and 37 °C **(right)**. Lanes span the time range from 0.1 or 0.3 s to 150 s after adding NTPs and heparin to OC formed by premixing RNAP and λP_R promoter DNA (See SI Methods). Gels shown are for the "high UTP" condition: 200 µM ATP and UTP, 10 µM GTP, 17.5 nM α-^{32}P-GTP. CTP is omitted causing transcription to halt at a 16-mer RNA, before read-through occurs to synthesize longer transcripts [1]. Representative "low UTP" (200 µM ATP and GTP, 10 µM UTP, 17.5 nM α-^{32}P-UTP) gels at 25 °C and 37 °C are shown in the main text Fig. 2.

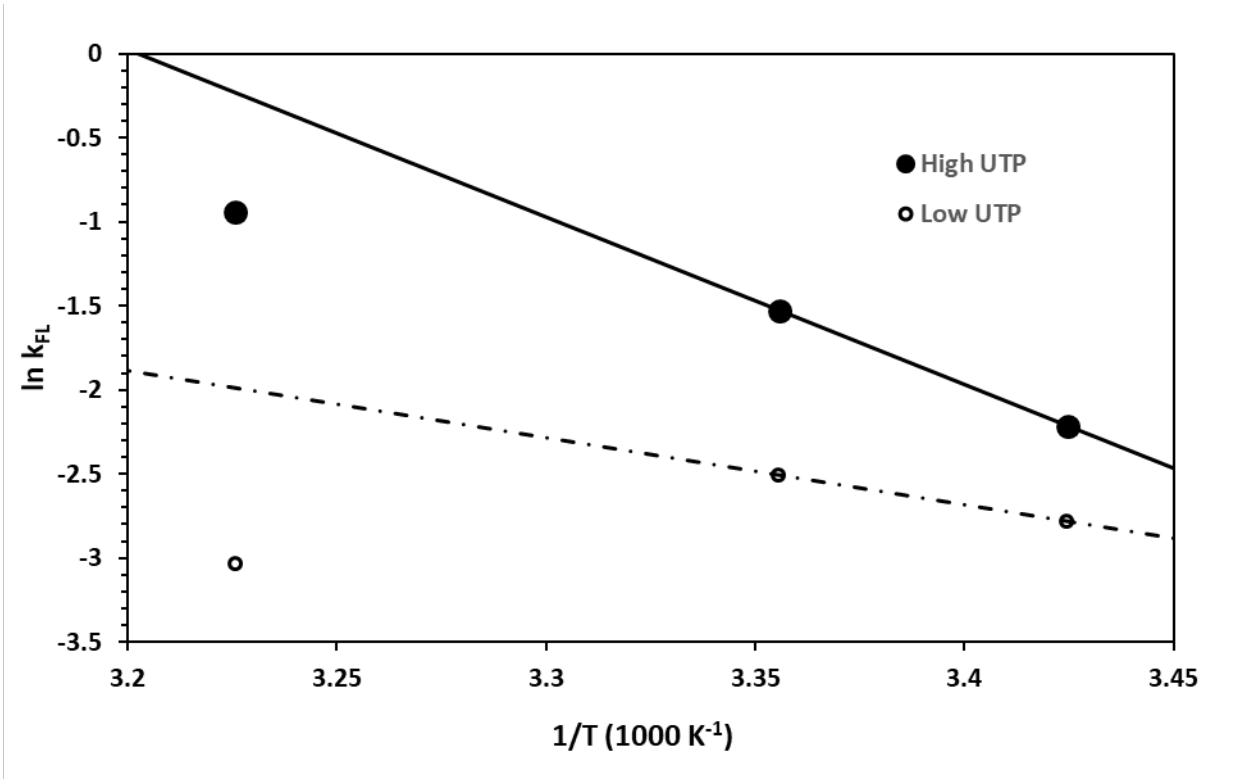


Figure S2: Arrhenius Plots of lnk_{FL} vs 1/T. Natural logarithms of first order rate constants for full length RNA production (k_{FL}, in s^{-1}) are plotted vs inverse temperature (K^{-1}) for high UTP (200 μM UTP, 10 μM GTP; filled circles) and low UTP (10 μM UTP, 200 μM GTP; open circles). At both NTP conditions (especially low UTP), the 37 °C rate constant is significantly smaller than predicted assuming a linear Arrhenius relationship defined by the 19 °C and 25 °C rate constants (predictions shown).

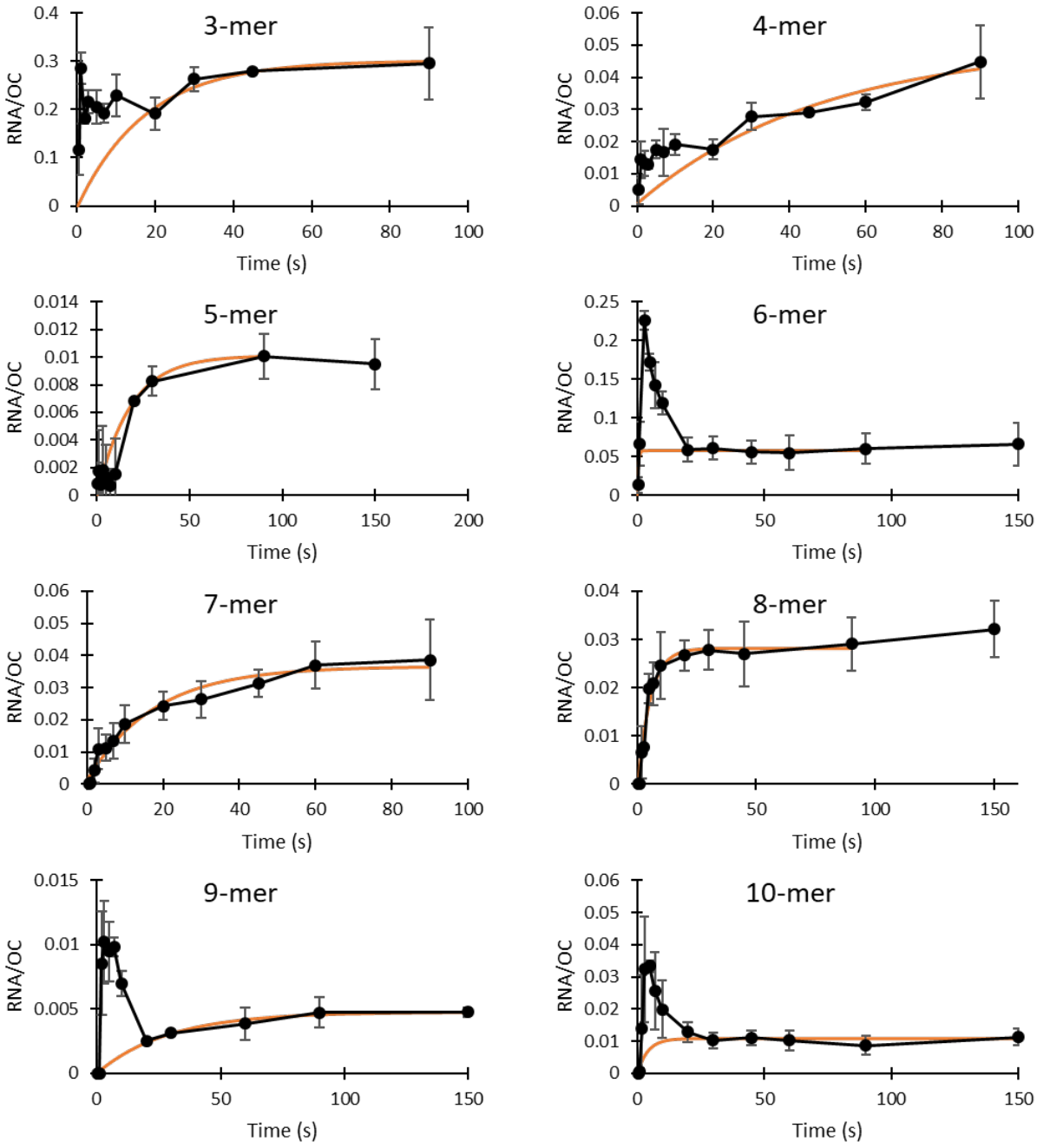


Figure S3: Average Amounts of Short RNAs vs Time at 25 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 4 independent experiments) are shown with one standard deviation error bars (black). Data for all detectable lengths up to the point of escape are shown. NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. For

each RNA length, the predicted contribution from the initial round of synthesis by non-productive complexes is also shown (orange).

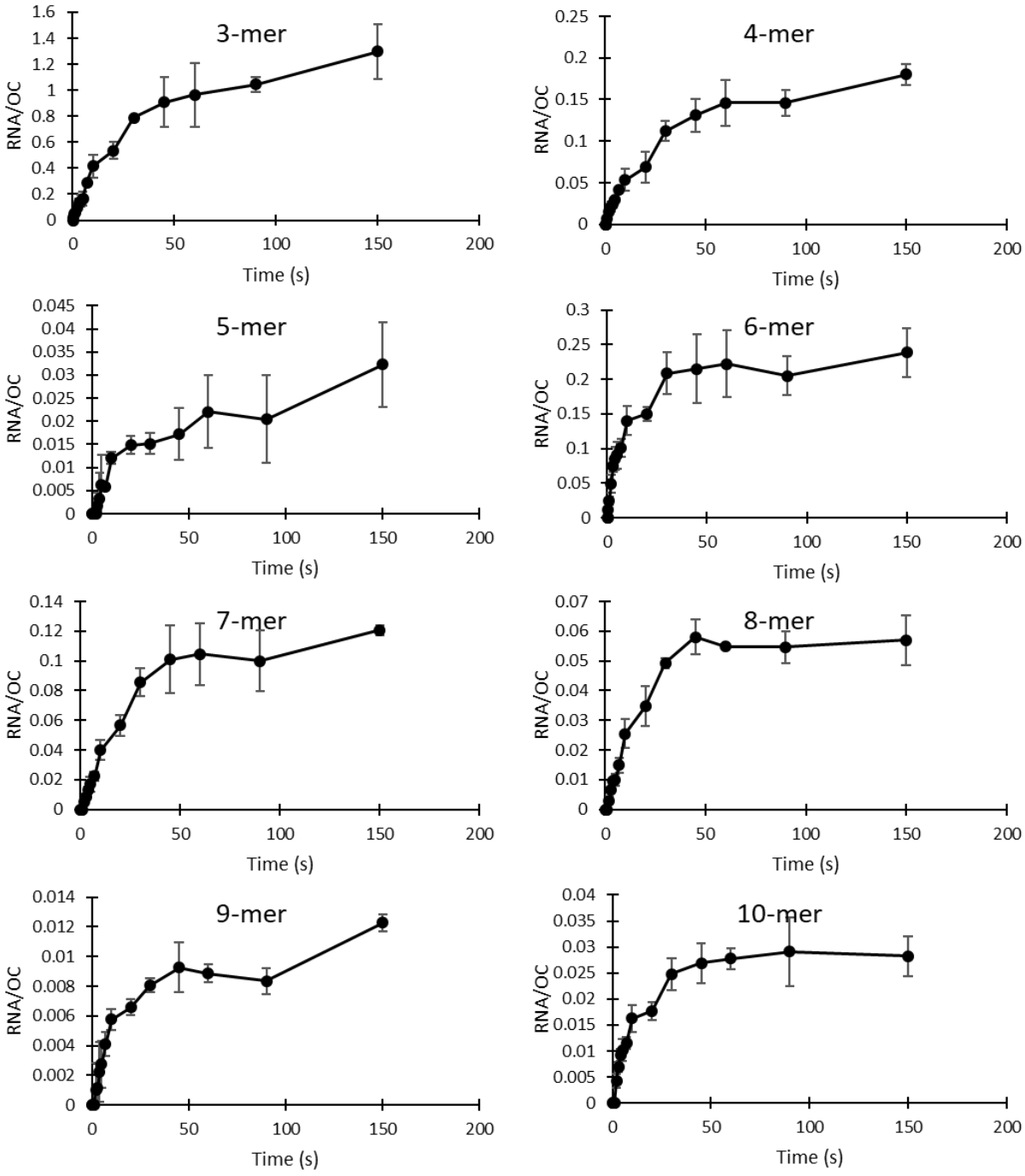


Figure S5: Average Amounts of Short RNAs vs Time at 37 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). Data for all detectable lengths up to the point of escape are shown. NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. Transient concentrations of short RNA synthesized by productive OC are too small to

detect at 37 °C, low UTP, and therefore these curves show the initial and subsequent rounds of RNA synthesis by the non-productive fraction of OCs.

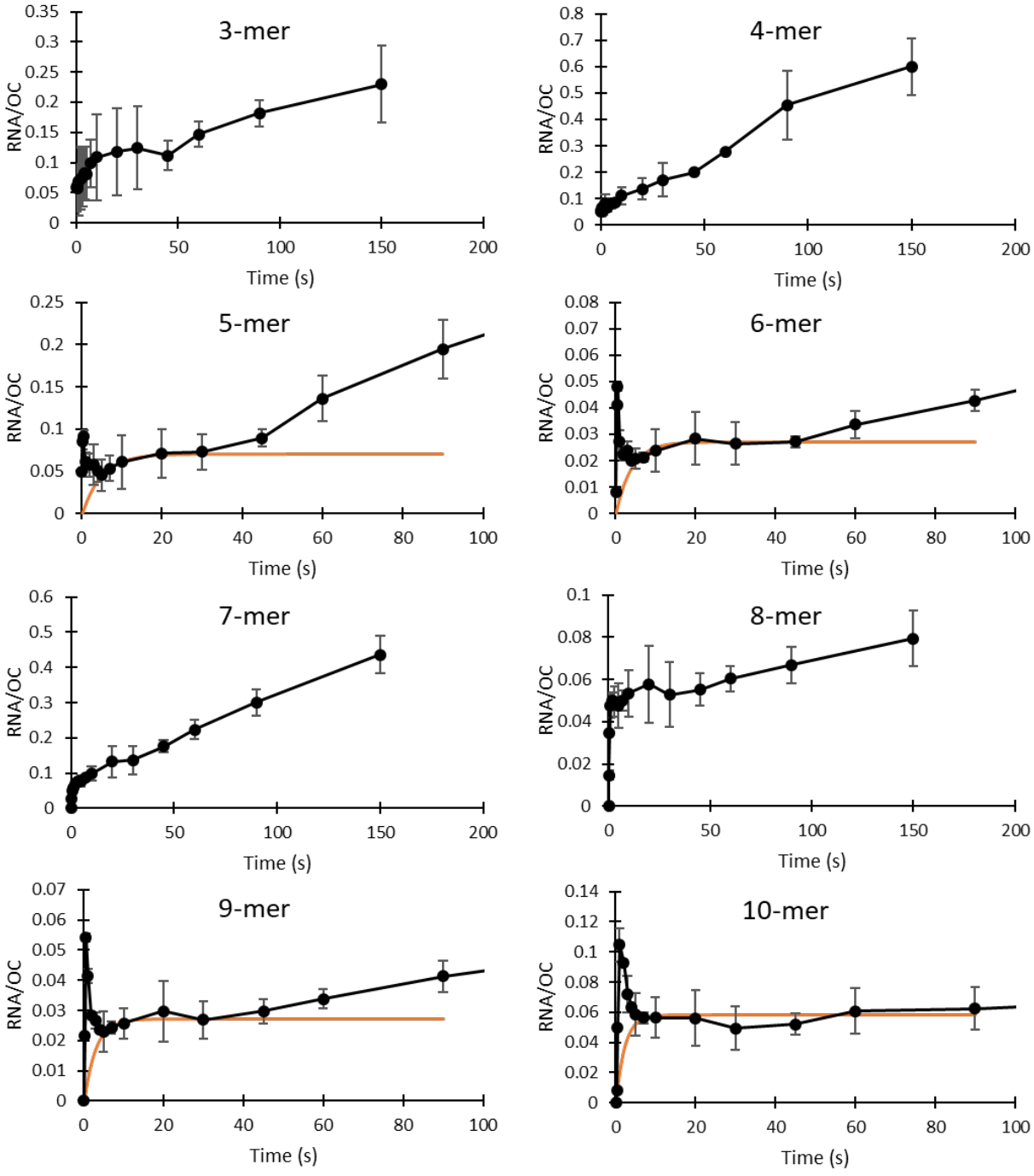


Figure S6: Average Amounts of Short RNAs vs Time at 37 °C, High UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 2 independent experiments) are shown with error bars representing the range (black). Data for all detectable lengths up to the point of escape are shown. NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. For

each RNA length where a transient concentration of RNA from synthesis by productive complexes is observed, the predicted contribution from the initial round of synthesis by non-productive complexes is also shown (orange).

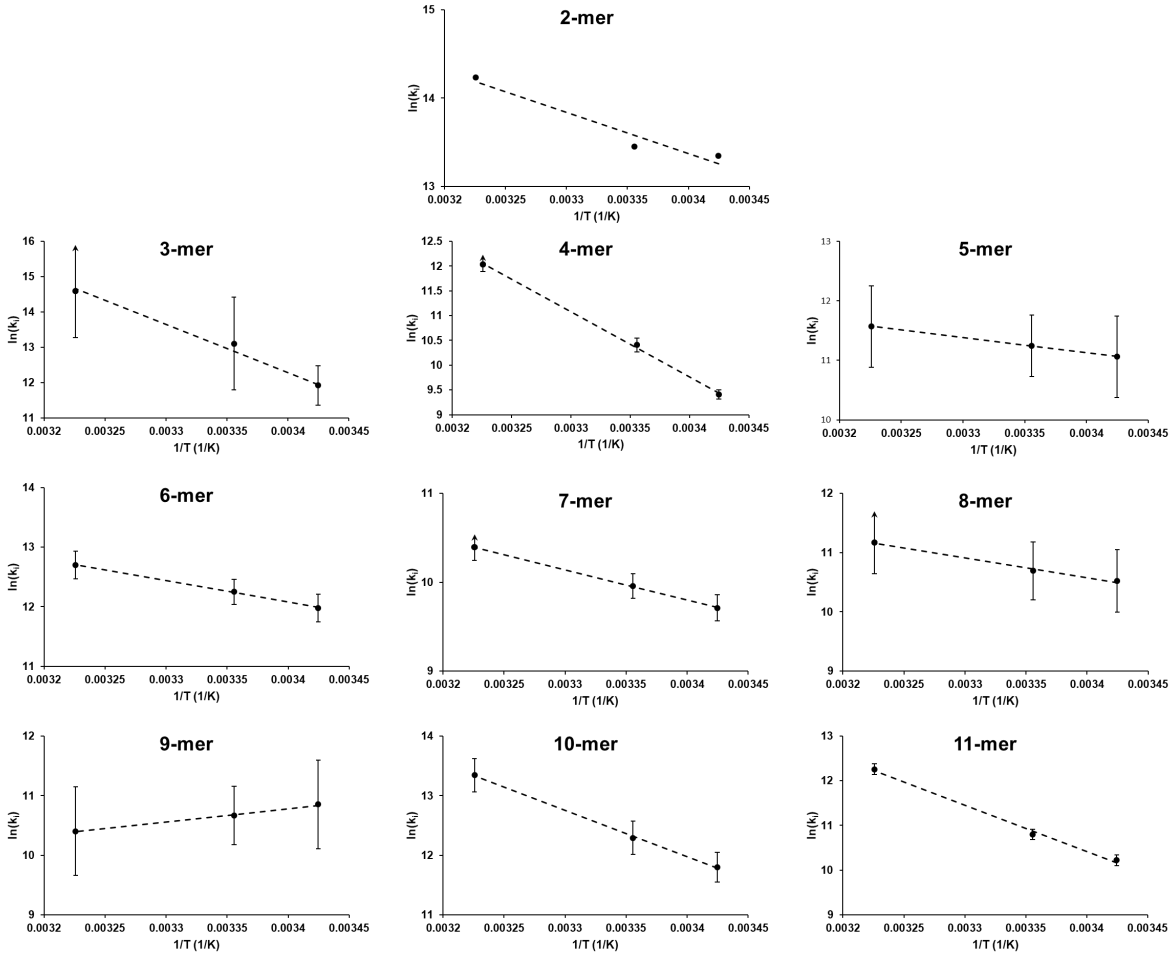


Figure S7: Temperature Dependences of Composite Initiation Rate Constants k_i: Arrhenius Plots

For each step i, values of $\ln k_i$ (with k_i in $M^{-1}s^{-1}$) are plotted vs 1/T. Uncertainties are listed in Table S4. Slopes of linear fits yield Arrhenius activation energies $E_{act,\,i}$ (slope = $-E_{act,i}/R$), reported in Table S5.

Supplemental References

1. Henderson KL*, et al.* (2019) RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. *Biochemistry* 58:2339-2352.

2. Henderson KL*, et al.* (2017) Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proceedings of the National Academy of Sciences* 114(15):E3032-E3040.

Chapter 3: Step-by-Step Regulation of Productive and Abortive Transcription Initiation by Pyrophosphorolysis

This chapter was accepted for publication in the Journal of Molecular Biology.

Experiments for this chapter were conceptualized by Kate Henderson, Claire Evensen, Tom Record, and Dylan Plaskon. Experiments were performed by Claire Evensen, Benjamin Palatnik, Taka Ishikuri, Sarah Doughty, and Olivia Hamzelou. Data analysis and interpretation were performed by Dylan Plaskon, Hao-Che Wang, and Tom Record. The manuscript was written by Dylan Plaskon and Tom Record.

Abstract

An understanding of the kinetics and mechanism of bacterial transcription initiation is needed to understand regulation of gene expression and advance fields from antibiotic discovery to promoter design. The step-by-step forward kinetics and mechanism of initiation and RNA-DNA hybrid growth, made irreversible by omitting pyrophosphate (PPi) byproduct, were determined recently for *E. coli* RNA polymerase (RNAP)-λP_R promoter complexes. Strong position-dependences of overall rate constants (k_{cat}/K_m analogs) for each nucleotide-addition step were observed because of coupling of hybrid growth to disruption of promoter contacts, bubble closing, and RNAP escape. Here we investigate reversal of these steps (pyrophosphorolysis) at PPi concentrations ([PPi]) found in exponentially-growing cells. We quantify [PPi] effects on the amount and rate of synthesis of long (>10-mer, post-escape) and short (stalled, abortive) RNA to determine how PPi regulates initiation. Physiological [PPi] makes uridine incorporation and some other initiation steps significantly reversible. Physiological [PPi] reduces the fraction of RNAP-promoter complexes that productively initiate and the rate of RNA synthesis per productive complex, while increasing the fraction of complexes that abortively initiate, affecting abortive rates, and shifting the abortive-product distribution to shorter RNAs. Pyrophosphorolysis rates for some initiation complexes are orders of magnitude larger than for removal of the same nucleotide from elongation complexes because of the strong bias toward the pre-translocated state in initiation, and exhibit even stronger dependences on nucleotide identity (pyrimidine >> purine). Because cytoplasmic [PPi] is much higher in exponential-phase than stationary-phase cells, these [PPi] effects on initiation rates and amounts of RNA synthesis must be physiologically-relevant.

Introduction

Transcription is a fundamental life process and nuanced transcriptional regulation is consequently vital. Because this regulation largely occurs in initiation, it is very important to understand the mechanism of initiation and how the kinetics and equilibria of individual steps are determined by the sequences of the promoter and initial transcribed region (ITR) and their interaction with RNA polymerase (RNAP), as well as by physiological solution variables including temperature and concentrations of reactants ([NTP]) and byproduct (pyrophosphate, PPi). Of these variables, effects of [PPi] on initiation have been least explored. This information is needed because [PPi] in the *E. coli* cytoplasm increases greatly from stationary phase (low μM) to exponential growth (~1 mM), and also varies with growth conditions [1-3].

Bacterial transcription initiation begins with recognition of each unique duplex DNA promoter sequence by RNAP [4-7]. Initial binding of *E. coli* RNAP to the λP_R promoter directs remodeling of the duplex to open 13 base pairs (bp) and form a relatively-unstable open-promoter intermediate (I_2) [7-9]. Interactions of the RNAP jaw and other mobile elements with the downstream duplex, directed by the discriminator region of the promoter, convert I_2 at some promoters to a more stable OC (I_3, RP_O) [10, 11]. The partially-stabilized open intermediate I_3 is the λP_R initiation complex (IC) that binds NTPs and begins RNA synthesis [12, 13]. Structures of two λP_R open complexes (OC) proposed to be I_3 and the 37 ^oC stable OC (RPo) were recently reported [13].

Kinetic-mechanistic studies (both ensemble [12, 14-16], and single molecule [17-19]) together with structural studies [13, 20-23], have clarified key aspects of the *E.coli* initiation mechanism in the presence of NTPs. Analysis of the kinetic data yields overall

[15, 16, 24] and step-by-step [12, 15] kinetics of irreversible hybrid growth and escape of RNAP from the promoter. After synthesis of the initiating dinucleotide, the cycle of steps to add a nucleotide to the 3' end of the RNA in initiation is well-described by the minimal mechanism of Fig. 1 [12, 15], which is the same as in elongation [25, 26]. After the previous catalytic step, reversible translocation of the RNA-DNA hybrid establishes an equilibrium distribution of pre- and post-translocated states. In the post-translocated state the active site is accessible to the incoming NTP as a result of unfolding of the trigger helix [26, 27]. Reversible NTP binding stabilizes the post-translocated state. In this step, refolding of the trigger loop closes the active site for catalysis and aids in properly positioning the incoming NTP [27]. In the catalytic step of Fig. 1 the triphosphate of the NTP is split to add the nucleotide monophosphate (NMP) to the hybrid and form the next pre-translocated state with PPi as byproduct. Without added PPi, the catalytic step in initiation is effectively irreversible. Addition of PPi increases the rate of the reverse catalytic step (Fig. 1; pyrophosphorolysis), regenerating the NTP from PPi and the 3'-NMP [26, 28, 29].

In elongation by bacterial RNAP, the observation of similar kinetics of translocation and of PPi release indicates that PPi release occurs shortly before or concurrently with translocation [30]. Pre- and post-translocated complexes are in rapid equilibrium on the timescale of NTP binding. In elongation, a bias for the post-translocated state is observed, indicating that the translocation equilibrium constant of an elongation step $K_{tr,i} > 1$ [26, 30]. In initiation the opposite situation holds. Recent results demonstrate that most translocation steps in initiation are highly unfavorable, with translocation equilibrium constants $K_{tr,i} << 1$ [12, 15]. These small $K_{tr,i}$ result from translocation stresses unique to

initiation, including opening of one downstream DNA bp (when not offset by upstream bp closing), steric stress [20, 21, 31], scrunching of the bubble strands [32-34], and disruption of RNAP-promoter contacts [12, 15].

Small $K_{tr,i}$ of initiation steps reduce the overall 2nd order forward rate constants k_i (the analog of k_{cat}/K_m for this situation) for incorporation of the next nucleotide and hybrid growth, relative to elongation. Values of k_i for most steps of nucleotide addition in initiation that begin with translocation are in the range 0.01 – 0.2 μM^{-1} s^{-1} at 19 and 25 ^oC [12], much smaller than those for initial dinucleotide synthesis (without translocation; ~ 0.6 μM^{-1} s^{-1} at 19 and 25 ^oC [12]) and for the NTP binding and catalytic steps of elongation (≥ 0.6 μM^{-1} s^{-1} at 24 ^oC [35, 36]) Conversely, translocation stress in initiation is predicted to greatly favor the reverse process of pyrophosphorolysis relative to elongation.

Effects of PPi on the amounts and rates of productive and abortive RNA synthesis and on the reversible kinetics of the individual steps of transcription initiation and escape of RNAP from the promoter have not previously been systemically investigated. Scaffolded promoter constructs with artificial mismatches that mimic the initiation complex and artificially paused complexes have been employed to investigate pyrophosphorolysis in initiation at or near the escape point [28]. Pyrophosphorolysis in elongation has been characterized in both scaffolded [30, 37, 38] and continuously-elongating complexes [39, 40]. The sensitivity of the elongation complex to [PPi] is highly sequence-specific for several bacterial RNAP, including *E. coli* [30, 38]. Elongating transcripts with a 3' U are the most sensitive to pyrophosphorolysis, with the rank order U >>> C > A > G. These preferences are proposed to arise from differences in the strength of the favorable interaction of the 3' base with the folded trigger helix [30, 38, 41]. These interactions

stabilize the pre-translocated state, increasing the rate of pyrophosphorolysis at that position. The penultimate 3' nucleotide has a smaller but significant effect, in the opposite order (G > A > C >> U). Overall 2nd order rate constants estimated from published results [38] for pyrophosphorolysis in elongation are very small (< 10^{-5} μM^{-1} s^{-1}; Table S2) for both *E. coli* RNAP and *T.th.* RNAP. For continuously-elongating complexes, addition of 1 mM PPi reduces the average elongation rate to approximately half the baseline (1 μM PPi) rate [39].

Previously we investigated initiation of a single round of long RNA synthesis at two model promoters (λP_R, T7A1 [14]). We found there are two phases of initiation and two approximately equal subpopulations of initiating complexes, designated productive and nonproductive. These findings are consistent with earlier studies of initiation time courses where long RNA synthesis was single-round [42]. Fixed-time experiments in which synthesis of long RNA is multi-round are incapable of distinguishing these subpopulations. In the faster phase of initiation (< 10 – 20 s for the NTP concentrations investigated at 37 °C), productive complexes initiate and processively extend the RNA-DNA hybrid without stalling or aborting. RNAP escapes from the promoter after synthesis of a 10-mer (λP_R) or 7-mer (T7A1) by these productive initiation complexes. In this faster phase of initiation, nonproductive initiation complexes stall after synthesizing a short RNA (10-mer or less for λP_R, 7-mer or less for T7A1). At longer times, in a second phase of initiation, some of these stalled nonproductive complexes release their RNA and reinitiate (abortive synthesis) at a rate which decreases in general with increasing RNA length and with decreasing temperature [12, 14, 15, 42].

We subsequently determined the step-by-step kinetics and mechanism of initial transcription at the model λP_R promoter [12, 15]. We found that in initiation at this promoter, overall 2nd order rate constants (k_{cat}/K_m analogs) for steps of nucleotide addition involving translocation exhibit a repeating pattern of large and small forward rate constants which we interpreted in terms of the stepwise disruption of RNAP-promoter contacts and collapse of the upstream initiation bubble. We deduced that in-cleft contacts are broken first, followed by disruption of -10 contacts and then upstream (-35) contacts, allowing promoter escape. These previous studies were done in the absence of PPi, where all catalytic steps are irreversible.

Here, from initiation kinetic measurements at 37 °C and 19 °C, we determine the large effects of [PPi] in the low-mM range on the fraction of RNAP- λP_R complexes that are productive and synthesize a long (post-escape) RNA, and on the rate of long RNA synthesis by these productive complexes. From rapid-mixing experiments at 19 °C we determine the large effects of [PPi] on each step of productive initiation. Analysis of these 19 °C kinetics in the context of our previous results for irreversible initiation without added PPi reveals that [PPi] effects result from pyrophosphorolysis and yield composite 2nd order rate constants for each pyrophosphorolysis (back reaction) step from 16-mer to 3-mer. We compare the magnitudes and nucleotide (ultimate, penultimate) dependences of pyrophosphorolysis rate constants with previous observations of PPi effects on elongation steps [38]. From manual-mixing experiments at 37 °C we determine the large effects of [PPi] on the fraction of the OC population that is nonproductive, on the length-distribution of the short RNAs synthesized by nonproductive complexes (NPC) before stalling, and on the abortive synthesis rates of individual short RNAs.

We conclude that effects of [PPi] on initiation result from pyrophosphorolysis and that PPi effects on steps of initiation parallel those observed in elongation, but are even more profound. Based on our results and analysis, we propose that physiological [PPi] affects amounts and rates of productive (and also abortive) initiation in a promoter-specific manner. Because cytoplasmic [PPi] varies with growth conditions and attains low mM concentrations in exponential growth, we deduce that [PPi] is a global regulator of transcription initiation and gene expression.

Results

Time Courses of Transcription Initiation by Productive and Nonproductive Complexes at 37 °C and Physiological Levels of Pyrophosphate (PPi)

The kinetics of transcription initiation by *E. coli* RNAP at the λP_R promoter were investigated over a wide range of [PPi] (0.1 mM – 3 mM) that bracket that observed in exponentially-growing *E. coli* [1-3]. OC were formed between RNAP and a 122 bp linear λP_R promoter (-80 to +42) with a modified ITR (Fig. 2A) to halt transcription at 16-mer when CTP is withheld. For Fig. 2 experiments, these pre-formed OC were manually mixed with NTPs at 37 °C to initiate transcription. To obtain efficient initiation, concentrations of the two initiating NTP (ATP, UTP) were 200 μM, while that of the labelling NTP (GTP) was 10 μM unlabelled, 17.5 nM α-^{32}P-GTP. Use of a low GTP concentration ensures efficient incorporation of the radiolabel and improves detection of transient intermediates in 16-mer RNA synthesis [12, 15]. RNAP escapes from its contacts with the λP_R promoter DNA upon synthesis of 11-mer RNA [14, 43]. Long (post-escape) RNA is defined as the sum of all RNA species ≥ 11-mer (referred to as 11+ RNA) observed on gels like Fig. 2B.

Figure 2B shows representative gel separations of quenched samples taken at the indicated times in initiation without added PPi and with 2 mM added PPi. The gel at left, from an experiment without added PPi, is similar to those obtained for a different set of NTP concentrations (200 μM ATP,GTP; 10 μM UTP, 17.5 nM α-^{32}P-UTP) and analyzed previously [14]. These experiments demonstrate two kinetic phases of initiation and two subpopulations of complexes (productive, nonproductive). In the faster phase, complete in the first 10-20 s without added PPi for these conditions, RNAP in productive complexes escapes from the promoter to synthesize an 11+ RNA, while nonproductive complexes stall after synthesis of a shorter RNA. At 37 ^{o}C the kinetics of this first phase are too fast to determine by manual mixing, but were previously determined in fast-mixing (rapid quench flow, RQF) experiments with a <10 ms deadtime [12]. Manual mixing experiments like those in the left panel of Fig 2B without added PPi provide information about the endpoint amounts of 11+ and shorter RNAs synthesized in this faster phase. At longer times release of some short RNAs from nonproductive complexes (e.g. 7-mer in Fig. 2B left panel) results in re-initiation and abortive RNA synthesis which continues for the duration of this 480 s experiment. In this slower kinetic phase, significant readthrough of the stop at +16 is observed, as in previous studies [14], and is accounted for in calculations of amounts of 11+ RNA.

The right panel of Fig. 2B shows the qualitatively different behavior of both short and long RNA synthesis in the presence of 2 mM PPi. This gel reveals slow kinetics of synthesis of 11+ RNA by productive complexes at 2 mM PPi, occurring over 480 s and accurately determined in these manual mixing experiments. In addition, the fraction of the OC population that synthesizes an 11+ RNA is visibly reduced at 2 mM PPi as

compared to 0 mM PPi, while the fraction of OC synthesizing only short RNA increases at 2 mM PPi.

Much more extensive readthrough of the stop at 16-mer is observed in these 2 mM PPi experiments than at 0 mM added PPi, resulting in synthesis of 31-mer at the next stop point where CTP is required [15]. No significant readthrough of the 31-mer stop (and no run-off 42-mer RNA) is observed. Readthrough bands are included in the calculations of amounts of 11+ RNA for Fig. 2. Other than this, readthrough does not affect the interpretation of these 37 ^{o}C experiments.

Effects of PPi on the Time Required for Long (11+) RNA Synthesis and on the Fraction of Productive OC at 37 ^{o}C

Results of the experiments in Fig. 2B and other 37 °C experiments at [PPi] from 0 – 3 mM are analyzed in Fig. 2C, which plots the amount of 11+ RNA synthesized (normalized to total OC) vs. time at these different [PPi]. Fig. 2C is truncated at 240 s for clarity; Fig. S1 extends these plots to include the 480 s time points. With increasing [PPi], the fraction of OC that synthesize 11+ RNA decreases and the time required for single-round synthesis of 11+ RNA increases greatly. Fig. 2D quantifies these trends.

Fig. 2D indicates that without added [PPi], for the λP_R promoter construct and conditions investigated, about half (40-50%) of OCs are productive, capable of synthesizing 11+ RNA upon NTP addition in the time interval examined (480 s). A similar result was obtained previously at a different NTP condition [14]. As [PPi] is increased to 1 - 2 mM, Fig. 2D shows that the fraction of OCs that behave as productive is reduced to

about 20% of the OC population, while at 3 mM PPi no OC behave as productive and 11+ RNA synthesis is eliminated.

Previous fast-mixing experiments performed in the absence of PPi for the conditions investigated here found that the approach of the amount of 11+ RNA to its plateau value exhibits first order kinetics with a $t_{1/2}$ of ~1.8 s (rate constant $k_{11+} \approx 0.4\ s^{-1}$) after an initial short lag (~0.4 s) [12]. Fig. 2D shows that addition of PPi drastically slows the synthesis of 11+ RNA, allowing $t_{1/2}$ values to be determined by manual mixing. At 1 mM PPi, $t_{1/2}$ of 11+ RNA synthesis is ~25 s, a ~14-fold increase in $t_{1/2}$ as compared to 0 mM added PPi. At 2 mM PPi, $t_{1/2}$ is ~70 s, almost a 40-fold increase as compared to 0 mM added PPi. For comparison, a ~2-fold effect of 1 mM PPi was observed on the rate of *in vitro* elongation relative to 1 μM PPi [39].

Fast-mixing initiation kinetic experiments, performed at 19 °C with a rapid-quench-flow (RQF) mixer and described in the following sections, are needed to determine step-by-step kinetics and mechanism of long RNA synthesis as a function of [PPi] and test the hypothesis that the large effects of PPi are the result of pyrophosphorolysis. These step-by-step kinetics are more accurately determined at 19 °C than at higher temperatures because of the larger number and greater amplitudes of short RNA transients and because rates of abortive synthesis by nonproductive complexes are reduced at this lower temperature (ref 12, 15). On the other hand, 37 ºC is a better choice of temperature to characterize PPi effects on initiation by nonproductive complexes because abortive rates are larger, transients in long RNA synthesis are reduced, and the rate of long RNA synthesis is larger (ref 12) so one obtains a better separation of contributions to initiation from nonproductive vs productive complexes.

Fast Kinetic Analyses: Calculation of Rate Constants for Step-Wise Conversion of 11-mer to 16-mer Without Added PPi at 19 °C

Forward rate constants (2nd order; k_{cat}/K_m analogs) for each nucleotide incorporation step of irreversible initiation by productive OC up to and including 11-mer synthesis were previously determined for the λP_R promoter without added PPi at 19 °C and higher temperature [12, 15]. In our previous research, all RNA lengths beyond the escape point were defined as full-length and their amounts as a function of time were summed to obtain the kinetics of synthesis of full-length (11+) RNA. An extension of that mechanism and step-by-step kinetic analysis to the induced pause after 16-mer synthesis in the absence of added PPi is now needed to accompany the quantitative analysis of stepwise pyrophosphorolysis rates. This extension of the minimal kinetic mechanism is given in Fig. S2; all forward steps are irreversible when no PPi is added. The left panels of SI Fig. S3 show the irreversible kinetic data for 12-mer to 16-mer at 19 °C and 0 mM added PPi. Amounts of 12-mer and 13-mer RNA increase to maxima at around 10 s and then decrease, 14-mer and 15-mer RNA increase to maxima at 15-20 s before decreasing, and 16-mer RNA increases to a stable plateau at times > 50 s because of the slow rate of readthrough (quantified below) for these 19 °C experiments without added PPi. Analysis of these results yields composite 2nd order rate constants k_i^+ for each post-escape step of nucleotide incorporation up to 16-mer (Table S1). Values of k_i^+ are plotted in Fig. 3A, and the good fits to the time courses at the two NTP conditions investigated in that research [15] using these rate constants are shown in Fig. 3B and C. Labelling of

these longer RNAs at multiple positions makes it possible to quantify their small mole amounts and determine rate constants accurately.

Values of rate constants k_{12}^{+} and k_{13}^{+} for the first two post-escape steps (synthesis of 12-mer, 13-mer) are progressively larger than k_{11}^{+} for the 11-mer escape step, but k_{14}^{+} is smaller

than k_{13}^{+} (Fig 3A; Table S1). The progression to larger k_{i}^{+} values resumes for 15- and 16-mer rate constants (k_{15}^{+} and k_{16}^{+}) (Table S1). While these post-escape rate constants (ranging from 0.08 μM^{-1} s^{-1} to 0.2 μM^{-1} s^{-1}) are at the high end of the range of pre-escape rate constants (Table S1), all are less than previously reported rate constants for initial dinucleotide synthesis (without translocation [12]; ~ 0.6 μM^{-1} s^{-1} at 19 and 25 °C) and for the combined NTP binding and catalytic steps of elongation (≥ 0.6 μM^{-1} s^{-1} at 24 °C [35, 36])

Fast Kinetic Analyses: Profound Effects of PPi on the Step-by-Step Kinetics of Initiation by Productive λP_R OC at 19 °C

Effects of PPi on the step-by-step kinetics of transcription initiation by productive complexes were investigated by fast-mixing (RQF) experiments at 19 °C. Fig. 4A shows representative PAGE gels comparing time courses of initiation of RNA synthesis at 0 mM and 2 mM added PPi, and a representative gel of an RQF experiment at 1 mM PPi is shown in SI Fig. S4. Phosphorimager analysis of these gels provides quantitative single-nucleotide resolution of ^{32}P-labelled RNA products as a function of time after NTP addition to preformed λP_R open complexes (which, at 19 °C are predominantly I_3, the initiation complex) [12]. Long RNA synthesis in these experiments is single-round because heparin is added with the NTPs to bind any RNAP that dissociates from the DNA.

Visual comparison of the 0 mM and 2 mM PPi gels in Fig. 4A reveals that the kinetics of long RNA synthesis by productive complexes are almost an order of magnitude slower at 2 mM added PPi than at 0 mM added PPi. Transients in short RNA intermediates on the pathway to long RNA synthesis by productive complexes are shifted to longer times and broadened at 2 mM added PPi. Additionally, the amount of short abortive products is much greater at 2 mM added PPi than at 0 mM added PPi. Quantitative comparisons are made in subsequent figures. More readthrough of the 16-mer stop, resulting in 31-mer synthesis, is observed at 2 mM PPi than without added PPi (Fig 4A; see also Fig 2B). A first order readthrough step with rate constant k_{rt} is included in the overall initiation mechanism in Fig S2, and values of k_{rt} are obtained at each [PPi].

Fig. 4B quantifies the kinetic behavior of two pre-escape RNAs (5-mer, 9-mer) and a post-escape RNA (12-mer) as a function of [PPi] (0, 1, 2 mM) at 19 °C. Total amounts of these RNA lengths, including initiation from both productive and nonproductive complexes, are plotted as a function of time. Results for 5-mer and 9-mer at 0 mM added PPi were analyzed previously [12, 15]. Results for 12-mer at 0 mM added PPi are from experiments analyzed in Fig. 3. Significant transients are observed at these three lengths, in part because the next base added is G and the low concentration of GTP in the reaction mixture reduces the rate of extension to 6-mer, 10-mer, and 13-mer respectively. With increasing [PPi], rates of formation and of subsequent decay of all three of these intermediate RNA species decrease because each step of RNA synthesis becomes more reversible. This is expected from the observed behavior of long RNA with increasing [PPi] in both the 37 °C experiments in Fig. 2 and in these 19 °C experiments in Fig. 4A.

In general, after the increase and decrease in amount of a short RNA ($\leq$ 10-mer for λP_R) in the transient phase, a non-zero plateau or slow increase is observed, where the plateau indicates the amount of RNA synthesized in the first round by NPC that stall at that length, and the slow increase is from re-initiation after slow release of this first short RNA by NPC (abortive initiation) [14, 15]. Without added PPi, Fig. 4B shows that at 0 mM added PPi there is little synthesis of 5-mer and 9-mer RNA by NPC, either initially or abortively, so the plots return to zero after the transient. This is because λP_R NPC at this this low-GTP, 19 °C condition primarily synthesize 3-mer, with smaller amounts of 4-, 7-, 8- and 10-mer, and abortive rates are only significant for 3-mer and 4-mer (Figs. S5-S6; see below). After the transient phase the amount of 12-mer (a post-escape RNA) decays to zero as expected.

As [PPi] is increased to 1mM and 2 mM. Fig. 4B shows that post-transient amounts of 5-mer, 9-mer, and 12-mer all increase greatly. Figs. S5-6 show similarly large increases for 3-mer and 4-mer, but not for most other RNA lengths. Abortive rates remain small for 5-mer and 9-mer at 1 mM and 2 mM PPi. For 12-mer, which is post-escape and therefore synthesized only by productive complexes, the return of the transient peak to zero at 0 mM added PPi indicates that synthesis of 13-mer is irreversible for this condition. Development of a significant 12-mer plateau at 1 mM and 2 mM PPi demonstrates unequivocally that 13-mer synthesis is reversible at these PPi concentrations, and that the bottleneck at 16-mer causes long-time stable plateaus (on the time scale of these experiments) in the amounts of preceding long RNA intermediates (11-, 13-, 14-, and 15-mer in Fig. S5-6).

Fig. S7 plots time courses of long (11+) RNA synthesis at 19 °C from the above RQF measurements at 0 mM, 1 mM and 2 mM PPi. The fraction of OC that are productive and synthesize an 11+ RNA is reduced by these PPi concentrations and the half time of 11+ RNA synthesis is increased, as observed at 37 °C (Fig. 2 C, D). The effect of PPi on the half time of 11+ RNA synthesis is not as large at 19 C as at 37 C. At 19 °C $t_{1/2}$ increases from approximately 7 s at 0 mM PPi to 20 s at 1 mM PPi and 34 s at 2 mM PPi, while the fraction of OC that are productive is halved at these mM PPi concentrations (Fig. S7).

*Rate Constants of Pyrophosphorolysis Steps of Initiation and Escape Span a 1000-fold Range*Second order rate constants k_i^- for the overall process of PPi uptake and pyrophosphorolysis are obtained from the analysis of the kinetic data (including that of Fig. 4A) at 1 and 2 mM PPi according to the mechanism of Fig S2, which includes the first order kinetics of readthrough of the 16-mer stop to form 31-mer with rate constant k_{rt}. These reverse rate constants (k_i^-) are plotted together with forward rate constants (k_i^+) on a linear scale in Fig. 5A, and k_i^- values are plotted on a logarithmic scale in Fig 5B to illustrate the wide range (>10^3) in values of k_i^- for different steps of initiation. All rate constants (k_i^+, k_i^-, k_{rt}) are given in Table S1. Observed time courses of all RNA intermediates in initiation by productive complexes (and of the readthrough 31-mer) are compared with predictions from the fitted rate constants at each [PPi] (0, 1 and 2 mM) in Fig. 5C.

We assume in these fits that the enzymatic Michaelis constant for PPi ($K_{m,\ PPi}$) is larger than 2 mM (i.e. a small PPi binding constant $1/K_{m,\ PPi} < 500\ M^{-1}$) so that a single composite rate constant k_i^- (the analog of k_{cat}/K_m) is sufficient to describe the kinetics up

to 2 mM PPi. This assumption is justified by the expectation that PPi binding should be weak and by the quality of the fits obtained. In these fits, newly-determined forward rate constants for post-escape steps and previously-determined forward rate constants for pre-escape [12] steps are fixed at their values without added PPi in order to test whether pyrophosphorolysis by itself, without other PPi effects, can explain all the observed changes in the kinetics of short and long RNA synthesis by productive complexes at 1 and 2 mM PPi.

Fig. 5B shows that pyrophosphorolysis k_i^- values span a very wide range (over 3 orders of magnitude). Steps 4, 7, and 15 have relatively large k_i^- values (0.01 – 0.1 $\mu M^{-1}s^{-1}$). Steps 6, 9, 10, 12, 13, and 16 are in the intermediate range of k_i^- values (0.0005 – 0.01 $\mu M^{-1}s^{-1}$). Steps 3, 5, 8, 11, and 14 have very small k_i^- values (<0.0005 $\mu M^{-1}s^{-1}$) (Fig. 5B). Values of k_i^- for this last group of steps are of similar magnitude to pyrophosphorolysis rate constants in elongation, which are on the order of 10^{-5} μM^{-1} s^{-1} or less (Table S1, calculated from published results [38]. Values of k_i^- for nine of the fourteen initiation steps quantified are much larger than for elongation. As discussed below, k_i^- values of these initiation and post-escape steps correlate well with the identity of the 3' RNA base, as previously reported for elongation steps [36].

Effects of PPi on Synthesis of Short RNAs by Nonproductive Complexes

For the reasons stated above, we chose 37 °C manual-mixing experiments [14] like those of Fig. 2 to investigate effects of [PPi] on synthesis of short RNA by nonproductive complexes (NPC). Results in Fig. 6 reveals that PPi at physiological concentrations has very large effects on the fraction of complexes that are nonproductive,

on the initial amount of short RNA of each length synthesized by NPC and on the rate of abortive synthesis of these RNA lengths. NPC initiate a first round of RNA synthesis but are incapable of escaping the promoter to make a 11+ RNA and stall after synthesis of a short, pre-escape length (i.e. 3-mer to 10-mer) RNA. The time required for the initial round of short RNA synthesis by NPC at 37 °C is not accurately known, but is probably similar to that for long (11+) RNA synthesis by productive complexes ($t_{1/2} \approx 1.8$ s without added PPi; Fig. 2D). Stalled NPC slowly release their RNA and reinitiate (abortive initiation). At 37 °C the nonproductive fraction of the OC population that stalls at each RNA length is quantified by extrapolating the linear abortive-synthesis kinetic data to time t = 0 [14]. The bar graph of Fig. 6A quantifies the effect of PPi on the fraction of λP_R initiating complexes detected in our assays that are nonproductive at 37 °C for the NTP concentrations investigated here and compares with results obtained at 19 °C from experiments like those in Fig 4.

Figure 6B shows the changes in the length distribution of the initial short RNAs synthesized by NPC before stalling as [PPi] is increased from 0 to 3 mM. (Values are listed in Table S2.) In the absence of PPi, this length distribution is broad, with a peak at 4- and 5-mer but with significant populations of all RNA lengths up to 10-mer. Increasing [PPi] increases the population fraction of NPC that initially synthesize a short RNA (3-, 4-, and 5-mer) before stalling, and greatly reduces the fraction that make a longer RNA (6-mer to 10-mer). Populations of 8-mer and 10-mer disappear at 2 mM PPi and no RNA longer than 5-mer is synthesized by NPC at 3 mM PPi (Fig. 6B). These trends in the length distribution of short RNAs at 37 °C can be explained qualitatively by reference to the 19 °C rate constants in Fig. 5 and Table S1. At 1 mM and 2 mM PPi, for the NTP

concentrations investigated here, step 4, 6, and 7 are predicted to be thermodynamically reversible and steps 9 and 10 are predicted to be somewhat reversible. At 3 mM PPi, steps 4, 6, 7 and 10 are all predicted to be reversible, and steps 5, 9, and 11 are predicted to be somewhat reversible. With so many reversible steps beyond 5-mer, it is not surprising that most NPC synthesize only 3-mer, 4-mer or 5-mer RNA at 3 mM PPi (Fig. 6B, C).

Fig. 6C summarizes abortive initiation rates at 37 ^{o}C, determined from manual-mixing experiments extending to 480 s as a function of [PPi] up to 3 mM. These rates are for one abortive cycle, including release of the initial short RNA, regeneration of the original OC without dissociation of RNAP from the promoter, and re-synthesis of another short RNA. (Values are listed in Table S3.) Abortive rates at 19 ^{o}C from rapid (RQF) mixing experiments extending to 150 s are smaller and less well determined, but trends with [PPi] appear similar to those at 37 ^{o}C in Fig 6 C. Amounts of RNA of a given length observed as a function of time are normalized by the total amount of NPC making RNA of any short length (3-mer to 10-mer), and expressed as RNA per total NPC per second.

In the absence of added PPi, each cycle of abortive initiation is a much slower process than productive initiation, which occurs on a time scale of a few seconds for the conditions investigated (Fig. 2D; [12]). Averaged for all short RNAs detected (3-mer to 10-mer) at the low U condition investigated previously, the time required for abortive synthesis by a nonproductive complex at 37 C is ~80 s [14]. The results in Fig. 6C for the high U condition investigated here are very similar (~80 s).

Fig. 6C shows that, without added PPi, the abortive synthesis rate is largest for 7-mer RNA. The rank order of abortive synthesis rates is 7-mer > 4-mer > 5-mer> 6-mer $\approx$

3-mer > 9-mer ≈ 8-mer > 10-mer. This rank order, obtained for a “high UTP” condition (200 μM ATP, UTP; 10 μM GTP), differs from the order obtained previously at a different NTP condition (“low UTP”: 200 μM ATP, GTP; 10 μM UTP): 3-mer > 7-mer > 6-mer ≈ 4-mer > 5-mer > 8-mer > 10-mer > 9-mer), and shows that multiple factors (including but not limited to NTP concentration and RNA length) determine the rank order of abortive rates. Addition of 1 mM PPi has modest effects on rates of abortive synthesis of most RNAs (Fig. 6C), except for 7-mer (large reduction) and 3-mer (moderate increase). The average time required for a round of abortive synthesis is ~130 s as compared to ~80 s for 0 mM added PPi.

Much larger effects on rates of appearance and amounts of short RNAs per NPC are observed at higher [PPi]. Fig. 6C shows that abortive rates of longer RNAs (6-mer to 10-mer) are greatly reduced at 2 mM PPi and become too small to measure at 3 mM PPi. Concurrently, abortive rates of 3-mer and 4-mer increase greatly. At least in part because of this shift to shorter RNAs, which are more easily released and where the conformational changes to re-form the initiation complex are presumably smaller, the time required for a round of abortive synthesis, averaged for all RNA lengths, is greatly reduced from 130 s at 1 mM PPi to ~30 s at 2 mM PPi and ~20 s at 3 mM PPi. An alternative normalization of abortive rates for each RNA length (Table S4), using the initial amount of the NPC of that length instead of using the total amount of NPC as in Fig. 6C, shows similar trends for each RNA length with increasing [PPi].

A minority of OC (variable between experiments but typically 20-30%) are not detected as either productive or nonproductive in both low-UTP [14] and low-GTP initiation assays at all PPi concentrations investigated. These OC may fail to initiate at all

or may stall during or after synthesis of the initial pppApU dinucleotide. Any pppApU synthesized in the low GTP assays reported here is unlabeled, and even when labeled in experiments at low UTP is difficult to detect [12, 15]. The fraction of undetected complexes does not exhibit a significant trend with PPi concentration.

Discussion

[PPi] Regulates Many Aspects of Initiation by Productive and Nonproductive Complexes

With increasing [PPi], all steps of initiation become more reversible, reducing the net forward rate of RNA synthesis per RNAP-promoter complex engaged in productive or nonproductive (abortive) initiation. From the rate constants reported here, the effects of any choice of PPi and NTP concentrations (below their K_m) on the rate of productive initiation at the λP_R promoter and ITR investigated here are readily calculated. Additionally, PPi facilitates stalling of RNAP prior to promoter escape, converting what were productive complexes without added PPi to nonproductive complexes that are incapable of promoter escape and the transition to elongation. Hence an increase in [PPi] in the physiological range reduces the total amount as well as the rate of long RNA synthesis, while increasing amounts and rates of short (abortive) RNA synthesis. An increase in PPi also shifts the length distribution of the short RNA made initially and in following abortive cycles by nonproductive complexes, reducing the fraction of nonproductive complexes that make 6-mer or longer RNA and increasing the fraction that make 5-mer or shorter RNAs.

In addition to PPi concentration, which is highly variable *in vivo* (as discussed below), factors investigated in this research and found to be important for regulation of initiation by PPi include RNA sequence, because pyrimidine steps are reversed at faster

rates than purine steps, and translocation stress, which determines the bias in initiation for the pre-translocated state as a result of translocation stress including the coupling of translocation to disruption of interactions between RNAP and the upstream regions of promoter DNA. The bias toward the post-translocated state in elongation reduces the pyrophosphorolysis rate because only pre-translocated complexes can undergo pyrophosphorolysis. Unlike NTP binding, which occurs only to the post-translocated state, PPi appears to bind with similar affinity to pre- and post-translocated states so the extent of translocation is not a function of PPi concentration [30, 38]. Hence PPi binding does not preferentially stabilize $\mathbf{IC}_{\mathbf{i+1}}^{\mathbf{pre-tr}}$ relative to $\mathbf{IC}_{\mathbf{i+1}}^{\mathbf{post-tr}}$.

Both productive and abortive initiation from other promoters should be affected by [PPi] in similar ways to those described here for the λP_R promoter. These [PPi] effects will certainly depend on the ITR sequence and probably also on the promoter sequence that determines the stability of the open complex. Translocation stresses in initiation are presumably larger for stable OC like λP_R than for unstable OC like those formed at ribosomal promoters [14], and this difference is predicted to affect both the escape point [14] and the bias in each step for the pretranslocated state [12, 15]. Consideration of these factors leads to the prediction that initiation from promoters with unstable OC should be less sensitive to PPi than initiation from promoters with stable OC like λP_R.

Pyrophosphorolysis Rates in Initiation Vary Greatly with RNA 3'-end Sequence

Second-order pyrophosphorolysis rate constants k_i^- of initiation and post-escape steps of long RNA synthesis span a wide range from < 10^2 M^{-1} s^{-1} to 6 x 10^4 M^{-1} s^{-1} (Fig. 5; Table S1). These rate constants correlate well with the identity of the 3' RNA base. All

three large and one moderate pyrophosphorolysis rate constants are for steps where the RNA initially ends in 3'U and UTP is the product. The U-excision steps with large pyrophosphorolysis rate constants are distributed between early initiation (step 4), late initiation (step 7) and post-escape (step 15). All five steps with negligibly-small pyrophosphorolysis rate constants and five of the steps with moderate pyrophosphorolysis rate constants are steps where the 3' end base is a purine and ATP or GTP is the product. These steps also are distributed throughout the ITR: steps 6, 9, 10, 12, 13, and 16 have moderate k_i^- values and steps 3, 5, 8, 11, and 14 have relatively small k_i^- values. Aside from the strong correlation between rate constant and the 3' end sequence of the RNA, we have found no other correlation of rate constant with position in the ITR. We deduce that the lack of a detectable dependence of k_i^- values for the same 3' RNA dinucleotide sequence on ITR position results from the strong pre-translocation bias present in all steps of initiation, as discussed quantitatively below.

In addition to the large differences in pyrophosphorolysis rate constants for pyrimidine vs. purine bases, second nucleotide effects are also present. The smallest rate constant for a 3'-U pyrophosphorolysis step and two of the smallest rate constants for a 3'-A pyrophosphorolysis step occur where the next base in the pyrophosphorolysis direction is U. All the larger rate constants for purine pyrophosphorolysis steps occur where the next base in the pyrophosphorolysis direction is another purine, regardless of the position within in ITR.

The dependence of pyrophosphorolysis rate constants on the identity of the 3' RNA nucleotides matches what has been reported for elongation. In elongation, transcripts ending in U are the most PPi sensitive, followed by C, A, and G [30, 38]. Furthermore, the

penultimate 3’ nucleotide has also been shown to influence PPi sensitivity, with roughly the opposite effect as the final 3’ nucleotide (e.g. a penultimate U making the RNA strand overall less PPi sensitive). These penultimate-nucleotide effects are also present in pyrophosphorolysis rate constants of initiation steps. The steps that are predicted to undergo slow but significant pyrophosphorolysis, excluding step 16, all have the 3’ sequence AG, AA, or GA. All steps predicted to undergo negligible pyrophosphorolysis have the 3’ sequence UG, UA, or GG. Step 16, the only step with a 3’ U that is not predicted to undergo fast pyrophosphorolysis, is the only step with the sequence UU. These sequence effects closely correspond to those reported for elongation.

This correlation of pyrophosphorolysis rate constants with the identity of the 3’ RNA base contrasts with the situation for RNA synthesis (forward) rate constants (Fig. 5A and Table S1), which as reported previously [12, 15] show no significant correlation with the identity of the nucleotide incorporated. Forward rate constants for pre-escape steps instead exhibit a pattern with three cycles of large and small values (3-mer to 5-mer, 6-mer to 9-mer, 10-mer to 11-mer). We previously interpreted these small forward rate constants in terms of difficulties in translocation resulting from coupling of translocation to sequential disruption of RNAP-promoter contacts with the discriminator strands, with the -10 region and with the -35 region [12, 15]. We also found that the corresponding step-by-step forward rate constants for initiation by T7 RNAP [44] exhibit a clear pattern with ITR position and proposed a structural interpretation in terms of disruption of RNAP-promoter contacts [15].

The translocation step is in rapid equilibrium so the contribution of translocation to each k_i^+ value is determined by the translocation equilibrium constant K_i^{tr} , defined as the

ratio of concentrations of post- and pre-translocated states [12, 15]. Each k_i^+ value contains a factor $K_i^{tr}/(1+K_i^{tr})$, interpreted as the fraction of complexes in the post-translocated state which binds the NTP. Likewise each k_i^- value contains the factor $1/(1+K_i^{tr})$, interpreted as the fraction of complexes in the pre-translocated state, to which PPi must bind before pyrophosphorolysis. We previously deduced that $K_i^{tr} << 1$ for all initiation steps, varying by more than 10-fold for different steps. For this situation, applicable in initiation but not in elongation, k_i^+ is proportional to K_i^{tr} while k_i^- is independent of K_i^{tr}, so translocation stresses in initiation affect only k_i^+, not k_i^-.

Pyrophosphorolysis Rates in Initiation Greatly Exceed those for the Same Pair of 3' Nucleotides in Elongation, Amplifying Differences in Rate for Different Sequences

Pyrophosphorolysis rate constants (Table S1) for most if not all initiation steps greatly exceed those estimated from the published *E. coli* elongation results [38]. A numerical comparision can be made for GU and UG 3' dinucleotide sequences. For excision of 3' U when the adjacent base is G, the ratio of pyrophosphorolysis rate constants for initiation and elongation is $\geq 10^4$, while for excision of G with a neighboring U it is $\geq 10^2$. For these GU vs UG 3' dinucleotide sequences, pyrophosphorolysis k_i^- values differ by 6-fold in elongation but by $>10^2$ fold in initiation (Table S1). Hence initiation appears to amplify the effects of the 3' nucleotides observed in elongation. In elongation, the difference in pyrophosphorolysis rates for GU vs UG is thought to arise from differences in interactions of 3' U and G with the folded trigger loop (TL) in the pre-translocated state, with 3' U – TL interactions being more favorable than 3' G – TL interactions. If and how this preference might be amplified in initiation is unknown.

[PPi] Regulates the Amount and Rate of Short (Abortive) RNA Synthesis by Nonproductive Complexes (NPC)

Evidence has been presented for two distinct sub-classes of initiation complexes: productive and nonproductive [14, 42, 45]. Productive complexes initiate without aborting, carrying out processive RNA synthesis which culminates in escape of RNAP from the promoter and conversion of the initiation complex to an elongation complex. NPC, on the other hand, stall before escape, presumably as the result of a translocation defect, slowly release the short RNA they synthesized and start over, making another short RNA in an abortive cycle. Abortive transcripts may have regulatory roles *in* vivo. The shortest abortive RNA have been reported to prime initiation, in some cases changing the transcription start site (TSS) [46, 47]. The molecular differences between productive and nonproductive initiation complexes remain unclear.

Addition of PPi makes complexes that were productive at 0 mM added PPi behave as nonproductive, stalling after synthesis of an initial short RNA, slowly releasing that RNA and re-forming the initiation complex to carry out another round of short RNA synthesis (abortive initiation). NPC presumably are defective in translocation [15]. [PPi] acts like an additional translocation defect by increasing the rate of the back reaction, reducing translocation especially at 3' dinucleotide sequences like GU in the ITR investigated here, and thereby converting productive complexes to nonproductive. *Predicted Consequences of Large Changes in* E. coli *Cytoplasmic [PPi] with Growth Conditions on Gene Expression*

The [PPi] in the *E. coli* cytoplasm increases greatly from stationary phase (low μM) to exponential growth (~1 mM), and also varies with growth conditions [1-3]. In stationary

phase where the gene expression rate is very small, eliminating most production of PPi from nucleic acid synthesis and tRNA aminoacylation, $[PPi]_{stationary}$ is in the low μM range [48], comparable to K_m of the *E. coli* pyrophosphatase [48]. In exponential growth, on the other hand, $[PPi]_{exponential}$ is in the low mM range [2]. This thousand-fold increase in [PPi] appears to be the direct result of gene expression and protein synthesis, which are capable of making PPi at a sufficiently rapid rate to overwhelm the ability of the constitutive pyrophosphatase to degrade it, resulting in a cytoplasmic [PPi] that increases greatly in the transition from stationary phase to exponential growth (i.e. with increasing growth rate).

The predicted large effects of mM PPi concentrations in vivo on tRNA aminoacylation and therefore on protein synthesis led to the proposal that much of this anionic PPi is bound in the cytoplasm [49]. A possible analogy for this would be Mg^{2+}, where the free concentration is much less than the total concentration because of strong binding of Mg^{2+} to polyanionic ribosomal RNA phosphates. No comparable polycation exists to bind anionic PPi, and if it were bound to individual anion binding sites of proteins these would have to be present on a significant fraction of all proteins, because the total cytoplasmic protein concentration is approximately 5 mM. No evidence exists for such widespread PPi binding sites on *E. coli* proteins, so it is likely that the total and free [PPi] are similar and are in the mM range in exponential growth, where PPi effects on initiation are large.

Conclusions

In this research we quantified the very large effects of changes in [PPi] in the physiological range (up to low mM) on productive and abortive transcription initiation by

E. coli RNAP at the λP_R promoter. This information was not previously available for any RNAP or promoter DNA. We found that, as in elongation, uridine incorporation and some other initiation steps become significantly reversible at physiological PPi and NTP concentrations. We also found that both the fraction of RNAP-promoter complexes that productively initiate and the rate of RNA synthesis per productive complex decrease with increasing [PPi]. Concomitantly the fraction of complexes that abortively initiate increases and the abortive product distribution shifts shorter RNAs. Pyrophosphorolysis rates for some initiation complexes are orders of magnitude larger than for removal of the same nucleotide from elongation complexes because of the strong bias toward the pre-translocated state in initiation, and exhibit even stronger dependences on nucleotide identity (pyrimidine >> purine). Initiation from promoters with unstable open complexes, which are predicted to build up less translocation stress and therefore exhibit less of a bias to the pre-translocated state and which may transition from initiation to elongation at a shorter RNA-DNA hybrid length [14], should be less affected by [PPi] and should behave more like elongation complexes in this regard. Given that cytoplasmic [PPi] is much higher in exponential-phase than stationary-phase *E. coli* and varies with growth conditions, these large [PPi] effects must be physiologically-relevant.

Methods

Reagents, Buffers, and Gels

Reagents for buffers and stock solutions were the highest available grade and were used as received. All solutions were prepared using 18 MΩ deionized water from a Barnstead EPure system. NTPs and dNTPs (Boston Bioproducts, Thermo Fisher, New England Biolabs) used in transcription assays and PCR reactions were 99% pure and

used as received. Enzymes for PCR reactions were purchased from NEB and used according to the manufacturer's protocols.

Storage buffer (SB) for core RNAP, σ^{70} and RNAP holoenzyme is 50% v/v glycerol, 0.01 M Tris (pH 8.0), 0.1 M NaCl, 0.1 mM EDTA, and 0.1 mM DTT, 0.05 mg/ml BSA. Transcription buffer (TB) is 40 mM Tris, 5 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, and 0.05 mg/mL BSA, adjusted to pH 8.0 at the experimental temperature (19 °C, 37 °C).

5X initiation solution (IS) for manually-mixed transcription assays at 37 °C assayed with α-^{32}P-GTP is 1 mM ATP, 1 mM UTP, 50 μM unlabeled GTP, 87.5 nM α-^{32}P-GTP, 0.25 mg/ml heparin in TB, and from 0 to 15 mM PPi (five times the final [PPi] concentration). 2X initiation solution (IS) for rapid quench flow (RQF) transcription assays at 19 °C assayed with α-^{32}P-GTP is 400 μM ATP, 400 μM UTP, 20 μM unlabeled GTP, 35 nM α-^{32}P-GTP, 0.1 mg/ml heparin in TB and from 0 to 4 mM PPi (two times the final [PPi]).

Quench Solution (QS) for manual mixing and RQF transcription assays is 8 M urea and 15 mM EDTA in TB. QS with added dyes (QSD) for polyacrylamide gel electrophoresis (PAGE) has 0.05% xylene cyanol and 0.05 bromphenol blue in QS. TBE buffer for PAGE is 90 mM Tris-Borate (pH 8.3) and 2 mM Na_2EDTA. All transcription gels are 20% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics).

RNA Polymerase and λP_R Promoter DNA

RNAP core enzyme ($\alpha_2\beta\beta'\omega$) is overexpressed from pVS10 plasmid and purified via Ni affinity chromatography [15]. The σ^{70} subunit is overexpressed from pIA586 and

purified using Ni affinity chromatography [15]. Holoenzyme is reconstituted using a 1:2 ratio of core to σ^{70}. Filter binding activity assays [14] performed on preparations of RNAP holoenzyme used here show that 50% ± 10% of RNAP molecules form a stable open complex with the λP_R promoter at 37 °C and 0 mM added PPi. All RNAP concentrations reported here refer to this active fraction. Open complexes of RNAP and λP_R promoter DNA are formed as previously described [14, 15] by incubating a 2:1 mole ratio of active RNAP and promoter DNA for 1 hour at the temperature of the experiment. For fast-mixing experiments the initial OC concentration was 100 nM; for manual mixing experiments the initial OC concentration was 12.5 nM. Sequences of PCR primers and λP_R promoter DNA fragments used to prepare the 122 bp promoter DNA fragment used in all kinetic studies reported here and previously are given in Table S5.

Initiation Kinetics Assays

Manually-mixed initiation kinetics assays are performed at 37 °C and analyzed as described previously [15]. Assays are designed to obtain single-round RNA synthesis by productive complexes. Briefly, 5X initiation solution (IS) is manually mixed 1:4 with a solution of pre-formed OC at time zero to obtain a final OC concentration of 10 nM. Reactions are quenched with QS and RNA is separated via PAGE.

Rapidly quenched initiation kinetics assays are performed at 19 °C and analyzed as described previously [14]. Assays are designed to obtain single-round RNA synthesis by productive complexes. Briefly, 2X initiation solution (IS) is mixed 1:1 with a solution of pre-formed OC at time zero using a KinTek Corp Rapid Quench Flow (RQF) apparatus to obtain a final OC concentration of 50 nM. Reactions are quenched with QS and RNA is separated via PAGE.

Transcription gels are transferred to a phosphorimaging cassette and analyzed using a Typhoon 9000 phosphorimager (18 h exposure). Peak area is converted to moles of observed product as described previously [12]. To obtain good labelling efficiency without the use of high concentrations of labelled NTP that cause background problems and limit the ability to detect small RNAs on the gels, we use a low concentration of α-^{32}P-GTP (17.5 nM) and a reduced concentration (10 μM) of the corresponding unlabeled NTP. The probability of incorporation of the radiolabeled NTP at each position is determined by the ITR sequence, and was accounted for using incorporation probabilities [12]. Quantitative results reported here are from the average of three RQF or manually mixed experiments at each set of NTP concentrations investigated, and the uncertainties reported are one standard deviation from the mean.

Quantifying Synthesis of Short RNAs by Nonproductive Complexes in 37 ºC Manual Mixing Experiments

Nonproductive complexes (NPC) are defined as those OC that stall after synthesizing a short (pre-escape length) RNA in the initial phase of the kinetics [12, 14, 15]. At 37 ºC many of the NPC release their RNA and reinitiate to begin a cycle of abortive synthesis. As previously described [14], amounts of NPC that stall at each RNA length after synthesis of an initial short RNA are determined in manual-mixing experiments from the intercepts of plots of the linear increase in the amount of that short RNA vs time from abortive synthesis in the time range after long RNA synthesis by productive complexes is complete.

Quantifying Synthesis of Short RNAs by Nonproductive Complexes in 19 ºC Fast Mixing Experiments

Without added PPi, at 19 °C, the kinetics of synthesis of the first short RNA of each length by NPC are well-described as a first order (single-exponential) approach to a plateau value [12, 15]. These fitted amounts were subtracted from the total amount of RNA of a given length present at each time to determine transient amounts of that RNA length in initiation by productive complexes [15]. Addition of PPi broadens transient RNA peaks from initiation by productive complexes and reduces RNA synthesis rates, introducing ambiguity in the above approach. As an alternative, the full time course of initial and subsequent rounds of RNA synthesis by NPC at 19 °C was modeled in Kintek Explorer [50] as reiterative synthesis of each RNA length by a subpopulation of NPC, according to the simple mechanism $NPC_i \rightarrow NPC_iRNA_i \rightarrow NPC_i + RNA_i$ where the first step is [PPi]-dependent RNA synthesis and the second, rate-determining step is RNA release and restoration of the initiating NPC, assumed to be [PPi]-independent. RNA synthesis and release rate constants were adjusted manually to best replicate the observed NPC behavior. Results of this analysis are shown as red curves in SI Figs. S5-S6. The breakover point between the initial exponential phase and the subsequent linear abortive phase was used to determine the initial amount of RNA synthesized by NPC and hence the amount of NPC synthesizing that RNA length for use in Fig. 6A.

Determination of Rate Constants for Forward (post-escape) Steps of RNA Synthesis

To determine forward rate constants k_i^+ for synthesis of 12-mer to 16-mer for Fig. 3, previously published 0 mM PPi RQF experiments [15] at final NTP concentrations of 200 μM ATP and GTP, 10 μM UTP, and 17.5 nM α-^{32}P-UTP were analyzed together with the data presented here to extend the transient analysis to 16-mer, instead of 11-mer. In conjunction with the 0 mM PPi experiments shown here, the forward rate constants for

steps 11-16 were determined by fitting the transient RNA produced in the two NTP concentrations studied to the mechanism in Fig. S2, an extension of the mechanism reported reported previously [12, 15]. Because these are post-escape species, no correction for synthesis by NPC is required for this analysis.

Determination of Rate Constants for Individual Steps of Pyrophosphorolysis (for 3-mer to 16-mer)

Second order pyrophosphorolysis rate constants at each [PPi] were determined at 19 °C using Kintek Explorer [50] to fit all lengths of RNA transients in 16-mer RNA synthesis to the mechanism in SI Fig. S2. These transients (shown in SI Figs. S3 and S5-S6) were determined by subtracting the predicted amounts of RNA synthesized by NPC from the total amount of RNA at each time. Forward rate constants, determined at 0 mM PPi as previously described [15] were held constant in these fits.

Acknowledgements
We thank the reviewers for their many helpful comments, including the inference that initiation from promoters with unstable open complexes would be less affected by PPi than λP_R. DP was supported by NIH Biotechnology Traineeship NIH 5 T32 GM008349 and KH by NIH NRSA postdoctoral fellowship NIH GM 122303. We gratefully acknowledge the support for this research from the above mentioned fellowships, UW-Madison, and NIH GM R35-118100 (MTR), as well as experimental contributions from Olivia Hamzelou.

Figures

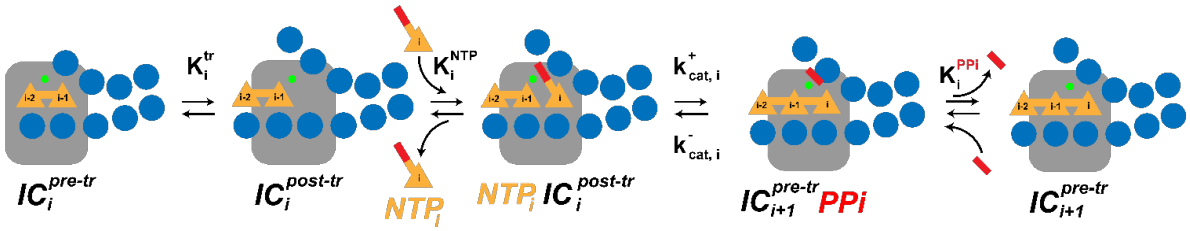


Figure 1: Mechanism of nucleotide addition and its reversal (pyrophosphorolysis) in initiation. In this schematic, DNA bases are represented by blue circles, RNA bases as orange triangles, PPi as a red rectangle, the incoming nucleotide triphosphate (NTP) by an orange triangle, and the active site Mg^{2+} as a green circle. The four substeps are: 1) Reversible translocation of the RNA-DNA hybrid relative to the active site. The pre-translocated state of the initiation complex (IC_i^{pre-tr}) in initiation step i is converted to the post translocated state ($IC_i^{post-tr}$) with equilibrium constant $\mathrm{K_i^{tr}}$, forming the binding site for the i-th NTP. 2) Reversible binding of the i-th NTP to the active site (grey) with equilibrium constant $\mathrm{K_i^{NTP}}$. 3) Catalysis of phosphodiester bond formation between the i-th NTP and the 3' end of the RNA with rate constant $\mathrm{k_{cat,i}^{+}}$, with PPi (red) as byproduct still bound in the RNAP active site. 4) Release of PPi. Addition of PPi makes the PPi binding step (equilibrium constant $\mathrm{K_i^{PPi}}$) significant and makes the catalyzed reaction reversible (pyrophosphorolysis), re-forming the i-th NTP with rate constant $\mathrm{k_{cat,i}^{-}}$.

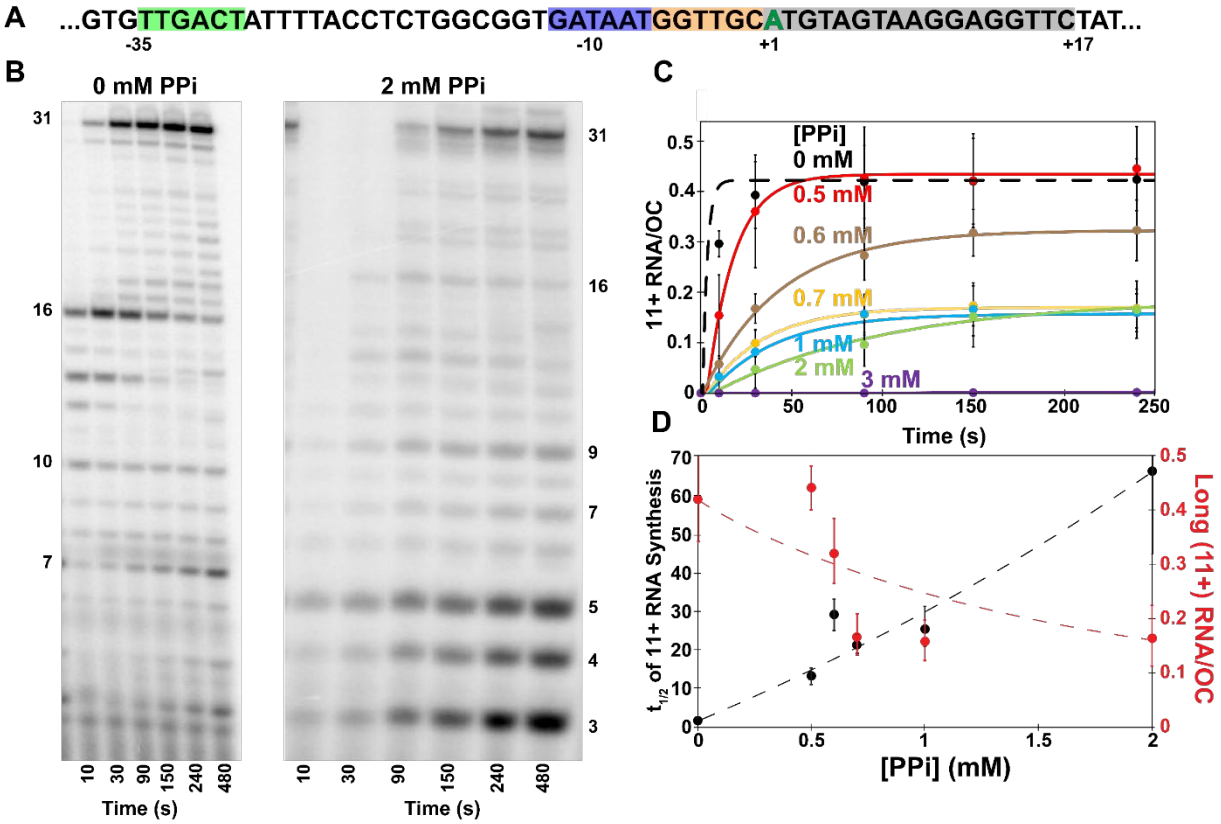


Figure 2: Effects of PPi on Synthesis of Long RNA at 37 °C. Panel A) The nontemplate-strand sequence of the core region of the modified λP_R promoter (in a 122 bp (-80 to -42) promoter fragment) used in this and previous studies of step-by-step initiation kinetics [12]. **Panel B)** Single-nucleotide resolution PAGE separations of ^{32}P-labelled RNA products at different times in initiation (10-480 s after addition of 200 μM ATP, 200 μM UTP, 10 μM GTP, 17.5 nM α-^{32}P-GTP to preformed OC) in manual-mixing experiments at 0 mM and 2mM PPi. Omission of CTP causes a stop at 16-mer, with subsequent readthrough to a second stop at 31-mer. Different run-times were used for the two gels. The partial lane at left is from an unrelated experiment. **Panel C)** Time courses of synthesis of long RNA (11-mer or longer, designated as 11+) in single-round experiments at ≤ 3 mM PPi at 37 °C. Amounts of 11+ RNA are normalized by the total amount of OC. Single exponential fits are shown as solid curves. Error bars show one

standard deviation from the average of normalized data (see Methods) from 3-4 experiments. **Panel D)** Half-times $t_{1/2}$ ($t_{1/2}$ = 0.69/k_{11+}) for synthesis of 11+ RNA (black) and amounts of 11+ RNA synthesized (red) are plotted for each PPi concentration at 37 °C. Experimental uncertainties are approximately $\pm$20% for all points. Without added PPi, $t_{1/2}$ is too small to determine by manual mixing. Fast-mixing experiments at 37 °C gave $t_{1/2} \approx 1.8$ s ($k_{11+} \approx 0.4$ s^{-1}) [12]. Dashed lines are included as a visual guide.

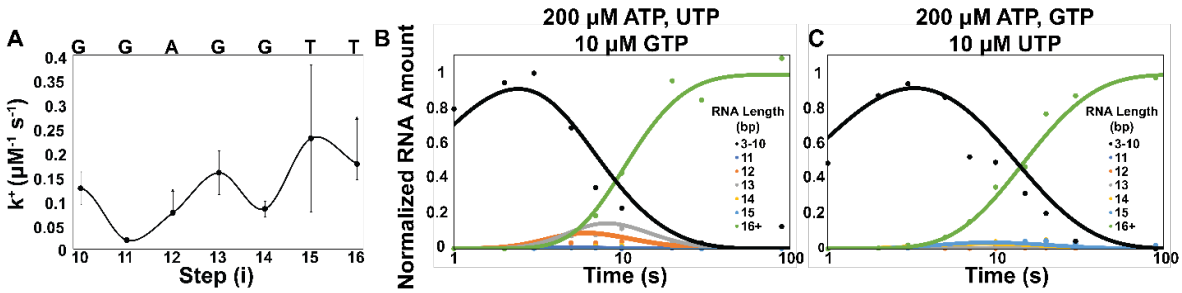


Figure 3: Composite 2nd-Order Rate Constants and Fits to Kinetic Data for Post-Escape Steps of Nucleotide Incorporation at the λP_R Promoter Without Added PPi at 19 °C. Panel A plots 2nd-order rate constants k_i^+ (k_{cat}/K_m analogs) for steps of nucleotide incorporation after promoter escape (11-mer synthesis) up to the pause at 16-mer induced by withholding CTP. These k_i^+ values are reported in SI Table S1 together with previously-determined k_i^+ for steps up to the escape point [12,15]. **Panels B and C** compare fast-mixing data (averages of 3 experiments) with fits (solid curves) of kinetic data at two sets of NTP concentrations, obtained using these composite rate constants. Amounts of RNA lengths shorter than 11-mer are consolidated for clarity.

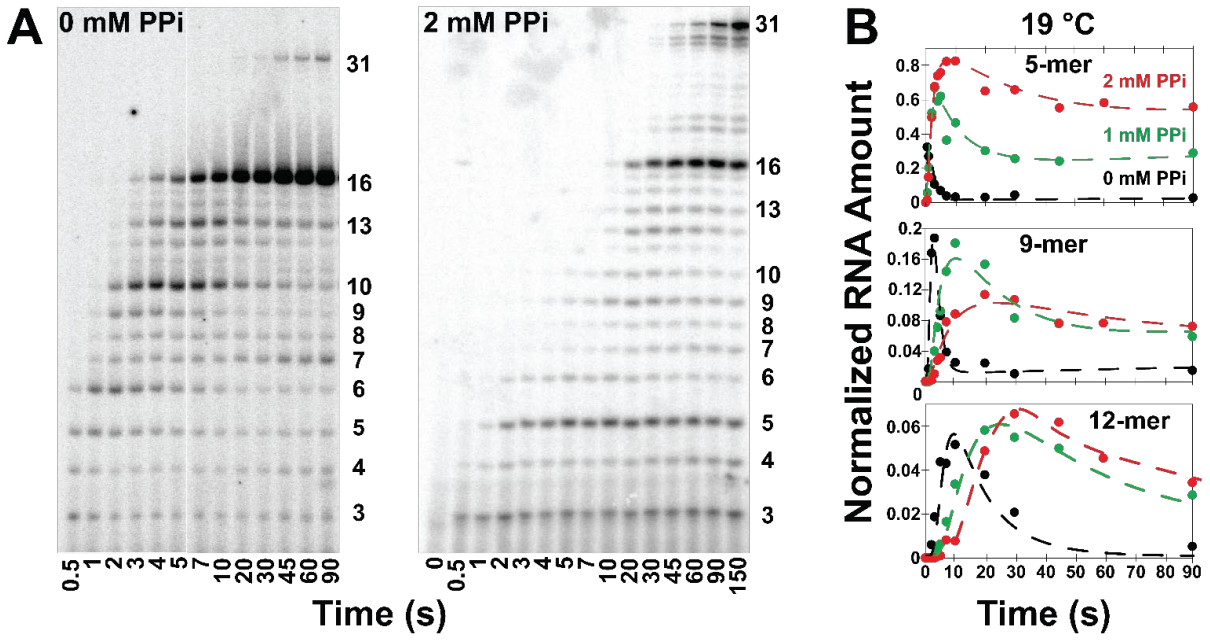


Figure 4: Effects of [PPi] on the Step-by-step Kinetics of Initiation at 19 °C. Panel A) Representative gels comparing single-nucleotide resolution PAGE separations of ^{32}P-labelled RNA products from RQF initiation experiments at the λP_R promoter at 2 mM PPi, 200 μM ATP, UTP, 10 μM GTP with results at 0 mM added PPi and the same NTP concentrations [12]. Synthesis of 16-mer RNA in these experiments is single-round. **Panel B)** Time courses of amounts of 5-mer **(Top)**, 9-mer **(Middle)**, and 12-mer **(Bottom)** RNA at 0 mM **(Black)**, 1 mM **(Green)**, and 2 mM PPi **(Red)**. Points shown are averages of 3 experiments. Amounts are normalized to the total amount of 11+ RNA synthesis in each single-round experiment before averaging. Different vertical scales are used for each RNA length. Dashed lines are included as a visual guide.

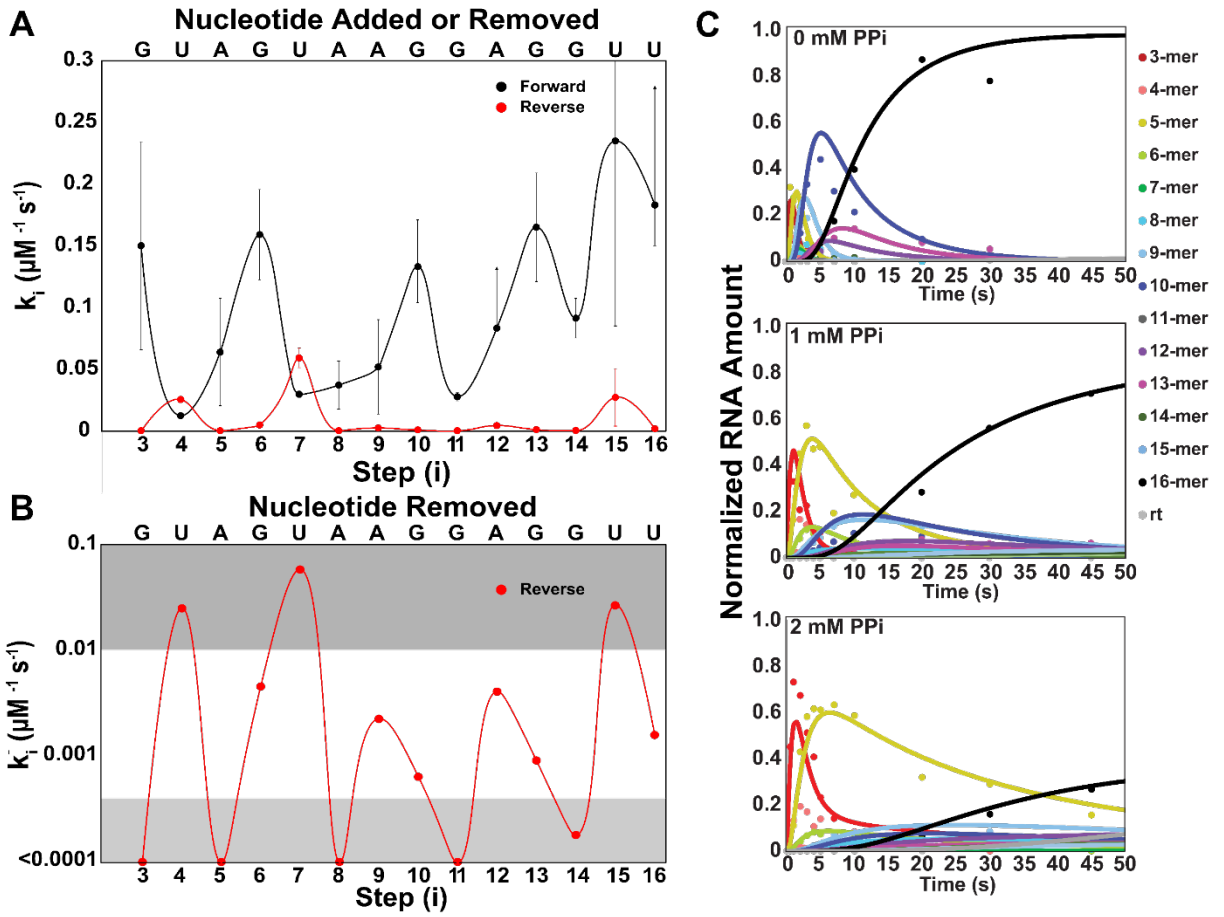


Figure 5: Step-by-step Variations in 2nd Order Pyrophosphorolysis Rate Constants in Initiation at 19 °C; Comparison with 2nd Order Nucleotide Incorporation Rate Constants. Panel A: Black data points and curve are composite 2nd order rate constants k_i^+ for each forward step of nucleotide incorporation that begins with translocation. Red data points and curve are composite 2nd order pyrophosphorolysis rate constants k_i^- for each reverse step (See Figs. 1 and S2 for mechanism; rate constants k_i^+ are from Fig 3A and Table S1). **Panel B:** Log scale plot of pyrophosphorolysis rate constants k_i^- from Panel A, separated by shading into steps with large, moderate, and small pyrophosphorolysis rate constants. **Panel C:** Comparison of RQF data (averages of 3 experiments) with fits (solid curves) for all RNA intermediates (3-15mer) in synthesis of

16-mer RNA (black) and subsequent slow readthrough (rt). Fits are shown for 0 mM PPi, 1 mM PPi, and 2 mM PPi.

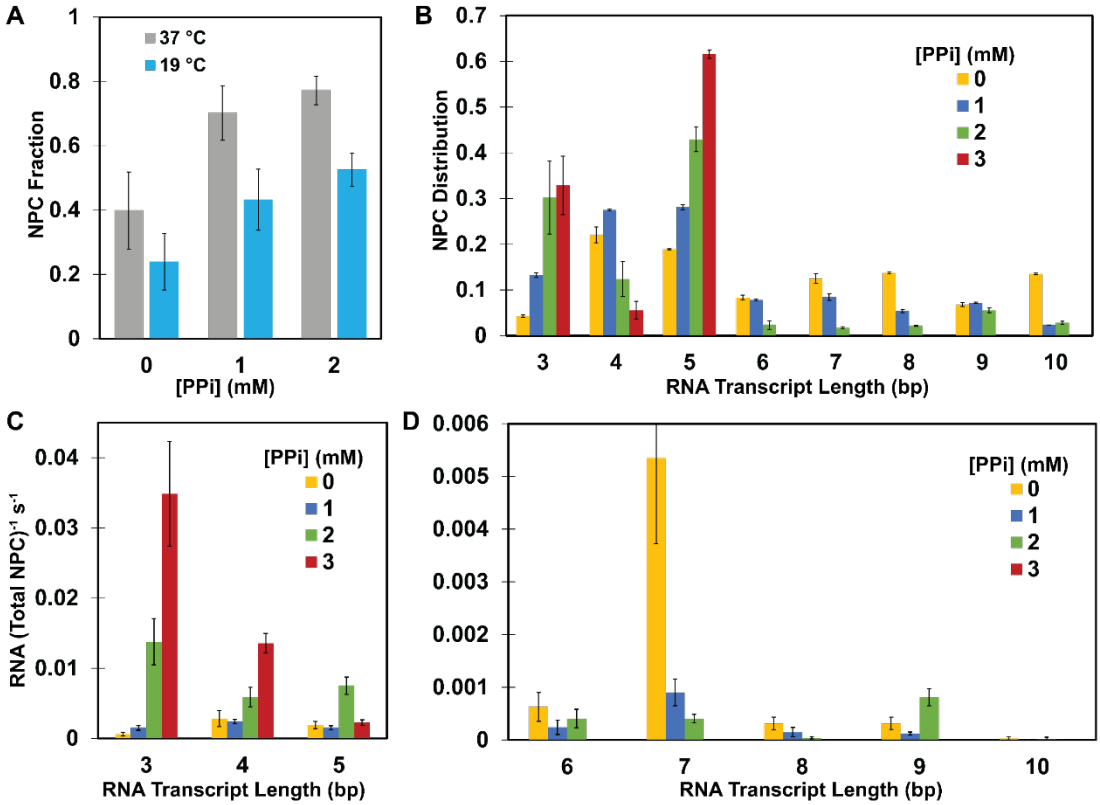


Figure 6: Effects of [PPi] on the Fraction of Nonproductive Complexes (NPC) and on the Initial RNA Length Distribution and Subsequent Abortive Rates of these NPC. Panel A) Fraction of detected initiating complexes that are nonproductive (stalling after synthesizing a short RNA (<11-mer) at 37 °C and 19 °C. **Panel B)** The fraction of detected NPC that stall after synthesis of each RNA length up to the point of escape in the initial round of short RNA synthesis at 37 °C and 0 mM added PPi **(Yellow)**, 1 mM PPi **(Blue)**, 2 mM PPi **(Green)** and 3 mM PPi **(Red)**. **Panel C)** Abortive synthesis rates, expressed as RNA synthesized per detected NPC per second, for each short RNA length as a function of [PPi]. Color scheme as in Panel B.

References

[1] Kukko-Kalske E, Lintunen M, Inen MK, Lahti R, Heinonen J. Intracellular PPi concentration is not directly dependent on amount of inorganic pyrophosphatase in Escherichia coli K-12 cells. Journal of Bacteriology. 1989;171:4498-500.

[2] Kukko E, Heinonen J. The Intracellular Concentration of Pyrophosphate in the Batch Culture of Escherichia coli. European Journal of Biochemistry. 1982;127:347-9.

[3] Heinonen J. Intracellular concentration of inorganic pyrophosphate in the cells of Escherichia coli: A method for its determination. Analytical Biochemistry. 1974;59:366-74.

[4] Buc H, McClure WR. Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry. 1985;24:2712-23.

[5] Li X-Y, McClure WR. Characterization of the closed complex intermediate formed during transcription initiation by *Escherichia coli* RNA polymerase. Journal of Biological Chemistry. 1998;273:23549-57.

[6] Heyduk E, Heyduk T. Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. Biochemistry. 2014;53:282-92.

[7] Ruff EF, Record MT, Jr. , Artsimovitch I. Initial events in bacterial transcription initiation. Biomolecules. 2015;5:1035-62.

[8] Gries TJ, Kontur WS, Capp MW, Saecker RM, Record MTJ. One-step DNA melting in the RNA polyerase cleft opens the initiation bubble to form an unstable open complex. Proceedings of the National Academy of Sciences. 2010;107:10418-23.

[9] Sreenivasan R, Shkel IA, Chhabra M, Drennan A, Heitkamp S, Wang H-C, et al. Fluorescence-Detected Conformational Changes in Duplex DNA in Open Complex Formation byEscherichia coliRNA Polymerase: Upstream Wrapping and Downstream Bending Precede Clamp Opening and Insertion of the Downstream Duplex. Biochemistry. 2020;59:1565-81.

[10] Drennan A, Kraemer M, Capp MW, Gries T, Ruff EF, Sheppard C, et al. Key roles of the downstream mobile jaw of *Escherichia coli* RNA polymerase in transcription initiation. Biochemistry. 2012;51:9447-59.

[11] Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record MT, Jr. *E. coli* polymerase determinants of open complex lifetime and structure. Journal of Molecular Biology. 2015;427:2435-50.

[12] Plaskon DM, Henderson KL, Felth LC, Molzahn CM, Evensen C, Dyke S, et al. Temperature effects on RNA polymerase initiation kinetics reveal which open complex

initiates and that bubble collapse is stepwise. Proceedings of the National Academy of Sciences. 2021;118:e2021941118.

[13] Saecker RM, Chen J, Chiu CE, Malone B, Sotiris J, Ebrahim M, et al. Structural origins of Escherichia coli RNA polymerase open promoter complex stability. Proceedings of the National Academy of Sciences. 2021;118:e2112877118.

[14] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[15] Henderson KL, Evensen CE, Molzahn CM, Felth LC, Dyke S, Liao G, et al. RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. Biochemistry. 2019;58:2339-52.

[16] Jensen D, Manzano AR, Rammohan J, Stallings CL, Galburt EA. CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. Nucleic Acids Research. 2019;47:6685-98.

[17] Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science. 2006;314:1139-43.

[18] Libing Y, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. Elife. 2017;6
.
[19] Duchi D, Bauer DLV, Fernandez L, Evans G, Robb N, Hwang LC, et al. RNA polymerase pausing during initial transcription. Molecular Cell. 2016;63:939-50.

[20] Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. Molecular Cell. 2015;58:534-40.

[21] Li L, Molodtsov V, Lin W, Ebright RH, Zhang Y. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. Proceedings of the National Academy of Sciences. 2020:201920747.

[22] Shin Y, Qayyum MZ, Pupov D, Esyunina D, Kulbachinskiy A, Murakami KS. Structural basis of ribosomal RNA transcription regulation. Nature Communications. 2021;12.

[23] Chen J, Chiu C, Gopalkrishnan S, Chen AY, Olinares PDB, Saecker RM, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. Molecular Cell. 2020;78:275-88.e6.

[24] Ko J, Heyduk T. Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochemistry Journal. 2014;463:135-44.

[25] Dangkulwanich M, Ishibashi T, Liu S, Kireeva ML, Lubkowska L, Kashlev M, et al. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. eLife. 2013;2.

[26] Belogurov GA, Artsimovitch I. The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. Journal of Molecular Biology. 2019;431:3975-4006.

[27] Mishanina TV, Palo MZ, Nayak D, Mooney RA, Landick R. Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. Proceedings of the National Academy of Sciences. 2017:201702383.

[28] Imashimizu M, Kireeva ML, Lubkowska L, Kashlev M, Shimamoto N. The Role of Pyrophosphorolysis in the Initiation-to-Elongation Transition by E. coli RNA Polymerase. Journal of Molecular Biology. 2019;431:2528-42.

[29] Kahn JD, Hearst JE. Reversibility of nucleotide incorporation by *Escherichia coli* RNA polymerase, and its effect on fidelity. Journal of Molecular Biology. 1989;205:291-314.

[30] Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS, Belogurov GA. Active site opening and closure control translocation of multisubunit RNA polymerase. 2012;40:7442-51.

[31] Liu B, Hong C, Huang RK, Yu Z, Steitz TA. Structural basis of bacterial transcription activation. Science. 2017;358:947-51.

[32] Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, et al. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science. 2016;351:1090-3.

[33] Winkelman JT, Winkelman BT, Boyce J, Chen AY, Ross W, Gourse RL. Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. Molecular Cell. 2015;59:768-80.

[34] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006;314:1144-7.

[35] Bai L, Fulbright RM, Wang MD. Mechanochemical Kinetics of Transcription Elongation. Physical Review Letters. 2007;98.

[36] Bai L, Shundrovsky A, Wang MD. Sequence-dependent kinetic model for transcirption elongation by RNA polymerase. Journal of Molecular Biology. 2004;344:335-49.

[37] Anand VS, Patel SS. Transient State Kinetics of Transcription Elongation by T7 RNA Polymerase. Journal of Biological Chemistry. 2006;281:35677-85.

[38] Hein PP, Palangat M, Landick R. RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. Biochemistry. 2011;50:7002-14.

[39] Wang MD. Force and Velocity Measured for Single Molecules of RNA Polymerase. Science. 1998;282:902-7.

[40] Righini M, Lee A, Cañari-Chumpitaz C, Lionberger T, Gabizon R, Coello Y, et al. Full molecular trajectories of RNA polymerase at single base-pair resolution. Proceedings of the National Academy of Sciences. 2018;115:1286-91.

[41] Liu B, Zuo Y, Steitz TA. Structures ofE.coliσS-transcription initiation complexes provide new insights into polymerase mechanism. 2016;113:4051-6.

[42] Kubori T, Shimamoto N. A branched pathway in the early stages of transcription by *Escherichia coli* RNA polymerase. Journal of Molecular Biology. 1996;256:449-57.

[43] Stackhouse TM, Telesnitsky AP, Meares C, F. Release of the s subunit from *Escherichia coli* RNA polymerase transcription complexes is dependent on the promoter sequence. Biochemistry. 1989;28:7781-8.

[44] Bandwar RP, Tang G-Q, Patel SS. Sequential Release of Promoter Contacts during Transcription Initiation to Elongation Transition. 2006;360:466-83.

[45] Susa M, Kubori T, Shimamoto N. A pathway branching in transcription initiation in Escherichia coli. Molecular Microbiology. 2006;59:1807-17.

[46] Skalenko KS, Li L, Zhang Y, Vvedenskaya IO, Winkelman JT, Cope AL, et al. Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in Escherichia coli. Proceedings of the National Academy of Sciences. 2021;118:e2106388118.

[47] Vvedenskaya IO, Sharp JS, Goldman SR, Kanabar PN, Livny J, Dove SL, et al. Growht phase-dependent control of transcription start site selection and gene expression by nanoRNAs. Genes and Development. 2012;26:1498-507.

[48] Wimmer JLE, Kleinermanns K, Martin WF. Pyrophosphate and Irreversibility in Evolution, or why PPi Is Not an Energy Currency and why Nature Chose Triphosphates. Frontiers in Microbiology. 2021;12.

[49] Wolfson AD, Uhlenbeck OC. Modulation of tRNAAla identity by inorganic pyrophosphatase. Proceedings of the National Academy of Sciences. 2002;99:5965-70.

[50] Johnson KA, Simpson ZB, Blom T. Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. Analytical Biochemistry. 2009;387:20-9.

Chapter 3 Appendix: Supplemental Tables and Figures

Table S1: Forward and reverse 2nd order rate constants (k_i^+, k_i^-) for nucleotide addition/excision steps 3 to 16 of initial transcription at the λP_R promoter at 19 °

Step (i)	k_i^+ ($M^{-1}\ s^{-1}$)	k_i^- ($M^{-1}\ s^{-1}$)	$K_{eq,\,i}$	3' Dinucleotide	Reversibility [a] ([PPi] = [NTP])
3	(1.5 ± 0.8) x 10^5	< 10^2	> 1000	UG	Very low
4	(1.2 ± 0.1) x 10^4	(2.5 ± 0.2) x 10^4	~0.5	GU	High
5	(6 ± 4) x 10^4	< 10^2	> 1000	UA	Very low
6	(1.5 ± 0.4) x 10^5	(4.6 ± 0.1) x 10^3	~3 x 10^1	AG	Low
7	(2.9 ± 0.2) x 10^4 [b]	(5.9 ± 0.8) x 10^4	~0.5	GU	High
8	(3 ± 2) x 10^4	< 10^2	> 1000	UA	Very low
9	(5 ± 4) x 10^4	(2.3 ± 0.9) x 10^3	~2 x 10^1	AA	Low
10	(1.3 ± 0.3) x 10^5	(6.4 ± 1.2) x 10^2	~2 x 10^2	AG	Very low
11	(2.7 ± 0.3) x 10^4	< 10^2	> 1000	GG	Very low
12	8.3 x 10^4 [b,c]	(4.1 ± 2.4) x 10^3	~20 [c]	GA	Low
13	(1.6 ± 0.4) x 10^5 [b]	(9.2 ± 3.8) x 10^2	~1.7 x 10^2	AG	Very low
14	(9 ± 2) x 10^4 [b]	(1.8 ± 0.4) x 10^2	~5 x 10^2	GG	Very low
15	(2.3 ± 1.5) x 10^5 [b]	(2.7 ± 2.3) x 10^4	~9	GU	Moderate
16 [d]	(1.8 ± 0.3) x 10^5 [b]	(1.6 ± 1.1) x 10^3	~1 x 10^2	UU	Very low
Elongation	≥ 6 x 10^5 [e]	~3 [f]	≥ 2 x 10^5	GU	Very low
Elongation	≥ 6 x 10^5 [e]	~0.5 [f]	≥ 1 x 10^6	UG	Very low

[a] Reversibility Scale: Very low ($K_{eq,\,i}$ >100); Low (20 < $K_{eq,\,i}$ < 100); Moderate (5 < $K_{eq,\,i}$ < 20); High ($K_{eq,\,i}$ < 5)

[b] Forward 2nd order rate constants k_i^+ for steps 12-16 were determined from previously published results without added PPi [12]. The fitting increased the k_i^+ value for step 7 by approximately 1.5-fold from that previously reported [12], improving the overall fit with no change in pattern of k_i values. All other forward rate constants were previously reported [12, 15].

[c] k_i^+ and $K_{eq,\,i}$ values are lower bounds

[d] Readthrough rate constants k_{rt} for extension of 16-mer RNA (first order with respect to [16-mer]) are approximately 0.0001 s^{-1} at 0 mM and 1 mM added PPi and 0.01 s^{-1} at 2 mM PPi.

[e] Calculated k_i^+ considering only NTP binding and catalytic steps of elongation at 24 °C [35, 36]. If the preceding translocation step rapidly equilibrates with an equilibrium constant of 1, this would reduce these k_i^+ values by 50%.

[f] Pyrophosphorolysis rate constants k_i^- for *E. coli* RNAP elongation steps are estimated from V_{max} and K_M values [37].

Table S2. Distribution of Initial Amounts of Each RNA Length Synthesized by Nonproductive Complexes[a,b]

[PPi]	3-mer	4-mer	5-mer	6-mer	7-mer	8-mer	9-mer	10-mer
0	0.042	0.22	0.19	0.083	0.12	0.14	0.068	0.13
1	0.13	0.27	0.28	0.078	0.084	0.054	0.072	0.023
2	0.30	0.12	0.43	0.023	0.017	0.021	0.055	0.028
3	0.33	0.056	0.61	0	0	0	0	0

[a] λP_R promoter, 37 °C, 200 μM ATP and UTP, 10 μM GTP, 17.5 nM α-^{32}P-GTP; see Fig. 6B.

[b] Population fractions are normalized by the total amount of nonproductive complexes; uncertainties are < 25%

Table S3. Normalized Abortive Rates for Subsequent Rounds of RNA Synthesis[a,b,c]

[PPi]	3-mer	4-mer	5-mer	6-mer	7-mer	8-mer	9-mer	10-mer
0	0.063	0.28	0.19	0.063	0.53	0.032	0.032	0.0031
1	0.15	0.24	0.15	0.024	0.090	0.015	0.012	0
2	1.4	0.59	0.75	0.041	0.041	0.0041	0.081	0.0014
3	3.5	1.3	0.22	0	0	0	0	0

[a] Units are RNA per NPC per 100 s; uncertainties are < 15%

[b] λP_R promoter, 37 °C, 200 μM ATP and UTP, 10 μM GTP, 17.5 nM α-^{32}P-GTP; see Fig. 6C

[c] Normalized by the total amount of nonproductive complexes

Table S4. Renormalized Abortive Rates for Subsequent Rounds of RNA Synthesis[a,b,c]

[PPi]	3-mer	4-mer	5-mer	6-mer	7-mer	8-mer	9-mer	10-mer
0	1.5	1.3	1.0	0.75	4.3	0.23	0.46	0.023
1	1.1	0.88	0.54	0.31	1.0	0.28	0.17	0
2	4.5	4.8	1.7	1.75	2.4	0.19	1.5	0.050
3	11	24	0.37	0	0	0	0	0

[a] Units are RNA per NPC per 100 s; uncertainties are < 40%

[b] λP_R promoter, 37 °C, 200 μM ATP and UTP, 10 μM GTP, 17.5 nM α-^{32}P-GTP;

[c] Tabulated rates are obtained by renormalizing entries in Table S3 using the corresponding entries in Table S2

Table S5: Primers used for λP_R template preparation

λP_R_wt_forward (-71 to -12)	CCACGAATTCGGATAAATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTG
λP_R_wt_reverse (-24 to 31)	ACAAAACCTTCATAGAACCTCCTTACTACATGCAACCATTATCACCGCCAGAGGT
HBOT	CACCTGCACCGACAAAACCTT
HTOP	CCAGCATTCCTCCACGAATTC

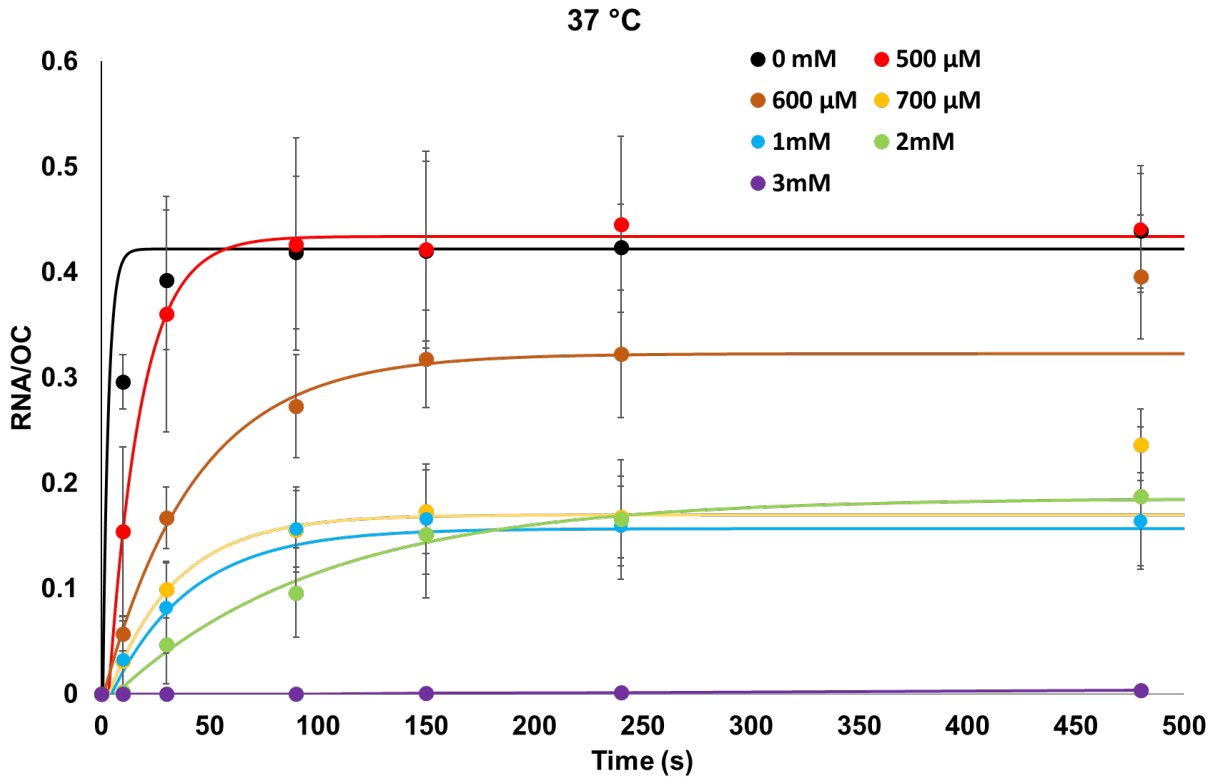


Figure S1: Long-Time Data Set (t ≤ 480 s) Used to Determine the Approach to a Plateau Amount of 11+ RNA per Open Complex (OC) in Single Round Synthesis at 37 °C. Time courses of synthesis of long RNA (11-mer or longer, designated as 11+) in single-round experiments at ≤ 3 mM PPi at the high-UTP condition at 37 °C. Single exponential fits are shown as solid curves. Error bars show one standard deviation from the average of normalized data (see Methods) from 3-4 experiments. This figure shows 480 s timepoints not plotted in Fig. 2C.

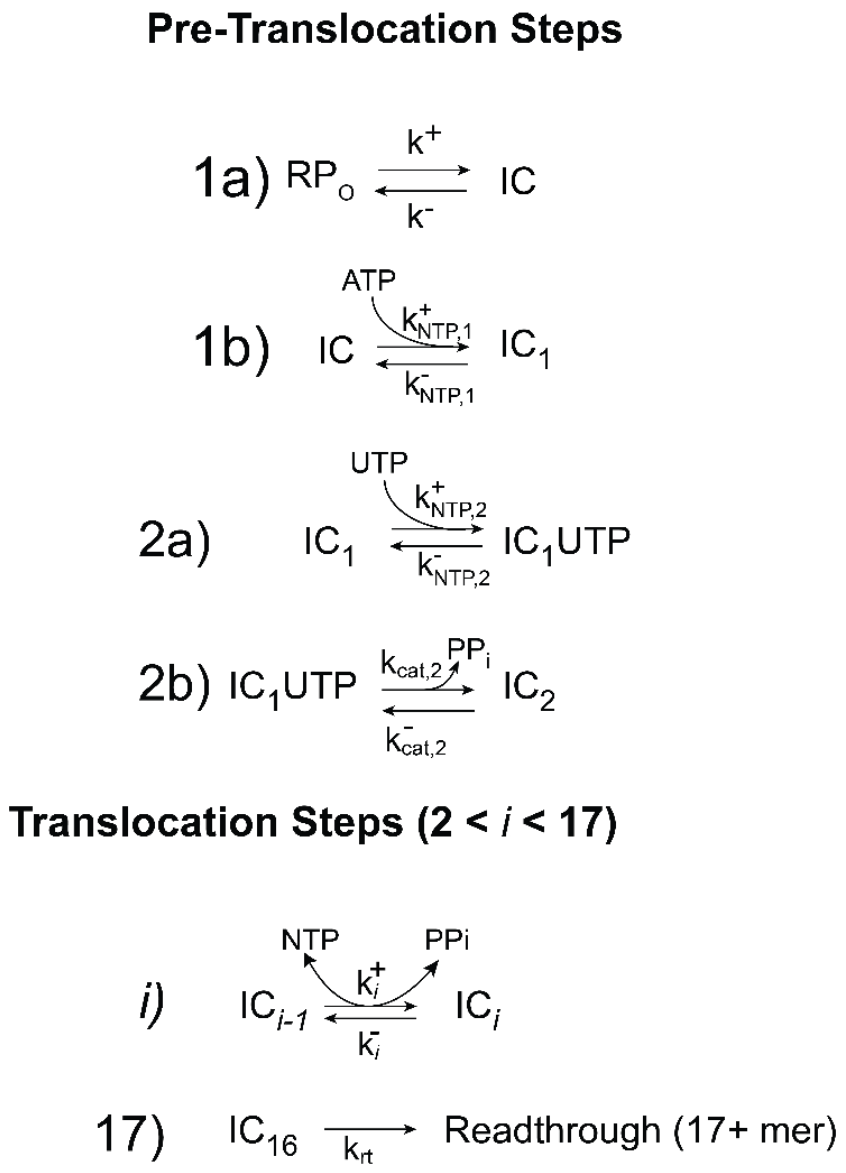


Figure S2: Extended mechanism of productive initiation at the λP$_R$ promoter. Extension of the mechanism used to fit forward rate constants at the λP$_R$ promoter [12, 15]. Briefly, a structural rearrangement step (1a) prior to NTP binding with a strongly T-dependent equilibrium constant is necessary to accurately fit initiation kinetic data, especially at 37 °C. Synthesis of 2-mer (step 2b) does not require translocation. All subsequent steps of RNA synthesis include an initial translocation step [12, 15]. The reversible kinetics of each step are described by 2nd-order rate constants k_i^+ in the forward

(synthesis) direction and k_i^- in the reverse (pyrophosphorolysis) direction. The kinetics of readthrough of the stop at 16-mer are accurately modeled as first order in [16-mer] with rate constant k_{rt}. Values of all these rate constants are given in Table S1 and those for the steps of 16-mer synthesis are plotted in Figs. 3 and 5.

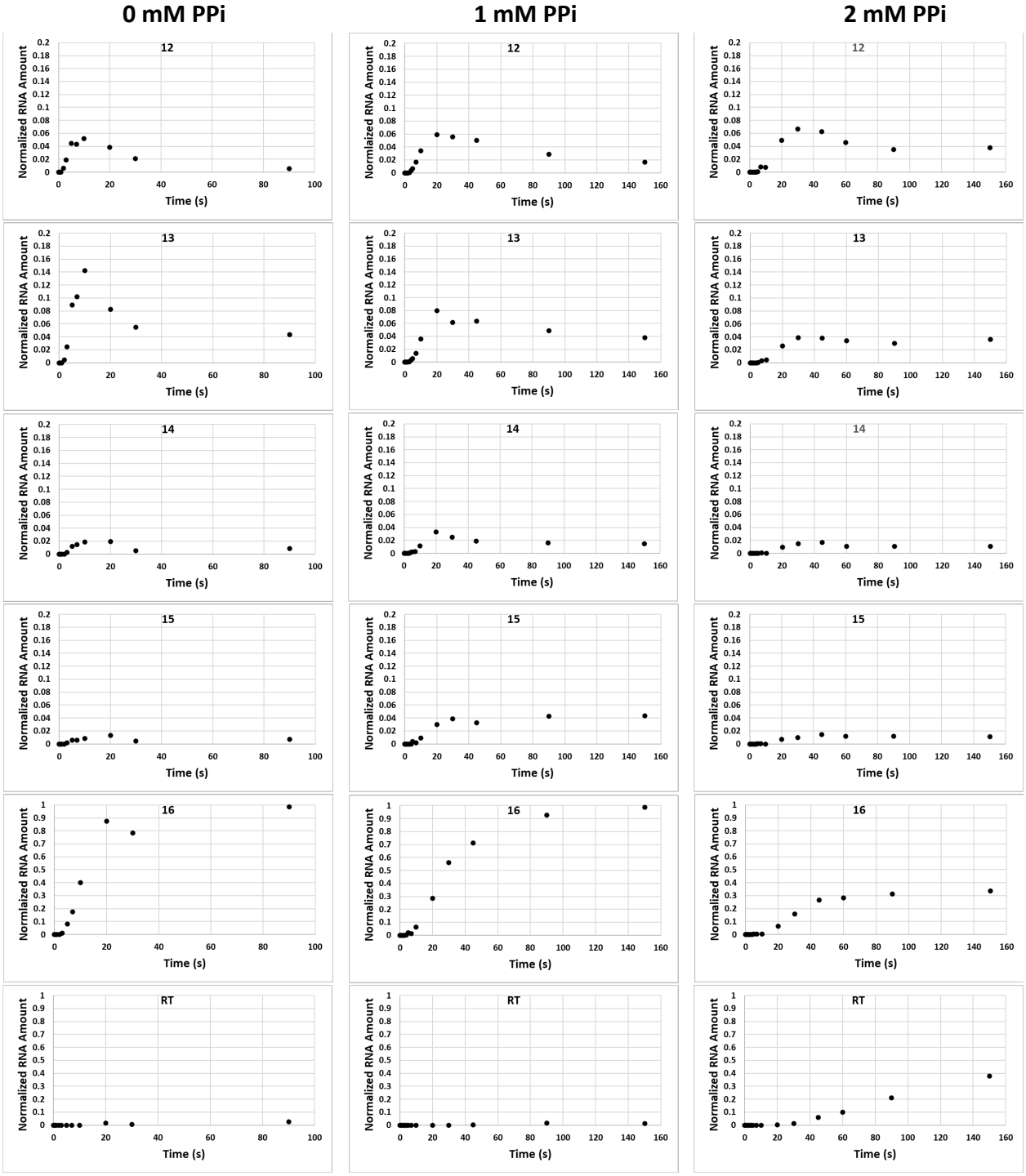


Figure S3: Comparison of Time Courses of Normalized Amounts of 12-mer to 16-mer and Readthrough (RT) RNA at 0 mM, 1 mM, and 2 mM added PPi by RQF Experiments at 19 °C. For each length and time point, the amount of RNA determined

in an experiment is normalized by the total amount of 11+ RNA synthesized in that experiment, averaged for triplicate experiments and plotted (black circles) vs time.

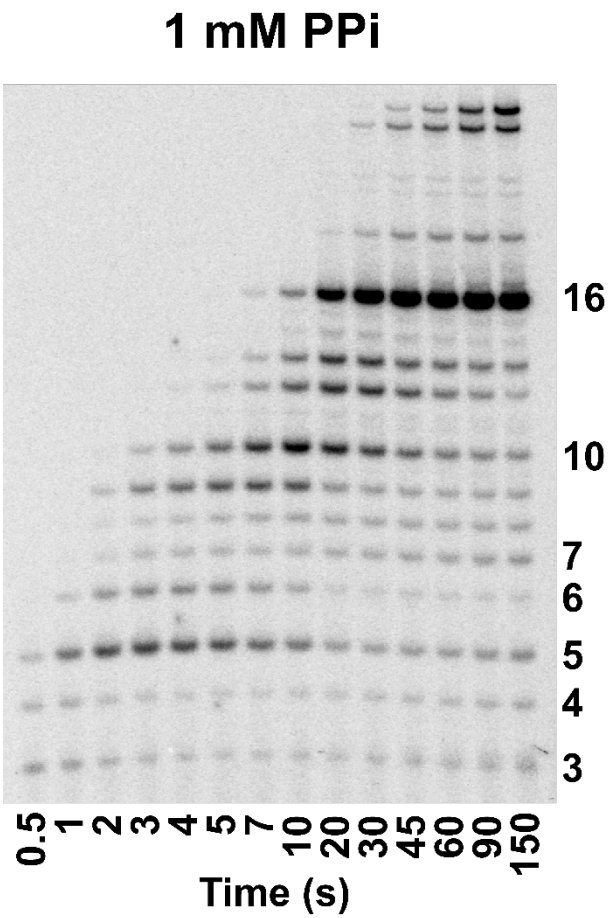


Figure S4: Representative 19 °C Initiation Kinetic Assay at 1 mM added PPi. Polyacrylamide gel electrophoresis separations with single-nucleotide resolution of ^{32}P-labelled RNA products from RQF initiation experiments at the λP_R promoter at 1 mM PPi. Synthesis of 16-mer RNA in these experiments is single-round. Corresponding gels at 0 mM and 2 mM added PPi are shown in Fig. 4.

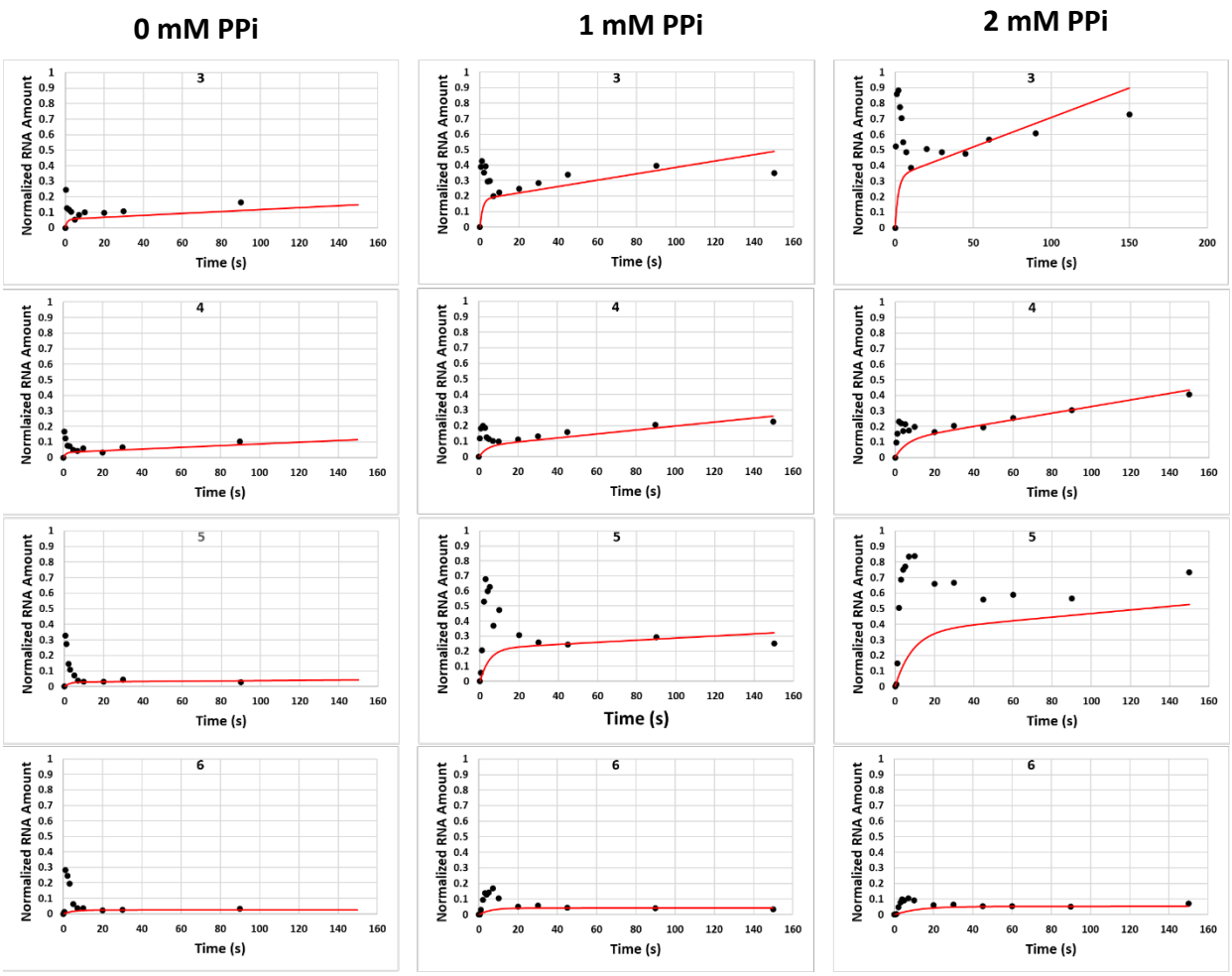


Figure S5: Comparisons of Short (3-mer to 6-mer) RNA Time Courses Determined at 0 mM, 1 mM, and 2 mM PPi by RQF Experiments at 19 °C. For each length and time point, the amount of RNA determined in an experiment is normalized by the total amount of 11+ RNA synthesized in that experiment, averaged for triplicate experiments and plotted (black circles) vs time. The predicted contribution to total RNA amount at each time (calculated as in Methods) for synthesis of these RNA lengths by nonproductive complexes that stall, slowly release their RNA and reinitiate another round of short RNA synthesis is shown in red. Subtraction of the nonproductive RNA contribution from total

RNA yields the transient population of each of these RNA lengths. Examples are shown in Fig. 4. Data at 0 mM added PPi are from ref [15].

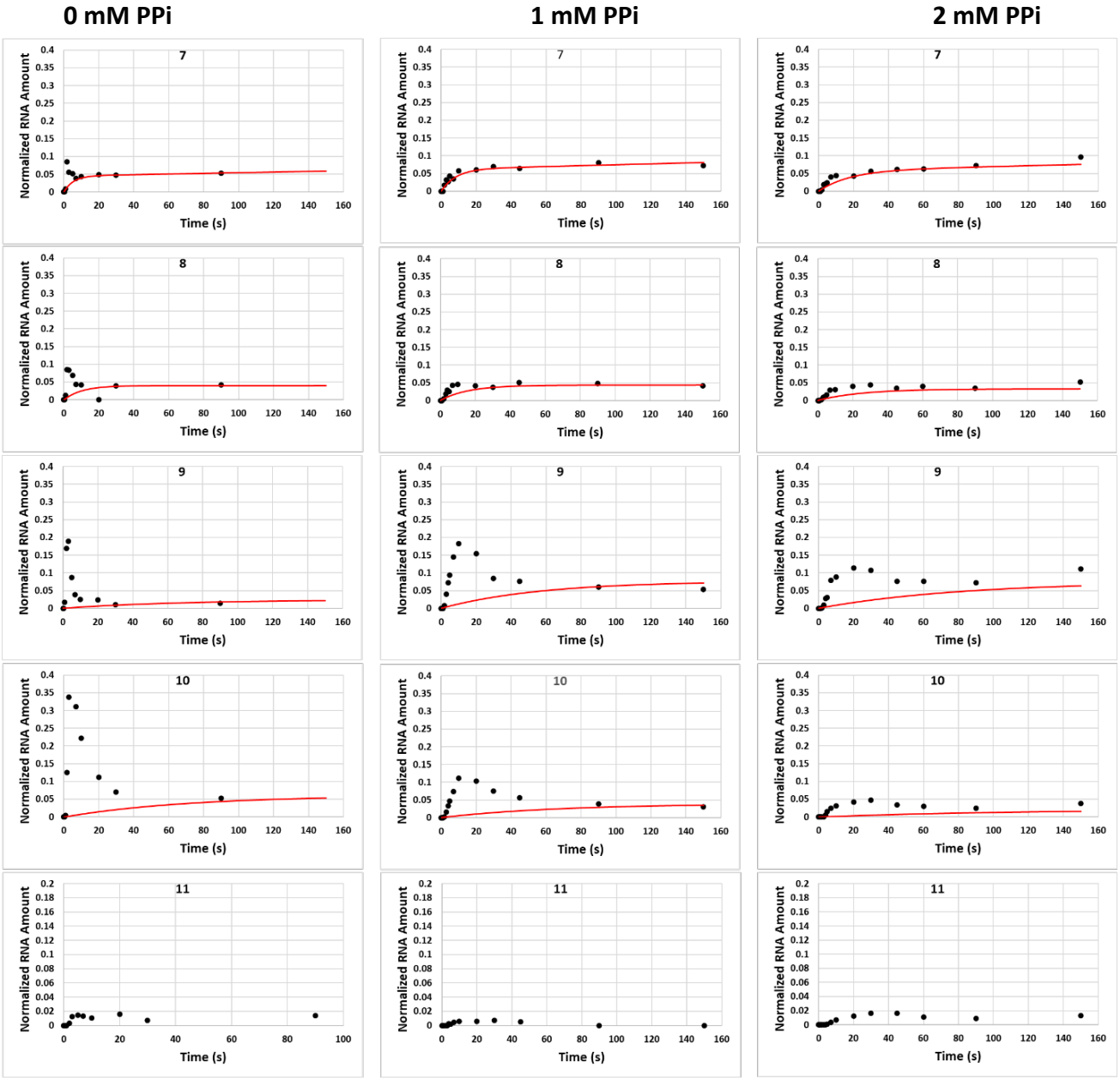


Figure S6: Comparison of Time Courses of Normalized Amounts of 6-mer to 11-mer RNA at 0 mM, 1 mM, and 2 mM added PPi by RQF Experiments at 19 °C. Black Points: averages of individual amounts of that RNA length at that time normalized by the final amount of 11+ RNA in that experiment. Red curves: predicted time courses of RNA synthesis by nonproductive complexes. Details are the same as in the caption for Fig. S5.

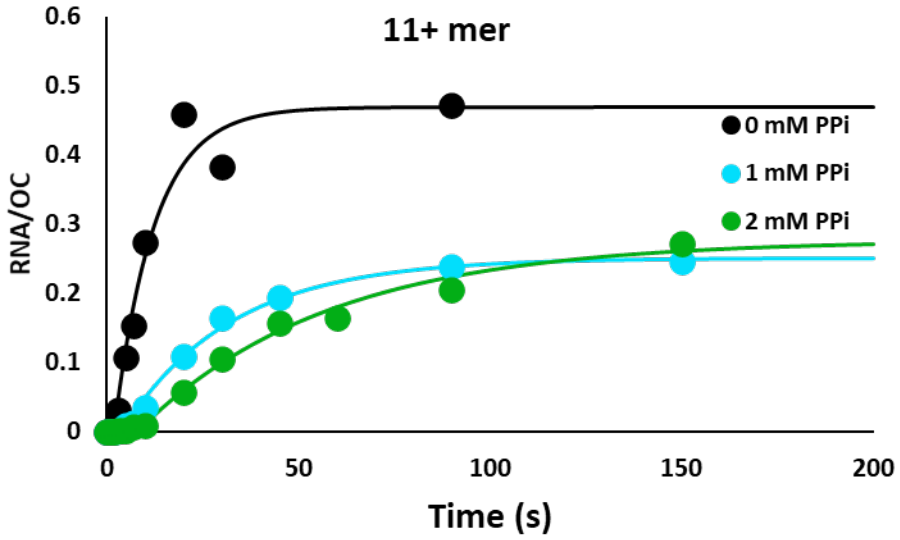


Figure S7: Comparison of Time Courses of 11+ RNA Synthesis at 0 mM, 1 mM, and 2 mM added PPi from RQF Experiments at 19 °C. Points are the sums of individual values for 11-mer to 31-mer, most of which are plotted in Fig S3. RNA amounts are normalized to total OC. Single exponential fits to these data (after a brief initial lag) are shown as solid curves. These fits yield $t_{1/2}$ values of approximately 7 s, 20 s, and 34 s for 0 mM, 1 mM, and 2 mM added PPi respectively and final normalized amounts of 11+ RNA/OC of 0.47 at 0 mM PPi, 0.25 at 1 mM PPi and 0.27 at 2 mM PPi. Uncertainties in these half times and amounts are approximately $\pm$ 20%

Chapter 4: Effects of the 8 bp *rrnB* P1 Discriminator on Productive and Abortive Transcription Initiation Kinetics

Experiments in this chapter were performed by Kate Henderson, Guanyu Liao, Tristan Gunther, and Dylan Plaskon. Analysis was performed by Dylan Plaskon, Tristan Gunther, and Tom Record.

Introduction

Attenuating energy-intensive transcription under stress conditions is important for *E. coli,* especially at the promoters for ribosomal rRNA operons which account for up to 70% of total RNA synthesis in the cell in exponential growth [1]. The most well studied of these promoters is *rrnB P1*, which is of great interest as an example of a promoter that evolved to respond to vastly different growth conditions and nutrient levels, regulated by factors including ppGpp and DksA [2-4] as well as supercoiling and NTP concentrations, especially those of the initiating NTP.

The unique promoter sequence and architecture of the *rrnB P1* promoter underpins this regulation, and understanding this interplay is critical to understanding regulation of transcription initiation. Upstream of the near-consensus -35 hexamer, the *rrnB P1* promoter contains an AT-rich UP element that binds the C-terminal domains of the RNAP α-subunits [5]. The spacer between the -35 and -10 element is unusually short at 15 bp, while the discriminator region is unusually long (8 bp) and GC rich. The combined length of the spacer and discriminator is 23 bp, the same as for the consensus (17 bp spacer, 6 bp discriminator). This contributes to selection of the proper transcription start site (TSS) [6].

The open complex (OC) formed by σ^{70} RNAP at the *rrnB P1* promoter is very unstable [7, 8]. Since a nontemplate strand purine TSS is very much favored over pyrimidine TSS [9, 10], the *rrnB P1* discriminator has to be either 8 bp or 5 bp. These are thought to entail similar free energy costs [11], but the TSS with an 8 bp discriminator is the observed TSS for this WT promoter. This 8 bp discriminator region is thought to be

the main source of OC destabilization at the *rrnB P1* [6, 7, 12]. This destabilization is the basis of ppGpp and DksA regulation of this promoter [8].

Correlations between OC instability and earlier escape points [7] and escape efficacy [13] have also been reported. OC instability at the *rrnB P1* promoter has also been suggested to be the source of reduced or undetected abortive synthesis from the *rrnB P1* promoter [6, 14]. Previous studies interpreted abortive synthesis as on the pathway to productive RNA synthesis, which has been challenged for promoters with more stable OC [15-19].

Initiation of RNA synthesis at the *rrnB P1* promoter is very sensitive to the concentrations of the initiating (+1 and +2) NTPs (iNTPs), more than expected for binding to an available initiation site. This sensitivity to iNTP concentration is a key part of regulation of this promoter to respond to high or low nutrient levels [20, 21]. The origin of this magnified effect of concentration of the iNTP on initiation from rRNA promoters is the instability of the binary open complex (OC) at ribosomal RNA promoters, especially *in vitro* on linear promoter DNA. At low iNTP concentrations for a linear promoter fragment *in vitro*, without additional factors, only a small (equilibrium) fraction of the *rrnB P1* promoter population have the open start-site region needed to bind the iNTP. The majority of RNAP- *rrnB P1* promoter complexes are closed complexes (CC), in which the region of the TSS is duplex DNA and can't initiate [3]. Binding of the initiating NTP to the small fraction of RNAP-promoter complexes that are OC shifts the CC ↔ OC equilibrium to favor OC and let all complexes initiate [20].

Studies of the *rrnB P1* discriminator variants reveal a strong dependence of OC lifetime and choice of TSS on discriminator sequence [6]. For example, mutation of the C

to a G 2 bp downstream of the -10 element (called +2 here; see Figure 1 below) stabilizes the OC and shifts the predominant TSS from an 8 bp to a 5 bp discriminator position [6]. This shift in TSS eliminates the prescrunching of the 8 base discriminator strands in the OC formed by the WT *rrnB P1* promoter, replacing that with stretching of the 5 base discriminator strands. This indicates that the contacts made within the cleft by the pre-scrunched DNA are sequence-dependent [6, 12, 22].

The consequences of this pre-scrunching, open complex stability, and TSS selection on the kinetics of promoter escape not yet well understood. The goals of this research are to better understand the roles of the unusually long 8 bp *rrnB P1* discriminator and the core *rrnB P1* promoter in start site selection, the transition from initiation to elongation, the rate of escape of RNAP from the promoter to synthesize long RNA, and the properties of abortive RNA synthesis.

Here, we utilize single-base resolution *in vitro* transcription assays to determine that promoter escape from two hybrid promoters containing the *rrnB P1* discriminator can be modeled as a two-step process, where the first step, presumably a conformational change in the OC in order to initiate, involves binding of the 2nd initiating nucleotide. We confirm the finding that initiation kinetics are heavily dependent on the concentration of the 2nd iNTP, and find that this dependence is largely overcome by a dinucleotide primer. Additionally, we investigate short RNA synthesis by nonproductive complexes from the *rrnB P1* promoter.

Results

Design of promoter constructs with the rrnB P1 discriminator.

Two promoter constructs containing the native *rrnB P1* 8-bp discriminator embedded in 124 bp fragments extending 82 bp upstream and 42 bp downstream of the

TSS were investigated. Construct R(R) has *rrnB P1* core promoter elements (-10, -35, spacer, UP), and construct L(R), has the corresponding λP_R core promoter elements (Fig. 1A). Both constructs have the *rrnB P1* discriminator and a modified λP_R initial transcribed region (ITR) with no G bases until position 17 so initial transcription stops at a 16-mer RNA when CTP is withheld. To enable comparisons with λP_R and other promoters with different discriminator lengths, the first base of the discriminator region (adjacent to the -10 hexamer) is designated position 1 for all constructs (Fig. 1). The WT *rrnB P1* promoter initiates at position 9 (corresponding to an 8-bp discriminator) to make the dinucleotide pppGpC from the WT *rrnB P1* ITR. Use of the modified λP_R ITR with R(R) and L(R) retains a purine-pyrimidine RNA sequence (A,U) at positions +9 and +10 and enables comparison with other promoters studied in constructs with the same ITR [15-17].

Different Start Site Distributions for R(R) and L(R) Promoters

Primer extension was used to determine the TSS distributions for R(R) and L(R) promoters at 200 μM all NTPs on linear template DNA at 37 °C. Over 99% of R(R) full length transcripts start at position +9A (with an 8-base discriminator), completely consistent with the observed behavior of WT *rrnB P1* [6] (Fig. 2). At the L(R) promoter, 33% of transcripts start at +9A (8-base discriminator), and 66% at +6A (5-base discriminator). For comparison, the L(L) TSS is entirely +6A (6-base discriminator). Comparisons with effects of upstream sequence variants on start site selection by *rrnB P1* [6] indicate that the observed difference in TSS selection between R(R) and L(R) promoters may arise from differences in their extended -10 and spacer regions (see Discussion). No initiation of long (post-escape, 8+) RNA synthesis at position 8(C) or

position 7(C) (corresponding to 7 and 6 base discriminators) is observed for either promoter.

Time Courses of Initiation by Productive and Nonproductive OC at R(R) and L(R) Promoters

CTP was withheld in previous L(L) transcription initiation kinetic assays to ensure that synthesis of long RNA by productively initiating complexes is single-round. Withholding CTP in R(R) and L(R) transcription assays also eliminates the possibility of initiation at position +6(A), the preferred L(R) start site (corresponding to a 5 base discriminator) when all four NTP are present (Fig. 2A), as well as at +7(C) and +8(C).

Manually-mixed transcription initiation assays were performed with RNAP preincubated with linear template DNA at 37 °C. Two different NTP conditions were used: 200 μM UTP and ATP, 10 μM GTP, 17.5 nM α-^{32}P-GTP (designated “high UTP”), and 200 μM GTP and ATP, 10 μM UTP, 17.5 nM α-^{32}P-UTP (designated “low UTP”). Two G bases in the template strand of the ITR serve as transcription stop-points at RNA lengths of 16-mer and 31-mer when CTP is omitted. Some readthrough of the stop at 16-mer is observed, but no readthrough of the 31-mer stop occurs, and a RNAP halted at these positions prevents binding of a second RNAP at this promoter. Hence it was not necessary to add a competitor to bind free RNAP and prevent reinitiation. Heparin could not be used because of the instability of these RNAP-promoter complexes.

Figure 3A shows representative gel separations of RNA populations synthesized over a 480 s time course of initiation at R(R) and L(R) promoters at the high UTP and low UTP NTP conditions. As observed previously for L(L) as well as the T7A1 promoter T(T)

and the variants L(T) and T(L), these gels provide evidence for two phases of initiation and two classes of promoter complexes, productive and nonproductive [17]. In the faster phase of initiation, RNAP in productive OC makes a long enough RNA-DNA hybrid to escape from the promoter, while RNAP in nonproductive OC stalls after making an initial short RNA. In the slower phase of initiation, nonproductive complexes may release their RNA, return to the initiation complex and reinitiate, making another short RNA to start cycles of abortive initiation.

In the slower phase of initiation (from Fig 2 gels defined as t > 30 s at the high U condition), abortive synthesis of 7-mer and some 3-mer and 4-mer is observed for both R(R) or L(R) at high U. By contrast, little if any abortive synthesis is observed for R(R) at low U. No significant abortive synthesis is observed for any RNA longer than 7-mer for either R(R) or L(R). We therefore designate the 7-mer to 8-mer step as the escape point and define long (post-escape) RNA as any length > 7-mer, abbreviated 8+ RNA. For L(L) the longest abortive RNAs are 8-mer, 9-mer and 10-mer and we therefore assume promoter escape of RNAP occurs at the 10-mer to 11-mer step [17].

Very few, if any, short RNA transients in synthesis of long RNA by productive complexes are detected at either R(R) or L(R) at either NTP condition, even though half times for long RNA synthesis by L(R) at both NTP conditions and for R(R) at low U NTP are long enough (30-50 s, see below) for significant transients to be detected in these manual mixing experiments. This indicates an unfavorable or slow conformational change in the RNAP-promoter complex (perhaps the conversion of a closed complex to an open initiation complex) prior to initiation of short RNA synthesis at these promoters and conditions. Other unfavorable conformational changes required for initiation are observed

for λP_R at 25 °C and 37 °C and for T7A1 at 19 °C (see Chapters 2 and 5). Transient buildup of 12-mer and 13-mer RNA is detected at the high UTP condition for both promoters. These transients at high UTP (which is also low GTP) but not low UTP (which is also high GTP) can be explained by the fact that GTP is the nucleotide that is bound to extend both 13-mer and 14-mer, but this does not explain why a similar effect is not seen in earlier, apparently post-escape, steps where GTP is incorporated (10-mer, 11-mer synthesis). Significant readthrough of the pause after 16-mer synthesis induced by withholding CTP is observed to various degrees at both NTP conditions for both promoters. This effect is relatively consistent between replicates, and may be the result of misincorporation or contaminated NTP stock. This readthrough does not affect the quantification of long RNA.

Long (post-escape) RNA synthesized by productive complexes, defined as 8 bp and longer, was quantified over 480 s from 3 or more independent experiments for each NTP condition and promoter construct as described previously. Phosphorimager amounts of different RNA lengths (12-mer to 31-mer; no RNAs from 8-mer to 11-mer are detected) were corrected for multiple incorporations of radiolabeled NTP as previously described [15]. To directly compare the kinetics of transcription initiation in all conditions, long RNA in each experiment was normalized by dividing by the long RNA plateau to report on the progress to the plateau over time (Fig. 3B). The plateau in the amount of long RNA (Fig. 3B) indicates that only a single round of long RNA synthesis occurs.

High Concentration of the 2nd Initiating NTP (UTP) Greatly Increases the Rate of Long RNA Synthesis at R(R), but not L(R)

At the high UTP condition Fig. 3B shows that synthesis of a long RNA by productive R(R) complexes is significantly faster than for L(R). At low UTP, long RNA synthesis from both promoters is similarly slow and comparable to that of L(R) at high UTP. The rate of productive initiation is quantified for each experiment via the single exponential decay constant k_{8+} and half time $t_{1/2} = 0.69/k_{8+}$. Single-exponential fits with these rate constants are plotted in Figure 3B, and k_{8+} is plotted in Figure 3D. Long RNA synthesis is 5 times faster for the R(R) promoter at high UTP ($t_{1/2}$ of about 7 s) than at low UTP ($t_{1/2}$ of about 35 s). For L(R), $t_{1/2}$ is about 35 s at high UTP and about 46 s at low UTP.

Low Concentration of the 2nd Initiating NTP (UTP) Greatly Reduces Abortive Transcription R(R), but not L(R).

Total amounts of short RNA (< 8-mer) synthesized as a function of time by nonproductive R(R) and L(R) complexes for both NTP conditions investigated at 37 °C are quantified in Figure 3C. All amounts are normalized to the amount of long RNA synthesized by productive complexes at that condition. Very little short RNA synthesis is detected from nonproductive complexes at the low U condition. Intercept values provide an estimate of the ratio of nonproductive to productive initiating complexes at each promoter and NTP condition. For L(L) this ratio was approximately 1 for both high U and low U NTP conditions [15, 17]. For R(R) this ratio is 0.19 at the high U condition and 0.07 at the low U condition. There are very few nonproductive RNA detected at the low U condition. This difference may indicate that nonproductive R(R) complexes are unable to synthesize the initial dinucleotide at low U. The situation is different for L(R), where the ratio of nonproductive to productive complexes is 0.21 at the high U condition and 0.27

at the low U condition. Table 2 lists these values for the different promoter variants investigated at 37 °C.

After this initial phase of synthesis of the first short RNA by nonproductive complexes at R(R) and L(R) promoters, a slower phase of RNA release and re-initiation (abortive initiation) is observed. For R(R) at the low UTP condition, the abortive rate from the slope in figure 3C is very small in comparison to L(R) and to R(R) at high U (Table 2). This is not simply because there are fewer total nonproductive complexes at this low U condition. The ratio of slope to intercept for the linear fits at each condition represents the average rate of abortive RNA synthesis per nonproductive complex, compared on the right axis of Figure 3D. This normalization shows that abortive rates per complex are similar for R(R) at high UTP, L(R) at high UTP and L(R) at low UTP, while R(R) at low UTP shows a much smaller rate of abortive synthesis per complex.

Dinucleotides CpA and CpC Prime Initiation at R(R) and L(R)

Primer extension experiments are performed with 80 μM dinucleotide primer and 200 μM of all four NTPs. The two primers investigated are CpA, which is complementary to positions 8 and 9 on the template strand, and CpC, which is complementary to position 7 and 8 on the template strand. Primers were preincubated with RNAP and promoter DNA before adding NTP as described in Methods. The presence of CpA causes both R(R) and L(R) to start primarily with CpA. This corresponds to a 7 bp discriminator if CpA is positioned in the active site upon initiation. For the CpA and NTP concentrations used here, about 2/3 of R(R) productive complexes (64%) initiate with CpA and the remainder initiate with ATP. A somewhat smaller percentage of L(R) productive complexes initiate with CpA (59%) while the other 41% initiate with ATP in a similar ratio (30% at position 6,

11% at position 9) to that observed in the absence of CpA (Figure 4B). Uncertainties in these values is approximately 10-15%.

At 80 μM, CpC also is able to prime RNA synthesis but to a smaller extent than for CpA, even though a CpC start corresponds to a 6 base discriminator if CpC occupies the active site. In CpC-primed R(R) start site experiments, 85% of long RNA started at the position 9 site with ATP and only 15% start with CpC at position 7. In L(R) CpC experiments, 39% of productive RNA start with CpC, 42% start at position 6 and 19% start at position 9 (Figure 4B). Only long RNA synthesis (longer than 32 bp) by productive complexes is detected via primer extension.

Dinucleotide Primer CpA Increases the Rate of Transcription Initiation from R(R) and L(R) when UTP is Limiting.

Figure 5A shows time-courses of long RNA synthesis at R(R) and L(R) at the high UTP condition with and without added CpA. There was not enough detectable transcription primed by CpC to accurately determine the long RNA synthesis kinetics. RNA synthesis that is primed with CpA can be quantified from the amounts of 14-mer, 17-mer, and 32-mer RNA detected in the primed kinetic experiments, each of which is 1 base longer than the most intense detectable bands from unprimed RNA synthesis at the R(R) and L(R) promoter constructs. In this way, the kinetics of long RNA synthesis for CpA primed transcription could be determined despite the multiple TSS. The overlaid RNA pattern of CpA primed and unprimed RNA, however, prevents accurate assignment of RNA lengths other than 13-mer, 16-mer, 31-mer, and their CpA primed counterparts, and thus for these experiments long RNA is defined only as the sum of the amounts of

these RNA species. This has only a small effect on the total amount of long RNA detected, because >90% of long RNA is detected at these three lengths.

Figure 5B shows incomplete priming at the high UTP condition for R(R) and L(R), calculated from the ratio of CpA primed long RNA to total long RNA. At the low UTP condition, there is near-complete (>93%) priming of long RNA synthesis at both R(R) and L(R). The differences between the percent of long RNA that is CpA primed in these kinetic experiments and in those reported in Figure 4 is within error. All four NTPs are present in equal amounts in the primer extension experiments in Figure 4.

Values of k_{8+} for primed long RNA synthesis from R(R) and L(R) with each NTP condition were determined using the same method as for unprimed k_{8+} and are reported in Figure 5C. For both R(R) and L(R) CpA increased k_{8+} at low UTP but did not significantly change k_{8+} at high UTP. The overall effect of this change is to reduce the UTP dependence of k_{8+} at R(R) from a 5-fold increase between low and high UTP to a 2-fold increase. At L(R), the low UTP condition initiates approximately 3 times faster than the high UTP condition when CpA is present at the concentration investigated here.

Discussion

Promoter Element Determinants of TSS Selection with the rrnB P1 Discriminator

An 8 bp discriminator requires two bases of each strand of discriminator DNA to prescrunch within the RNAP cleft, which is predicted to destabilize the *rrnB P1* open complex [6]. Several promoter sequence and architecture factors combine to facilitate the correct TSS selection despite thermodynamically unfavorable prescrunching. Mutational studies showed that the C-7G discriminator sequence mutation shifts the TSS to primarily position 6 (corresponding to a 5 bp discriminator), and the C-16G / C-17T mutations, which occur in the extended -10 element, also shift the TSS to position 6 [6].

To facilitate comparisons between previously-determined detailed kinetic data at the λP_R promoter [15-17], the hybrid promoters R(R) and L(R) were constructed and studied here. Over 99% of transcripts from the R(R) promoter, which contains the λP_R ITR and the *rrnB P1* upstream promoter and discriminator elements, started at the same +9 position as for the WT promoter. This indicates that sequence determinants of TSS preference for rrnB P1 are primarily upstream of the ITR. However, some sequence features that may be relevant for TSS selection are shared between the λPR and rrnB P1 ITRs, most notably the preference for purine (A and G) TSS shared by nearly all E. coli promoters [23].

Substituting the *rrnB P1* upstream promoter elements and ITR for the corresponding λP_R promoter elements to form the L(R) promoter construct favors selection of the TSS at position 6, although only 66 ± 3% of transcripts start at this position. The remaining 34 ± 3% of transcripts start from the expected position 9. A TSS preference of >99%, as seen at R(R), indicates a standard free energy difference between IC with different TSS of -2.8 kcal/mol (or more negative) at 310K. The 2:1 TSS preference seen at L(R) is a standard free energy difference of ~ -0.4 kcal/mol. The nucleotide differences at the positions 2 and 3 bp upstream of the -10 element, and the length of the λP_R spacer region are the most likely sources of the TSS difference between R(R) and L(R). Changes in the *rrnB P1* spacer at these positions to more closely resemble the λP_R spacer favored TSS selection at position 6 in previous research at the WT *rrnB P1* promoter [6]. The promoter sequence determinants of TSS selection at the L(R) promoter construct require further study.

Productive Initiation at R(R), but not L(R), is Strongly Dependent on the Second iNTP

In previous studies of L(L) transcription initiation kinetics, single-round RNA synthesis was ensured by addition of heparin with the NTPs to preformed RNAP-promoter complexes to bind excess RNAP and any RNAP that dissociates from the promoter [15-17]. At R(R) and L(R) no initiation was observed when heparin was added with the NTP, presumably because of the instability of open complexes at these promoters. However, by withholding CTP from the reaction, single-round RNA synthesis was still achieved, confirmed by the behavior of long RNA amount vs. time, which is well-modeled by a single-exponential approach to a plateau with rate constant k_{8+}. The amount of long RNA would not plateau if transcription were multi-round, and instead would continue in an approximately linear manner over the time of this experiment. Single-round synthesis is further confirmed by the absence of a significant amount of runoff products (RNA length 42+). Withholding CTP from the reaction mixture also prevents any RNA synthesis initiating from position 6 at both R(R) and L(R), as positions 7 and 8 are both CTP. Significant readthrough of the +17C is seen at all conditions and promoter constructs studied, but does not affect quantification of long RNA, which includes all post-escape RNA products. This readthrough may be the result of CTP contamination or misincorporation.

For R(R), k_{8+} shows a dramatic UTP dependence increasing approximately 5-fold between the low and high UTP conditions. This dependence is similar to that seen in previous studies of the *rrnB P1* promoter, which report an iNTP concentration effect of approximately the same magnitude [20, 21]. A ~10-fold difference in k_{8+} is also seen between the high and low UTP conditions at the L(L) promoter (Table 1) [15] at 37 °C, which in contrast to R(R) forms a very stable OC [17]. At L(L), the UTP dependence of

k_{8+} at 37 °C is well explained by the stabilization of an intrinsically-unfavorable I_3-like OC form which is capable of productive initiation, and is in equilibrium with the more stable RP_O which is incapable of productive initiation. At R(R), the heparin sensitivity of the equilibrated 37 °C complex indicates the stable RNAP-promoter complex may be closed, but in equilibrium with the unstable but initiation-competent OC. In this way the stable complex formed by R(R) at 37 °C behaves identically to the one formed at the WT *rrnB P1* promoter [21, 24].

For L(R), however, both the high UTP and low UTP conditions exhibit similarly small k_{8+} and therefore no dependence on UTP concentration. Several possible mechanisms can be proposed to explain this. The L(R) promoter construct may be sufficiently stabilized in an initiation-competent form by reversible binding of ATP alone, the concentration of which does not change between our two NTP conditions. Alternatively, the overall first step in the initial transcription mechanism at the L(R) promoter may not be a rapid equilibrium step stabilized by reversible UTP binding to the initiation competent, but unstable state as in L(L) and R(R). Instead, L(R) may form an initiation competent OC that is naturally slow to initiate and is therefore not strongly dependent on stabilization by iNTPs. A third possible interpretation is that reversible GTP binding at position +3 to enable 3-mer synthesis is required to stabilize the initiation-competent OC form at L(R). We discuss which of these mechanisms best fits our data in our discussion of dinucleotide priming kinetics below.

Initiation is significantly slower at R(R) and L(R) than L(L). This is surprising because the *rrnB P1* promoter on which these constructs are based is one of the most transcribed promoters in the cell. Slower initiation here may be because we are using a

linear promoter fragment, while the in vivo promoter is embedded in negatively supercoiled DNA. Negative supercoiling reduces the energetic barrier to forming the single-stranded initiation bubble and therefore should stabilize the R(R) and L(R) OC and increase the rate of initiation.

Short RNA Synthesis by Nonproductive Complexes at Both R(R) and L(R) Promoters Extends to 7-mer, and is [iNTP]-Dependent at R(R) but not L(R)

Two distinct classes of complexes are detected in these experiments. RNAP in productive complexes successfully escape from the promoter and synthesizes a long RNA, as described in the previous section. Nonproductive complexes stall after synthesizing a short (< 8-mer) RNA, then release that RNA and reinitiate. At R(R) and L(R), detectable amounts of these short RNAs are observed at all lengths up to and including 7-mer. The total amount of short RNA increases linearly with time for all constructs at all conditions investigated here, well past the time where the amount of long (productive) RNA reaches a plateau. This demonstrates unequivocally that these must be at least two separate populations of complexes, as has been previously reported at other promoters [15-17, 19].

The length of the longest RNA synthesized in abortive initiation by nonproductive complexes has been used to define the point of escape for a given promoter construct with the rationale that a nonproductive complex cannot re-initiate after all promoter contacts have been broken, and that these contacts break at the same RNA length in productive and nonproductive complexes. The longest detected abortive RNA at both the R(R) and L(R) promoters is 7 bp, indicating that by this reasoning, the escape point for these promoters must be at the 7-mer to 8-mer extension step. However, predictions of

escape point based on pre-scrunching energetics suggest that promoters with the OC instability of *rrnB P1* and R(R) may escape as early as the 4-mer extension step [17]. These two proposals for the escape point are difficult to reconcile. It is our view that it would be difficult for RNAP to escape from the R(R) or L(R) promoters before displacing $\sigma_{3.2}$ from the cleft [25, 26], and that at the R(R) and L(R) promoters this occurs after 7-mer synthesis together with the disruption of any remaining upstream promoter contacts. This proposal is consistent with crosslinking analyses of *rrnB P1* initial transcribing complexes which indicate that scrunching (which does not occur post-escape in elongation) occurs at least until 6-mer RNA synthesis [22].

The total amount of nonproductive complexes is equivalent to the amount of RNA produced in the initial round of abortive RNA synthesis and can be approximated from extrapolation of the linear phase of abortive synthesis to time 0 (Table 2). Normalization to the linear abortive synthesis rate to this intercept for each condition yields the rate of abortive synthesis per individual nonproductive complex (figure 3D). At the R(R) promoter this per-complex abortive synthesis rate exhibits an approximately 8-fold increase between the low and high UTP conditions, a similar UTP dependence as k_{8+}. At the L(R) promoter, the per-complex abortive synthesis rate is similar at both NTP concentrations (0.015 s^{-1} vs 0.014 s^{-1}). The similarity in fold changes between k_{8+} and per-complex abortive rate at each NTP condition for both R(R) and L(R) indicates similarities in the initial stages of initial transcription for both classes of complex. For R(R), this likely indicates an initial equilibrium between a stable complex that cannot bind NTPs and an unstable nonproductive OC that is stabilized by iNTP binding. Speculatively, because all rounds of abortive synthesis are slow at the low UTP condition for the R(R) promoter, this

equilibrium state may be re-established every time an abortive RNA is released, and the nonproductive complex subsequently reinitiates. For L(R), the consistency of the nonproductive synthesis rate between NTP conditions is further evidence for either the absence of an equilibrium step analogous to that at R(R) that defines the overall initiation kinetics, or that GTP is required to stabilize the productive OC. Compared to L(L), both the initial amount of nonproductive complexes and the abortive RNA synthesis rate at both low UTP (rate: 0.014 s^{-1}, initial amount: 1 nonproductive complex/productive complex) and high UTP (rate: 0.019 s^{-1}, initial amount: 0.8 nonproductive/productive complex) are lower for both R(R) and L(R) (Table 2). This indicates a potential correlation between the stability of the OC and amount of abortive RNA synthesis.

Dinucleotide Primer Directed TSS Selection at R(R) and L(R)

As indicated, the rules of TSS selection for a given promoter and ITR sequence are still being established. It is clear that multiple factors are involved, including sequence characteristics of the region downstream of the -10 region that constitute the discriminator and ITR, as well as NTP concentrations. Addition of a very short RNA primer (2 or 3 bp) can shift the TSS. These short RNA are the products of abortive RNA synthesis in vivo and drive large shifts in the TSS profile of the transcriptome especially in the shift from exponential to stationary phase where NTP concentrations are reduced [27-29].

Dinucleotide CpA (complementary to positions 8 and 9) and CpC (complementary to positions 7 and 8) were included in primer extension reactions at the R(R) and L(R) promoters to determine how initiation affected the kinetics of initiation at these promoter constructs. Neither dinucleotide primer was incorporated into 100% of full-length transcripts produced from either promoter. At both R(R) and L(R), CpA is much more

likely to prime initiation (as a percentage of total detected transcripts) than CpC. A recent massively-parallel study of primed transcription initiation indicates that primer efficacy is linked to promoter and primer sequence. GpG and UpA are the most favored primer sequences, and $YNN_{TSS}W$ is the most likely nontemplate strand sequence to be primed, at the TSS and TSS-1 positions [27]. Both R(R) and L(R) have CCAT as the relevant sequence, which is consistent with the reported motif. This preference may therefore be the origin of the higher total percentage of transcripts primed by CpA compared to CpC reported for R(R) and L(R).

Interpreting the Effect of CpA Priming on Initiation Kinetics

Concentrations of the +1 and +2 NTP were found to have a much larger effect on initiation kinetics than NTPs incorporated at other positions for promoters like R(R) and L(L) because binding of both these NTPs stabilize an otherwise unstable open complex [20]. It is reasonable to assume that CpA fills a similar role in primed initiation from the R(R) promoter. The ability of CpA primed initiation at these promoters to bypass the second nucleotide effect under low iNTP conditions suggests an additional layer of regulation at the ribosomal promoters in vivo, which are very responsive to nutrient starvation via high dependence on the concentration of the first and second iNTPs.

A simple mechanism can be used to interpret CpA effects on the kinetics of long RNA synthesis at R(R) and L(R) promoters. Figure 5 summarizes values of the first order rate constant $k_{obs,\ 8+}$ for the first order approach of the long RNA population to its plateau value in single round synthesis. These observed rate constants can be interpreted as the product of the fraction f_{init} of the population of productive OC that are in the correct conformation to initiate synthesis of a long RNA and the intrinsic rate constant $k_{obs,\ 8+}$ long-

RNA synthesis by these complexes. As a starting point we assume that f_{init} = 1 at [CpA] = 80 μM. Comparison of $k_{obs,8+}$ values +/- CpA in figure 5C for each promoter and NTP condition indicates that f_{init} ≈ 1 for R(R) and L(R) at high U (200 μM UTP) but f_{init} ≈ 0.3 for R(R) and L(R) at low UTP (10 μM). Interpreted as UTP binding, the value f_{init} ≈ 0.3 at 10 μM UTP corresponds to $K_{D,\,UTP}$ = 23 μM. This $K_{D,\,UTP}$ predicts 90% occupancy of the initial complex by UTP at 200 μM UTP (the high U condition) and explains why values of $k_{obs,8+}$ values +/- CpA are the same within uncertainty at high U. Surprisingly, this $K_{D,\,UTP}$ is very similar to that used for binding of UTP to the initiation complex with bound ATP in fitting L(L) step-by-step initiation kinetic data (Chapter 2).

Conclusions

In this research we described the TSS selection and escape kinetics of the R(R) and L(R) hybrid promoters. The replacement of *rrnB P1* promoter elements with the λP_R ITR has a profound impact on TSS selection at the L(R) promoter compared with R(R) in a manner consistent with previous studies of TSS selection at the *rrnB P1* promoter. This indicates that certain upstream promoter elements, such as spacer length, are key components of normally unfavorable 8 bp discriminator. The distribution of TSS at these promoters changes in the presence of dinucleotide primers. Long RNA synthesis is more likely to be primed by CpA (positions 8 and 9) than CpC (positions 7 and 8), consistent with previous studies on the favorability of dinucleotide primers [27]. The 2nd nucleotide effect whereby initiation efficacy is highly dependent on the concentration of the 2nd initiating nucleotide (here, UTP) previously reported for the WT *rrnB P1* promoter is also apparent at the R(R) promoter [20]. Analysis of the length distribution of RNA synthesized at these promoters by nonproductive complexes indicates that promoter escape may not occur until after 7-mer synthesis for either construct, and that the first steps of initial

transcription are similar for the productive and nonproductive complexes. Finally, we determined that CpA can overcome the second-nucleotide effect at R(R) and L(R).

These findings raise several interesting questions about the nature of the complexes formed at unstable promoters, and how the initiation mechanisms at these promoters may differ from the well-described mechanisms at promoters that form more stable OCs, like L(L). Further kinetic studies of these promoters are warranted, especially in the time regime that will allow for detection of transient productive RNA species, as this will allow for more detailed information on the mechanisms of initiation and escape at these promoters. Furthermore, because R(R) and L(R) have the same ITR sequence as L(L), the kinetics of initial transcription at these promoters can be easily compared, making R(R) and L(R) as excellent models of promoters that form unstable open complexes (R(R)) and initiate from multiple TSS (L(R)).

Materials and Methods

Reagents, Buffers, and Gels

Reagents for buffers and stock solutions were the highest available grade and were used as received. All solutions were prepared using 18 MΩ deionized water from a Barnstead EPure system. NTPs and dNTPs (Boston Bioproducts, Thermo Fisher, New England Biolabs) used in transcription assays and PCR reactions were 99% pure and used as received. Enzymes for PCR reactions were purchased from NEB and used according to the manufacturer's protocols.

Storage buffer (SB) for core RNAP, σ^{70} and RNAP holoenzyme is 50% v/v glycerol, 0.01 M Tris (pH 8.0), 0.1 M NaCl, 0.1 mM EDTA, and 0.1 mM DTT, 0.05 mg/mL BSA. Transcription buffer (TB) is 40 mM Tris, 5 mM MgCl2, 30 mM KCl, 1 mM DTT, and 0.05 mg/mL BSA, adjusted to pH 8.0 at the experimental temperature (37 °C).

5X initiation solution for primer extension assays (IS_{PE}) is 1 mM all NTPs in TB. 5X initiation solution for manual mixing initial transcription assays (IS) with α-^{32}P-GTP ("High UTP") is 1 mM ATP and UTP, 50 μM GTP, and 87.5 nM α-^{32}P-GTP in TB. IS with α-^{32}P-UTP ("Low UTP") is 1 mM ATP and GTP, 50 μM UTP, and 87.5 nM α-^{32}P-UTP in TB.

Quench Solution (QS) for manual mixing transcription assays is 8 M urea and 15 mM EDTA in TB. QS with added dyes (QSD) for polyacrylamide gel electrophoresis (PAGE) has 0.05% xylene cyanol and 0.05 bromphenol blue in QS. TBE buffer for PAGE is 90 mM Tris-Borate (pH 8.3) and 2 mM Na_2EDTA. All gels for initial transcription assays are 20% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics). All gels for primer extension assays are 12% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics).

RNA Polymerase and Promoter DNA

The RNAP holoenzyme is overexpressed (as core RNAP and σ^{70} separately), purified, and reconstructed as described previously [15]. Filter binding activity assays performed on preparations of RNAP holoenzyme used here show that 50% ± 10% of RNAP molecules form a stable open complex with the λP_R promoter at 37 °C. All RNAP concentrations reported here refer to this active fraction. Sequences of PCR primers and promoter DNA fragments used to prepare R(R) and L(R) promoter DNA fragments (RRNB_ALL_F, RRNB_ALL_R_TXN, HBOT, HTOP), and primers for primer extension assays (DMP_PE23OL_LPRHBOT) are given in Table 3.

TSS Determination

RNAP (25 nM) was incubated with promoter DNA (12.5 nM) for 1 h at 37 °C. In assays with dinucleotide primers, 100 μM CpA or CpC was included in this incubation.

5X IS_{PE} is manually mixed 1:4 with the RNAP-promoter DNA solution to obtain a final promoter concentration of 10 nM. The reaction is quenched with QS 300 s after IS_{PE} addition. RNA product length was determined via primer extension with Avian Myeloblastosis Virus (AMV) Reverse Transcriptase (NEB) primed by a 5'-end-labelled 23 base primer that anneals 9 bp downstream of the predicted TSS (Table 3). Primer extension products were separated via PAGE and imaged via phosphorimager as previously described [17]. Primer extension products were quantified via ImageQuant software.

Single-Round Transcription Assays

RNAP (25 nM) was incubated with promoter DNA (12.5 nM) for 1 h at 37 °C. In assays with dinucleotide primers, this incubation was performed in the presence of 100 μM CpA or CpC. 5X IS is manually mixed 1:4 with the RNAP-promoter DNA solution to obtain a final promoter concentration of 10 nM at time zero. Reactions are quenched with QSD at the indicated timepoints (10 s – 480 s). Transcription gels are imaged via phosphorimager and quantified as previously described [17].

Figures

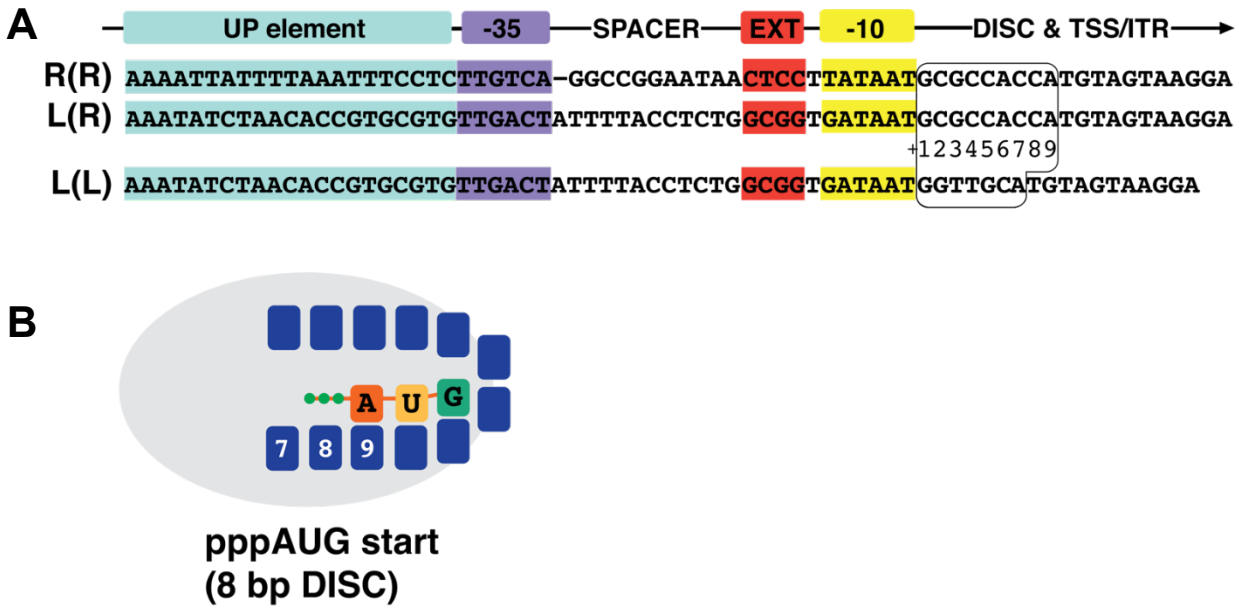
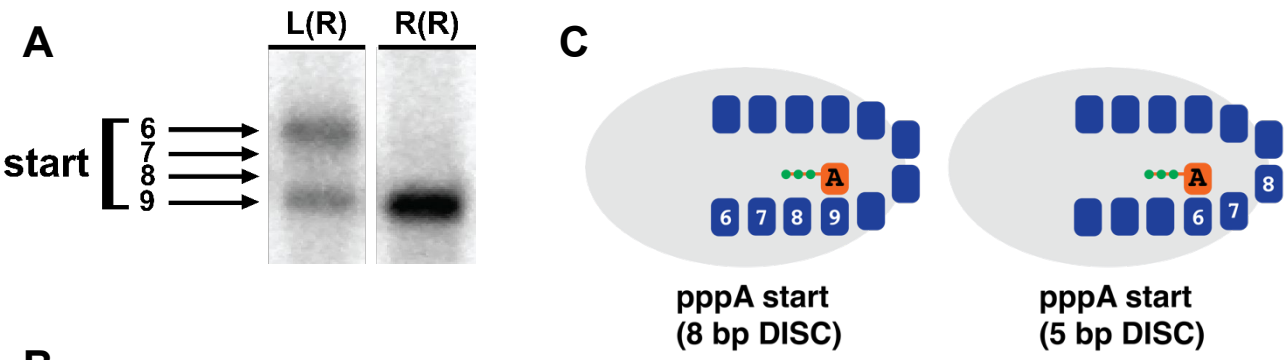


Figure 1. R(R) and L(R) promoter sequences. *(A)* Sequence and promoter elements comparing the previously-investigated λP_R promoter construct (abbreviated L(L)), to the investigated promoter constructs with the 8 base-pair *rrnB P1* discriminator and *rrnB P1* core promoter elements (abbreviated R(R)) or λP_R core elements (abbreviated L(R)). Important sequence elements are color-coded and labeled (UP element, -35, spacer, extended -10, -10, discriminator region (DISC) followed by the transcription start site (TSS) and initial transcribed region (ITR)). The sequence is numbered from +1 to +9 immediately following the -10 region and this numbering is referenced throughout the text to identify transcription start sites. The initial transcribed region (ITR) was engineered from the λP_R ITR and is identical for all constructs. *(B)* A simple cartoon shows a 3-mer synthesized from a +9 start (corresponding to an 8-base pair *rrnB P1* discriminator). Base pair numbering is consistent with (A).



B

	% Starting at +9 (A)	% Starting at +8 (C)	% Starting at +7 (C)	% Starting at +6 (A)
R(R)	>99%	<1%	<1%	<1%
L(R)	34(± 3)%	<1%	<1%	66(± 3)%

Figure 2. Primer extension reveals different start site distributions for R(R) and L(R).

(A) Transcription experiments were initiated at 37°C with 200 uM ATP, UTP, GTP and CTP and quenched after 300 seconds. Primer extension used α-^{32}P-labeled primer to identify FL RNA lengths produced in transcription experiments. Lengths were separated on 12% acrylamide gels. *(B)* Start site distributions were quantified and percentages are reported in (B). R(R) has a unique +1 start (corresponding to an A start and an 8 bp discriminator) with no detected initiation from -1(C), -2 (C) or -3 (A). L(R) has two major start sites: +1 (A), corresponding to an 8- bp discriminator and yielding 1/3 of FL transcripts, and -3 (A), corresponding to a 5 bp discriminator and yielding 2/3 of FL transcripts. No initiation is detected from -1 (C) or -2 (C). *(C)* A cartoon representation shows a +9 ATP start and a +6 ATP start for the R(R) and L(R) constructs.

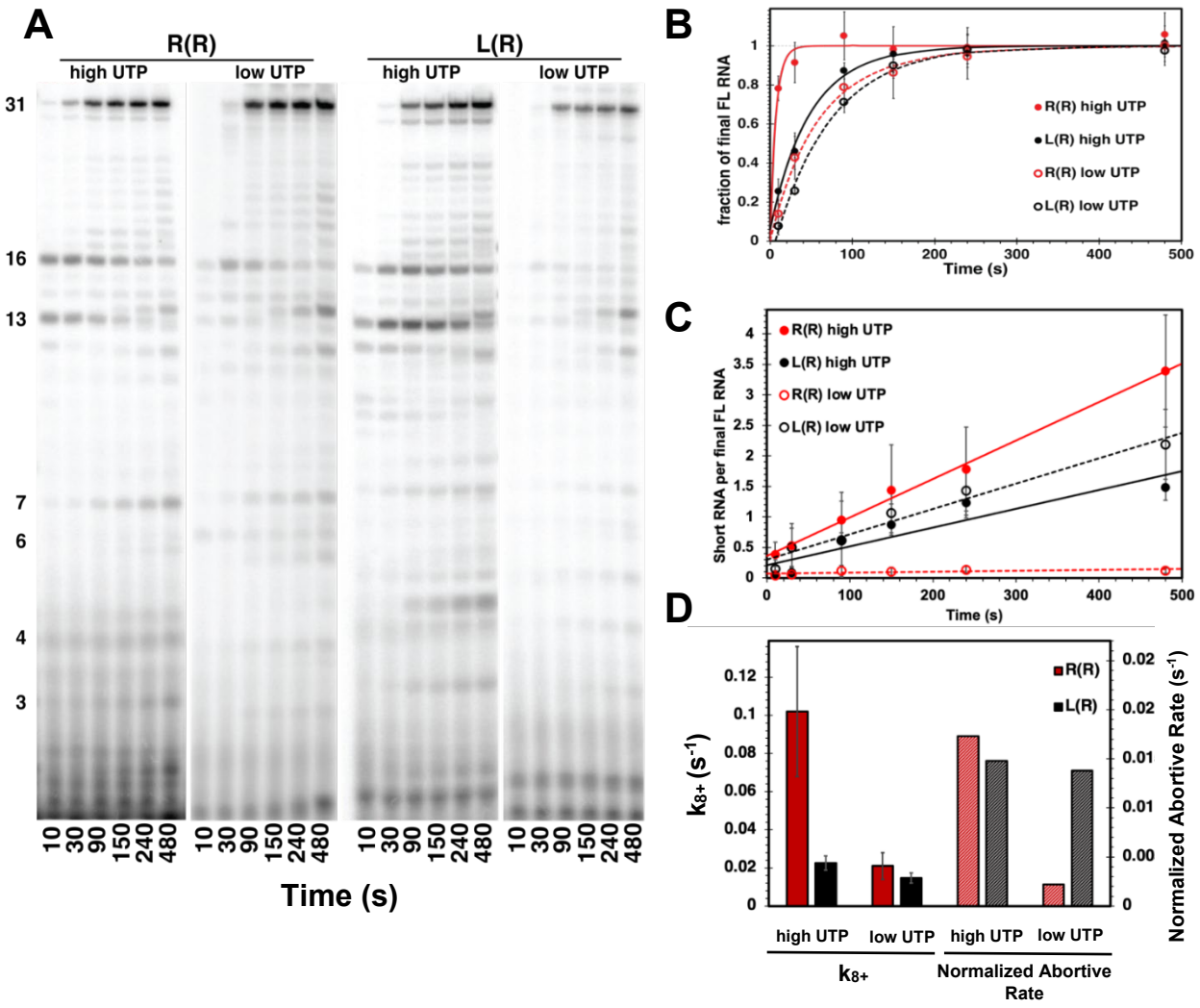
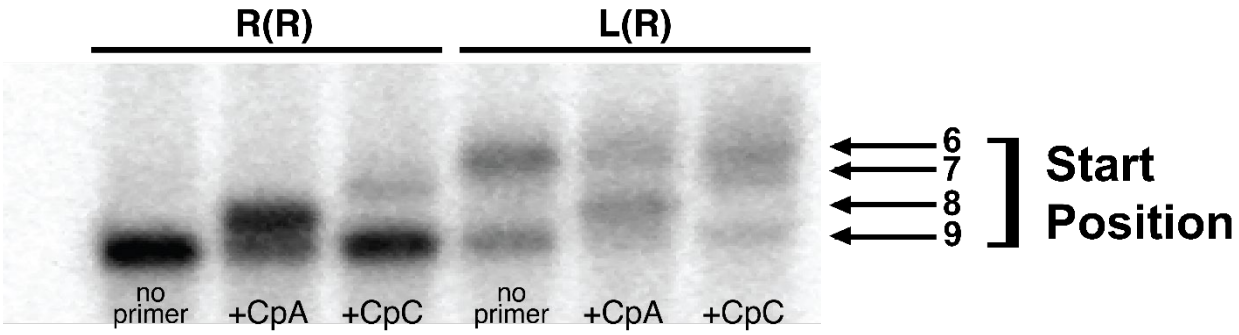


Figure 3. The kinetics of productive and nonproductive transcription initiation from R(R) and L(R). *(A)* Representative polyacrylamide gels with time courses (10-480 seconds) of α-^{32}P-labeled RNA synthesis are compared side-by-side. R(R) or L(R) DNA template was incubated with RNAP at 37°C and initiation solution (containing 200 μM: high UTP or 10 μM: low UTP) was added at time 0. Samples were quenched at designated timepoints and RNA products were separated by length using gel electrophoresis. *(B)* Long RNA (approximated as the sum of 13-, 16- and 31-mer RNA,

the major lengths accumulated post-escape) production over time is plotted as a fraction of final productive RNA for R(R) (red) and L(R) (black), averaged from 3-4 independent experiments. Experiments from each condition were fit with single exponential decay equations (see chapter 2) and average fits are plotted. Dashed lines and open circles refer to the low UTP condition, solid lines and closed circles refer to the high UTP condition. *(C)* Short RNA (<8-mer) is plotted over time for R(R) and L(R) with best-fit lines for each condition. *(D)* The full-length rate constants (k_{8+}) determined in the best-fit equations in *(B)* are graphed with error for both promoters at high and low UTP (solid bars); short RNA slopes from (C), normalized to short RNA intercepts from (C), are plotted for R(R) and L(R) at high and low UTP.

A



B

Start Site %	No primer	+CpA	+CpC	No primer	+CpA	+CpC
+9 (A)	**>99%**	**36%***	**85%***	**34(± 3)%**	**11%***	**19%***
+8 (C)	-	**64%***	-	-	**59%***	-
+7 (C)	-	-	**15%***	-	-	**39%***
+6 (A)	-	-	-	**66(± 3)%**	**30%***	**42%***

C

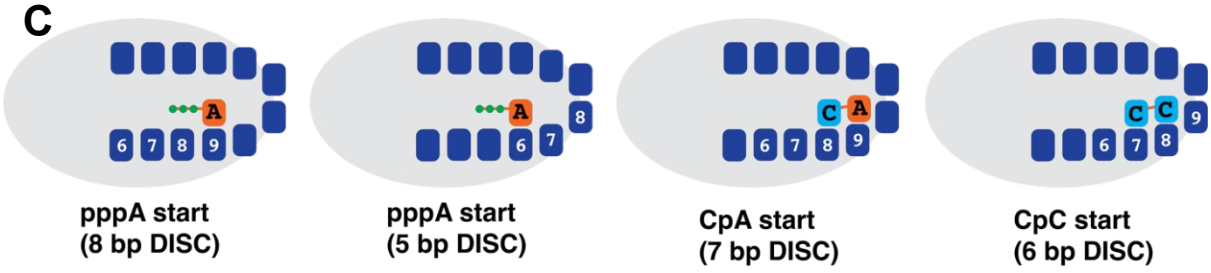


Figure 4. Start site selection is influenced by dinucleotide primers CpA and CpC. *(A)* Transcription experiments were initiated with the addition of all four NTPs (200 μM final) and CpA or CpC (80 μM final) where indicated and run at 37 °C for 300 seconds. Primer extension with α-^{32}P-labeled primer was followed by gel electrophoresis with a 12% acrylamide gel to separate products using length as an indicator of the selected start site. *(B)* Start site distributions for unprimed transcription from R(R) and L(R) are compared to start site distributions with the addition of CpA or CpC primer. * value is estimated to have ~10% error based on similar experiments *(C)* A cartoon representation

compares a possible CpA and CpC start in the active site to an ATP start at +9 and +6. The triphosphate of ATP is symbolized by three dots; CpA and CpC lack a 5’ phosphate.

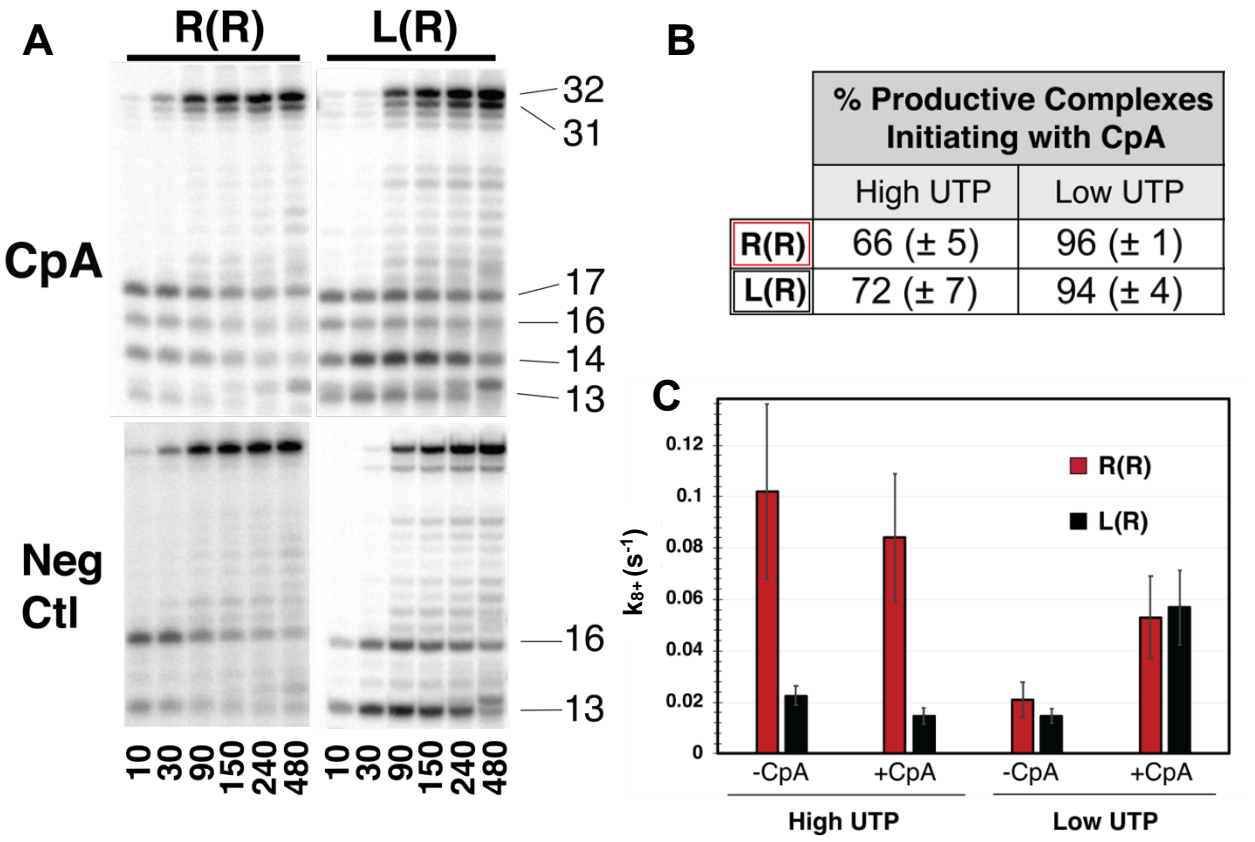


Figure 5. CpA primed transcription from R(R) and L(R). *(A)* Representative gels of long α-^{32}P-labeled RNA production over time compare high UTP transcription experiments without primer (bottom) to experiments with CpA (top). 100 μM dinucleotide primer (80 μM final) was incubated with promoter DNA and RNAP at 37 °C followed by the addition of initiation solution at time 0. CpA incorporation (+8, +9 start) resulted in long RNA accumulating at 14-, 17- and 32-mer. *(B)* Primed long RNA amounts were divided by total long RNA in each experiment, and the results are reported as a percent of productive complexes initiating with dinucleotide primer. *(C)* The long single exponential decay constants describing primed transcription initiation from R(R) (red) and L(R) (black) were determined with best-fit single exponential equations and are compared to unprimed

initiation rate constants at high UTP (200 μM ATP and GTP, 10 μM UTP) and low UTP (200 μM ATP and UTP, 10 μM GTP).

Table 1: Comparison of $t_{1/2}$ values between L(L), R(R), and L(R)

	L(L)[a]	R(R)	L(R)
High UTP	1.8 s	7 s	35 s
Low UTP	14 s	35 s	46 s

[a]Values were determined from RQF assays in chapter 2.

Table 2: Comparison of linear abortive slope and intercept between L(L), R(R), and L(R)

	L(L)[a]	R(R)	L(R)
High UTP Intercept (nonproductive/productive)	0.8	0.19	0.21
High UTP Slope (nonproductive RNA/productive RNA/s)	0.019	0.009	0.0031
Low UTP Intercept (nonproducive/productive)	1	0.07	0.27
Low UTP Slope (nonproductive RNA/productive RNA/s)	0.014	0.0002	0.0044

[a]Values were determined from RQF assays in chapter 2.

Table 3: Primers used for R(R) and L(R) template preparation

RRNB_ALL_F	*CCACGAATTC*GTCAGAAAATTATTTTAAATTTCCTCTTGTCAGGCCGGAATAACTCC
RRNB_ALL_R TXN	CCTTATTGAGGGATATTACGCGGTGGTACATCATTCCTCCAAGATACTTCCAAAACA
HBOT	TTCCAAAACAGCCACGTCCAC
HTOP	CCAGCATTCCTCCACGAATTC
DMP_PE23OL_LPR HBOT	CGACAAAAGCTTCAGGGAACCTC

References

[1] Bremer H, Dennis PP. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus. 2008;3.

[2] Gourse RL, Chen AY, Gopalkrishnan S, Sanchez-Vazquez P, Myers A, Ross W. Transcriptional Responses to ppGpp and DksA. Annual Review of Microbiology. 2018;72:163-84.

[3] Shin Y, Qayyum MZ, Pupov D, Esyunina D, Kulbachinskiy A, Murakami KS. Structural basis of ribosomal RNA transcription regulation. Nature Communications. 2021;12:528.

[4] Kriel A, Bittner AN, Kim SH, Liu K, Tehranchi AK, Zou WY, et al. Direct Regulation of GTP Homeostasis by (p)ppGpp: A Critical Component of Viability and Stress Resistance. Molecular cell. 2012;48:231-41.

[5] Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, et al. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science. 1993;262:1407-13.

[6] Winkelman JT, Chandrangsu P, Ross W, Gourse RL. Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of E. coli ribosomal RNA promoters. Proceedings of the National Academy of Sciences. 2016;113:E1787-E95.

[7] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[8] Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA promoter regulation by nonoptimal binding of s region 1.2: an additional recognition element for RNA polymerase. Cell. 2006;125:1069-82.

[9] Liu J, Turnbough CL. Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in Escherichia coli. Journal of Bacteriology. 1994;176:2938-45.

[10] Lewis DEA, Adhya S. Axiom of determining transcription start points by RNA polymerase in Escherichia coli. Molecular Microbiology. 2004;54:692-701.

[11] Yu L, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. eLife. 2017;6:e32038.

[12] Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, et al. Multiplexed protein-DNA crosslinking: scrunching in transcription start site selection. Science (New York, NY). 2016;351:1090-3.

[13] Ko J, Heyduk T. Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochemistry Journal. 2014;463:135-44.

[14] Winkelman JT, Gourse RL. Open complex DNA scrunching: A key to transcription start site selection and promoter escape. BioEssays. 2017;39:1600193.

[15] Plaskon DM, Henderson KL, Felth LC, Molzahn CM, Evensen C, Dyke S, et al. Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. Proceedings of the National Academy of Sciences. 2021;118:e2021941118.

[16] Henderson KL, Evensen CE, Molzahn CM, Felth LC, Dyke S, Liao G, et al. RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. Biochemistry. 2019;58:2339-52.

[17] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[18] Susa M, Kubori T, Shimamoto N. A pathway branching in transcription initiation in Escherichia coli. Molecular Microbiology. 2006;59:1807-17.

[19] Kubori T, Shimamoto N. A branched pathway in the early stages of transcription by *Escherichia coli* RNA polymerase. Journal of Molecular Biology. 1996;256:449-57.

[20] Lew CM, Gralla JD. Mechanism of Stimulation of Ribosomal Promoters by Binding of the +1 and +2 Nucleotides *. Journal of Biological Chemistry. 2004;279:19481-5.

[21] Gaal T. Transcription Regulation by Initiating NTP Concentration: rRNA Synthesis in Bacteria. Science. 1997;278:2092-7.

[22] Winkelman JT, Winkelman BT, Boyce J, Chen AY, Ross W, Gourse RL. Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. Molecular Cell. 2015;59:768-80.

[23] Libing Y, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. Elife. 2017;6.

[24] Barker MM, Gaal T, Josaitis CA, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. Journal of Molecular Biology. 2001;305:673-88.

[25] Pupov D, Petushkov I, Esyunina D, Murakami KS, Kulbachinskiy A. Region 3.2 of the σ factor controls the stability of rRNA promoter complexes and potentiates their repression by DksA. Nucleic Acids Research. 2018;46:11477-87.

[26] Li L, Molodtsov V, Lin W, Ebright RH, Zhang Y. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. Proceedings of the National Academy of Sciences. 2020:201920747.

[27] Skalenko KS, Li L, Zhang Y, Vvedenskaya IO, Winkelman JT, Cope AL, et al. Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in Escherichia coli. Proceedings of the National Academy of Sciences. 2021;118:e2106388118.

[28] Druzhinin SY, Tran NT, Skalenko KS, Goldman SR, Knoblauch JG, Dove SL, et al. A Conserved Pattern of Primer-Dependent Transcription Initiation in Escherichia coli and Vibrio cholerae Revealed by 5′ RNA-seq. 2015;11:e1005348.

[29] Nickels BE. A new way to start: nanoRNA-mediated priming of transcription initiation. Transcription. 2012;3:300-4.

Chapter 5: The Distribution of Discriminator Lengths in the *E. coli* Genome and the Effects of the 7 bp T7A1 Discriminator on Initial Transcription Kinetics

Experiments in this chapter were performed by Dylan Plaskon, Kate Henderson, Taka Ishikuri, Sarah Doughty, Quinn McBride, Will Langholz, Jack Prazich and Savannah Peterson. Analysis and interpretation was performed by Dylan Plaskon, Taka Ishikuri, Quinn McBride, and Tom Record.

Introduction

Promoter sequence elements provide a first level of regulation of all aspects of transcription initiation, but our understanding of the full range of these effects and the underlying mechanisms is incomplete. The discriminator region is defined as the sequence separating the -10 element and the transcription start site (TSS). The questions of when a TSS preference for a purine followed by a pyrimidine determines discriminator length vs. when a preferred discriminator length determines TSS selection, and the underlying energetics, are largely unresolved. As summarized and further analyzed in this chapter, discriminator regions of *E. coli* σ^{70} promoters lack a consensus sequence and range in length from 4 to 8 bp or more. These discriminator regions are found to play key roles in TSS selection, open complex (OC) stability, and escape point [1-4]. Despite the lack of a consensus sequence, sequence-specific contacts have been identified at the upstream end of the discriminator, in particular between a nontemplate strand G immediately downstream of the -10 element and $\sigma_{1.2}$ [5].

Promoter elements have evolved alongside RNAP to optimize initiation regulation. This regulation is largely mediated by RNAP-promoter interactions, in particular interactions of the -35 and -10 elements with $\sigma_{4.2}$ and $\sigma_{2.4}$ respectively [5-7]. When these contacts are in place and the initiation bubble has been opened to form the OC, the single-stranded discriminator DNA is positioned in the RNAP cleft between the -10 contacts and the active site, where the TSS must be properly positioned for NTP binding to occur [7]. Structural [7] and biochemical [2, 3, 8] studies of the OC reveal that 6-base lengths of the discriminator DNA strands span the distance between the -10 region of the promoter and the active site containing the transcription start site (TSS) without distortion of the DNA or cleft. This may explain why promoters with 6 bp discriminators form more stable OCs

compared to promoters with other discriminator lengths [1-3]. The consequences of OC destabilization introduced by longer or shorter discriminators on promoter efficacy are not fully understood, although correlations between OC instability and promoter efficacy has been reported [9, 10].

For discriminators longer than 6 bp to be accommodated within the RNAP cleft, the single stranded discriminator DNA must undergo a thermodynamically unfavorable distortion called “pre-scrunching”. For the rrnbp1 promoter (8 bp discriminator) the path of pre-scrunched DNA within the cleft has been mapped [3]. Pre-scrunching has been proposed to destabilize the OC by approximately 1 kcal/mol per pair of bases scrunched [1, 2]. The lifetime of the 37 °C λP_R promoter OC is reduced by at least 4 orders of magnitude (a >5 kcal destabilization) when the λP_R 6 bp discriminator (GGTTGC) is replaced by the 8 bp rrnbp1 discriminator (GCGCCACC), which introduces prescrunching [1] and removes the G at position 2 which interacts with sigma region 1.2 [5, 7, 11, 12]. OC destabilization may reduce the length of RNA that must be transcribed before promoter escape by lowering the energy required to break the upstream contacts of RNAP with the promoter [1]. The dramatic effects of discriminator length and sequence on OC stability and its proposed roles in initiation kinetics suggest that it plays an important regulatory role. However no systematic study of discriminator length and OC lifetime on initiation rate and escape kinetics had been performed.

The T7A1 promoter is a well-studied model promoter with a 7 bp discriminator. Two variants of this promoter are investigated here. Both use the modified λP_R ITR studied in previous chapters of this thesis. The promoter with all the upstream elements of T7A1 (discriminator, core promoter, UP element) and the λP_R ITR is referred to as here

as T(T). The λP_R variant in which the λP_R 6 bp discriminator is replaced by the 7 bp T7A1 discriminator is referred to as L(T). At 37 °C, the lifetime of L(L), the promoter with the λP_R promoter and discriminator, is ~13 hrs while that of T(T) is ~6 minutes. The hybrid L(T) promoter has an intermediate lifetime of about 0.5 hours [1]. In initiation from L(L) promoters, RNAP appears to escape (based on the length of the largest abortive product) in the step synthesizing 11-mer RNA, while RNAP escapes from both L(T) and T(T) (by the same criterion) in the step synthesizing 8-mer [1].

Here we use *in vitro* transcription assays to determine the time course of appearance and disappearance of RNA transients of each length on the path to escape at the T(T) and L(T) promoters. We use this information to determine the stepwise kinetics of initial transcription for the T(T) and L(T) promoter constructs at 19 °C, 25 °C, and 37 °C to compare with studies of L(L) in chapters 2 and 3. We interpret these kinetics in terms of the breaking of RNAP-promoter contacts necessary for promoter escape.

Experimental and computational studies are expanding the set of sequences known to be *E. coli* σ^{70} promoters. These are compiled in the RegulonDB database of >1800 confirmed and putative σ^{70} transcription start sites [13]. Here, we use a motif-prediction algorithm to locate the -10 region and thereby determine the most likely discriminator length of each member of this set of known and predicted *E. coli* transcription start sites. We determine correlations between discriminator length and promoter and ITR sequences. We also use *in vitro* transcription and primer extension assays to investigate start site selection for hybrid λP_R promoters constructed by replacing the 6 bp λP_R discriminator by a shorter (4 or 5 bp) discriminator from another promoter.

Results

Prediction of the Most Probable -10 Element from the Consensus -10 Element for E. coli σ^{70} Promoters

Multiple methods are available to determine the TSS site of an *E. coli* gene, including high-throughput transcriptomic analysis [13], low-throughput TSS mapping [4], and computational prediction [14]. Many such studies have been performed, compiled, then tabulated in the RegulonDB TSS database [13]. For *E. coli*, the database of σ^{70} promoters contains 1896 individual TSS. Additionally, genome coordinates and the surrounding sequence is documented as in figure 1A.

To predict discriminator length, the location of the -10 hexamer relative to the TSS must be known. The consensus sequence for the -10 element of σ^{70} promoters is 5'-TATAAT-3' [15]. Variations in the -10 element sequence are common despite the well-defined consensus sequence. Only ~7% (130/1896) of the σ^{70} promoters in the RegulonDB database have this consensus sequence within 14 bp of the TSS.

The most probable position of the -10 element at each promoter was determined by building a position weight matrix from total occurrences of the consensus -10 element sequence TATAAT and sequences with one difference from the consensus sequence (e.g. CATAAT) (Figure 1, Methods).

E. coli σ^{70} Promoter Discriminator Length is Usually 4 – 8 bp

Two subsets of the total promoter dataset were analyzed based on their designation as "strong" or "weak" confidence in the RegulonDB promoter set. TSS of promoters in the strong confidence dataset have been determined experimentally by high-

throughput transcriptomics and confirmed computationally. We include in our analysis of these promoters the set of promoters with the “confirmed” designation, indicating a TSS confirmed via direct experimental evidence other than high-throughput transcriptomics. The TSS of promoters in the weak confidence dataset are predicted computationally but have not been detected experimentally, and do not overlap with the strong confidence dataset.

To date, 798 strong confidence promoters and an additional 1063 weak confidence promoters for *E. coli* σ^{70} RNAP have been identified. From the predicted -10 hexamer and observed or predicted TSS, the distribution of discriminator lengths in Fig. 1D is predicted. The majority of promoters from both the strong and weak TSS sets are predicted to have a discriminator length between 4-8 bp, though some discriminators of each length from 0 to 14 bp are predicted based on predicted -10 regions from sequence considerations alone. Discriminator lengths for the strong confidence promoters are relatively evenly distributed around 6 bp. The distribution of discriminators for the weak confidence promoters is skewed towards 4 and 5 bp, with 5 bp being the most common discriminator length (Fig. 1D). Significantly more 4 bp discriminators are predicted for weak confidence promoters than strong confidence promoters.

Promoter Sequence Correlations with Discriminator Length

To determine correlations between promoter sequence and discriminator length, all promoter sequences with the same predicted discriminator length were aligned and the nucleotide distribution at each position was determined. These are compiled in figure 2. The most consistent sequence feature is the preference for purines in the +1 (TSS) position, observed for promoters with all discriminator lengths. However, details of this

purine preference vary with discriminator length. For example, A starts are preferred for all discriminator lengths except 5 bp, where a G start is preferred by a small margin. The ratio of purine to pyrimidine starts is largest for promoters with a 4 bp discriminator, next largest for promoters with a 7 bp discriminator, and comparably large for promoters with 5, 6- and 8-bp discriminators. The preference for pyrimidines at the adjacent position is also largest for promoters with 4 bp discriminators. For 5 bp discriminators, where the start site nontemplate strand base is most likely to be G, the next base is equally likely to be C, A or T, while for all other discriminator lengths where the start site base is most likely to be A, the rank order at the second position is T>C>A>G or T>A>C>G. For promoters with 7 and 8 bp discriminators, pyrimidines are favored at the two positions before the observed TSS (-1 and -2). However promoters with 4 and 5 bp discriminators exhibit no corresponding preference for pyrimidines at +2 and +3 respectively in the strong confidence set. Analysis of this sequence information, its significance, and its energetic consequences is continuing.

The overall promoter sequences are 60% AT vs 40% GC. This ratio is the same for the ITR. The -35 consensus sequence is not immediately apparent upstream of the -10 because of variations in spacer length, which average the likelihood of the -35 hexamer occurring over multiple positions relative to the -10 element.

Roles of Discriminator Length and Sequence in TSS Selection

Given a well-defined -10 region, what aspects of downstream promoter sequence and other information dictate TSS and resulting discriminator length between the -10 region and TSS? To examine the role of downstream (ITR) sequence in determining TSS and discriminator length in context of the λP_R promoter, fourteen sequences of 4 bp

discriminator regions and twenty one sequences of 5 bp discriminator regions were chosen from figure 2 to replace the 6 bp λP_R discriminator (Figure 3A) [16, 17]. The sequences chosen were discriminators of promoters that either had the “confirmed” confidence designation or occurred at multiple promoters in the “strong” confidence designation data set.

Primer extension assays at 200 μM NTPs were used to determine the TSS of the hybrid promoters, with λP_R as a comparison (Figure 3B). None of these thirty five hybrid promoters initiated transcription exclusively with pppApU (the +1 position predicted from their original context), as would be expected if the discriminator alone determined the TSS (Figure 3C). Instead all fourteen λP_R variants constructed with a 4 bp discriminator sequence and the LPR ITR initiated exclusively at +3G, resulting in a 6 bp discriminator. Most of the 21 promoters with a predicted 5 bp discriminator exhibited multiple TSS. A wide range of behaviors is observed (Fig. 3C) where up to 75% of transcripts used the +1A as the TSS and so preserved a 5 bp discriminator (Figure 3C), while the remaining transcripts starting at +3G resulting in a 7 bp discriminator. This indicates that discriminator sequence alone is insufficient to determine the TSS for short (< 6 bp) discriminators.

We also examined TSS selection for the gcdp1 promoter (5 bp discriminator) in the context of its WT gcdp1 ITR using primer extension. Comparison between the primer extension products from the native GcdP1 promoter and hybrid promoters with the GcdP1 discriminator and λP_R ITR are shown in figure 3D, and show that at these hybrid promoters GcdP1 initiates from both the +1 and +3 positions. Mutation of the +3G to T

(Fig 3D center lane), allowing purine start only with a 5 bp (or 9 bp) discriminator, causes initiation to occur from many different start sites, with no clear preferred site.

Kinetics of Long (Post-Escape) RNA Synthesis from Promoters with a 7 bp Discriminator

Fast mixing transcription assays were performed on a timescale ≤ 0.5 s to ≥ 90 s for both the T(T) and L(T) promoter constructs (Figure 4A and B) at two different NTP conditions used elsewhere in this thesis for quantitative initiation kinetic assays (Low UTP: 200 μM ATP and GTP, 10 μM UTP, 17.5 nM α-^{32}P-UTP; High UTP: 200 μM ATP and UTP, 10 μM GTP, 17.5 nM α-^{32}P-GTP). Assays with both T(T) and L(T) constructs and both NTP conditions were carried out at 19 °C, 25 °C, and 37 °C. Qualitatively the initiation kinetics observed for T(T) and L(T) constructs with 7 bp discriminators are similar to that reported previously for L(L) with a 6 bp discriminator, including two types of OC and two kinetic phases of initiation. In particular, both T(T) and L(T) promoters exhibit detectable transient short RNA buildup on the pathway to productive RNA synthesis at 19 °C and 25 °C, with little or no detectable transient RNA synthesis at 37 °C, especially at low U. For both constructs, the total amount of long RNA synthesis (the sum of all post-escape (≥ 8-mer) RNA lengths, designated 8+ RNA) by productive complexes at any temperature can be modelled as a single-exponential decay to a plateau value after a short lag (Figure 4C). Rate constants k_{8+} for these fits are plotted in Figure 4D.

Both T(T) and L(T), like L(L), show a monotonic increase in k_{8+} with temperature at the high UTP condition. In the low UTP condition, again like L(L), both T(T) and L(T) exhibit an increase in k_{8+} from 19 °C to 25 °C, and a dramatic decrease in k_{8+} between 25 °C and 37 °C, resulting in the relationship $k_{8+,\ 37\ °C} < k_{8+,\ 19\ °C} < k_{8+,\ 25\ °C}$ at the low UTP condition for L(L), L(T), and T(T) (Figure 4C and D).

The half times $t_{1/2} = 0.693/k_{8+}$ for T(T) and L(T) are compared with L(L) in Table 1, and illustrate the dramatic slowdown of initiation at the low UTP condition. Briefly, compared to L(L) $t_{1/2}$ are higher at 37 °C at both NTP conditions for T(T) and at the low UTP condition for L(T). At 25 °C for T(T) $t_{1/2}$ is lower at low U and higher at high U compared to L(L), and for L(T) $t_{1/2}$ is lower at low U and lower at high U. At 19 °C $t_{1/2}$ at T(T) for low U is higher than L(L) and L(T) and for high U is lower than for L(L) and higher than L(T).

Individual time-courses of 3-mer to 10-mer RNA amounts in initiation at T(T) and L(T) promoters at both high UTP and low UTP conditions and at 19 °C, 25 °C, and 37 °C are plotted in figures 5-16. Stepwise synthesis of long RNA is detectable at 19 °C and 25 °C by observing the transient buildup and decay of short RNA products within the first 10 s of initiation. At 37 °C, there are not enough detectable transient intermediates on the pathway to long RNA synthesis for T(T) or L(T) to determine stepwise initiation constants.

Both productive and nonproductive OCs synthesize RNA of length 3 to 7 bp at the T(T) and L(T) promoters. For each length of RNA and each time, it is necessary to separate the total amount of RNA and determine amount that is a transient in long RNA synthesis by productive OC. To accomplish this, the initial phase of RNA synthesis by nonproductive complexes is modeled as a single-exponential increase to a plateau. For each length of RNA and each time the amount of RNA from productive complexes is determined by subtracting that synthesized by the nonproductive fraction from the total RNA as previously described [16, 17] . T(T) and L(T) show overall similar, but not identical, trends for pre-escape RNA time courses. Significant transients are detectable at 3-mer and 6-mer at the low UTP condition for both constructs at 19 °C, because the

next base incorporated is U, and fewer transients are observed at the high UTP condition than high UTP condition for L(L).

T(T) Stepwise Rate Constants

Stepwise T(T) initial transcription rate constants k_i were determined by fitting the transient productive RNA buildup to the initiation mechanism originally determined for the λP_R promoter. Fits were performed using Kintek Explorer [18]. Kinetic parameters for ATP binding, UTP binding, and 2-mer catalysis used in the L(L) fits were retained here [16, 17]. An initial conformational change in the OC prior to nucleotide binding, analogous to that used to fit initial transcription transients at the L(L) promoter [16] was included in order to accurately reproduce the experimental data. At 19 °C good fits, which accurately reproduce the experimental data were acquired (Figure 17A). At 25 °C, the best fit accurately reproduced the kinetics of 11+ RNA production and the behavior of the most significant short transient species but did not fully reproduce the behavior of less significant transient species (Figure 17B). There are insufficient amounts of any transient intermediates in long RNA synthesis at 37 °C, so stepwise rate constants could not be determined.

At 19 °C all but one overall 2nd order forward rate constant k_i^+ (k_{cat}/K_m analog) for RNA extension steps up to and including the putative escape point at 8-mer formation are relatively small (< 0.1 μM^{-1} s^{-1}) in comparison to elongation rate constants (> 0.6 μM^{-1} s^{-1}). The exception is 6-mer, for which the rate constant approaches that of an elongation step (Figure 17C). In addition to 6-mer, steps forming 9-mer and 10-mer also are close to the lower end of the range of elongation rate constants, while 11-mer is smaller, which as

discussed previously may reflect displacement of σ^{70} or separation of the upstream RNA into the exit channel in this step.

For most steps, rate constants k_i^+ are larger at 25 °C than at to 19 °C, as is normally the case and which indicates a positive activation energy for the overall reaction. Exceptions are steps forming 6-mer (pre-escape) and 10-mer and 11-mer (post-escape). The k_i^+ value for 6-mer formation is smaller at 25 °C than at 19 °C, indicating a negative activation energy. For L(L), in chapter 2, a negative activation energy was observed for step 9 and attributed to the final stages of collapse of the upstream initiation bubble and duplex formation prior to escape. This is a possible explanation of the behavior of step 6 for T(T), since this step is positioned similarly relative to the escape point. Values of k_i^+ for post-escape steps 10 and 11 have no significant temperature dependence (no detectable activation energy). These steps may involve σ^{70} displacement or other late events of the transition from initiation to elongation and warrant further study.

An initial unfavorable conformational change (RP_O to I_3) needed in an analysis of the kinetics of initiation at the λP_R promoter is also necessary to fit the T(T) kinetic data. At 19 °C, $K_{eq,1a} = 0.33$. At 25 °C, $K_{eq,\ 1a} = 1.5$. $K_{eq,1a}$ could not be determined at 37 °C due to the lack of detectable transient productive intermediate RNA. Because the RNAP-T(T) complex exhibits a closed complex to open complex transition near 20 °C [19], a likely interpretation of these $K_{eq,1a}$ values is in terms of this closed to open equilibrium. A factor of 5 increase in $K_{eq,1a}$ in 5 ºC is a van't Hoff enthalpy change for this conformational change of about 57 kcal, which is the expected enthalpy change for melting 12 bp at this temperature.

Discussion

E. coli σ^{70} Discriminators are 4 – 8 bp Long

Structural and biochemical studies of the σ^{70} open complex have determined that 6 bp is the optimal discriminator length to bridge between the -10 region and the TSS in the RNAP cleft without distortion [6, 7]. Initial transcription of the first bases of the ITR, which occur with the upstream RNAP-promoter contacts in place, require scrunching additional DNA into the cleft to properly position the correct template base in the active site [20, 21]. Discriminators longer than 6 bp must therefore fit in the active site by scrunching of the in-cleft discriminator DNA prior to nucleotide addition (pre-scrunching) [3].

Discriminators of length 5 bp have been included in some studies of TSS selection. At the rrnbp1 promoter, mutation of some discriminator bases favors initiation from the -3A, the equivalent of a 5 bp discriminator [3]. Based on the efficacy of initiation when primed to begin at different positions at the LacCons promoter, a 5 bp discriminator (requiring stretching of the discriminator strands to position the TSS in the active site) is predicted to be approximately as unfavorable as an 8 bp discriminator when compared to a 6 bp discriminator baseline [2].

The consensus-sequence based prediction of the -10 element at all detected and predicted σ^{70} TSS predicts a discriminator length distribution between 4 bp and 8 bp, agreeing well with recent reports of discriminator length distribution in *E. coli* [22, 23]. The heavy skew towards shorter discriminator lengths in the weak confidence promoter set suggests either that promoters with short discriminators are more likely to be poor initiators, and thus evade detection in transcriptomic studies. It is also possible that not

all predicted TSS in the weak confidence dataset are true promoters, and instead may act as non-initiating RNAP binding sites.

Pre-scrunching is a well-defined process for proper active site positioning in long (7 bp – 8 bp) discriminators, but the opposite process for short discriminators is not well understood. Proper positioning of the active site for short discriminators may require "stretching" (also referred to as anti-scrunching [24]), but this phenomenon has not yet been directly observed. Although it is predicted to be approximately twice as unfavorable to stretch one base than to scrunch one base, the distribution of discriminator lengths in the strong confidence dataset definitely does not reflect this.

Promoter Sequence Correlations with Discriminator Length

Figure 2 shows sequence probabilities from the -10 element alignments of promoter and ITR sequences for the subset of promoters with each predicted discriminator length between 4 bp and 8 bp. Both the promoter and ITR sequences are AT rich (~60%). This contrasts with the overall *E. coli* genome, which is ~50.8% GC [25]. This cannot be explained just by the presence of the AT-rich -10, -35, and UP elements as almost every position shows at least a small preference for A and T.

For promoters with long discriminators, pyrimidines are preferred at the -1 position (for 7 bp discriminators) and the -2 position (for 8 bp discriminators). This is in good agreement with previous studies of *E. coli* TSS selection [3, 4]. Nontemplate strand pyrimidines at these positions disfavor transcription starting from the position that would result in a 6 bp discriminator and thus require no distortion, thus favoring transcription from the annotated start site.

At promoters with short discriminators, this same effect would predict nontemplate strand pyrimidines being favored at the +2 position for promoters with a 5 bp discriminator and at the +3 position for promoters with a 4 bp discriminator, and a slight preference for pyrimidines at these positions relative to other promoter positions is indeed seen. This preference is less than the corresponding preference for long discriminators. This is unexpected due to the predicted increased energetic cost of stretching a discriminator to accommodate the correct start site compared to scrunching. A thorough investigation of transcription start site selection at promoters with short discriminators is necessary to understand the mechanism of TSS selection at these promoters. One possibility is that promoters with predicted short discriminators are less likely to start from a single start site, despite being reported and annotated as such, or that stretching is stabilized in other ways, like distortions near at the -10-RNAP contacts.

No further significant sequence correlations with discriminator length were observed. Additionally, no clear link between gene functionality and discriminator length was observed.

Discriminator Length and Sequence is not the Sole Determinant of TSS Selection

None of 40 chimeric promoters containing the λP_R promoter and ITR and a short discriminator chosen from the strong confidence promoter set started exclusively from the predicted +1A TSS. Chimeric promoters formed by swapping the λP_R discriminator with a predicted 4 base discriminator started only from the +3G (6 bp discriminator). Chimeric promoters formed by swapping the λP_R discriminator with a predicted 5 base discriminator started at either the +1A (5 bp discriminator) or the +3G (7 bp discriminator) and ratio of +1A starts to +3G starts varied greatly from 0% to 75%, showing the influence of

discriminator sequence on TSS selection. Primer extension analysis of the WT GcdP1 promoter, which has a 5 bp discriminator and an initial ITR sequence of GCUA, confirmed that the annotated TSS was the only start site detected for the WT promoter. hese results clearly demonstrate that discriminator length and ITR sequence by themselves are insufficient to determine the TSS distribution, and that other promoter features contribute as well [1]. An example of this effect at a promoter with a long discriminator is at the *rrnB* P1 promoter, where changes in sequence of the discriminator or upstream and changes in spacer length between -10 and -35 elements influences TSS selection [3]. Although unsurprising, this result demonstrates the difficulty in determining the effects of short discriminators on TSS selection and distribution.

Swapping in the T7A1 Discriminator Greatly Changes the Kinetics of Escape at the λP_R Promoter

In a previous chapter, effects of temperature on the promoter escape kinetics at the model λP_R promoter was determined [16]. These results are reproduced in Fig 4C of this chapter. At limiting concentrations of the second iNTP (UTP), the first order rate constant k_{8+} for the approach to a plateau level of long (post-escape) RNA synthesized in a single round by productive OC increases as expected from 19 °C to 25 °C but then decreases and is smaller at 37 °C than at 25 °C and 19 °C. This initially counterintuitive result is one indication that the very stable 37 °C λP_R OC (RP_O) is not the initiation complex. RP_O was known to be destabilized by lower temperature, converting to the intermediate OC designated I_3. We therefore concluded that the λP_R OC that initiates is the I_3 OC intermediate or an I_3-like OC, and that this is the primary conformation of the λP_R OC at 19 °C [16].

Stability and downstream interactions of the T7A1 (abbreviated T(T)) OC at 37 °C are deduced to be similar to the I_3 intermediate formed at the λP_R promoter [16, 26]. Escape of RNAP from T(T) is thought to occur at the 7-mer to 8-mer step, with overall rate constant k_{8+}. We initially hypothesized that k_{8+} for long RNA synthesis at T(T) would increase monotonically with temperature and be greater at 37 °C than at 25 °C. However, k_{8+} for T(T) initiation is much smaller at 37 °C than at 25 °C and 19 °C at the low UTP condition. For comparison, between 25 °C and 37 °C k_{11+} at low U decreases nearly 2-fold at the λPR promoter. Under the same condition, k_{8+} at low U decreases more than 10-fold at the T(T) promoter

The promoter construct referred to here as L(T), with the 6 bp λP_R discriminator replaced with the 7 bp T7A1 discriminator, forms an OC with similar lifetime to T(T) [1]. Both are much shorter-lived and less stable OC than λP_R at 37 °C. Temperature effects on k_{8+} for long RNA synthesis at L(T) resemble those for the T(T) promoter, decreasing 5-fold between 25 °C and 37 °C at the low UTP condition. At the high UTP (low GTP) condition, for both the T(T) and L(T) promoters increases monotonically with temperature, although ln(k_{8+}) does not follow the expected linear Arrhenius relationship with 1/T.

Mechanistic details cannot be extracted from trends in k_{8+} alone because it is a composite of many steps that may change in different ways with increasing temperature. For L(L), step-by-step rate constants k_i^+ for all but one step of initiation to the escape point increase with increasing temperature between 19 °C and 37 °C. Therefore overall rate constants like k_{8+} for T(T) and L(T) and $k_{11+}$for L(L) would be expected to exhibit a monotonic temperature dependence unless there are additional steps like pre-initiation conformational changes that are necessary to bind the iNTP. One such conformational

change, identified below as a likely factor in initiation at 19 °C and 25 °C for T(T) and L(T), is formation of the open complex from the more stable closed complex. This change is favored by higher temperature, and so cannot explain the behavior of k_{8+} for T(T) and L(T) at 37 °C. The similarity in magnitude of the temperature effect on k_{8+} for T(T) and L(T) suggest that the yet-unidentified conformational change responsible for this effect is similar for each promoter.

Interpretation of Variations in Stepwise Initiation Rate Constants at the T(T) Promoter at 19 °C in Terms of Translocation Stresses

At the λP_R promoter (L(L)), the three cycles in the periodic pattern of large and small rate constants on the pathway to 11-mer RNA formation and concurrent promoter escape are interpreted as stepwise disruption of contacts of RNAP with the three regions of the promoter DNA (Chapter 2, [17]). Contacts in the cleft with the discriminator strands are deduced to be disrupted first, primarily in forming 4-mer and 5-mer RNA. Stronger contacts with the -10 region are broken subsequently, coupled to synthesis of 7-mer, 8-mer and 9-mer RNA. Contacts with the -35 duplex are broken last, in synthesis of 11-mer RNA and accompanying escape of RNAP from the L(L) promoter.

Transients in long RNA synthesis by productive complexes at the T(T) promoter were fit to the same stepwise initial transcription mechanism as previously described (Chapter 2, Figure 3) [16]. Briefly, an initial reversible conformational change is followed by reversible ATP binding and reversible UTP binding, followed by irreversible catalysis to synthesize pppApU. Each subsequent step of RNA-DNA hybrid extension includes reversible translocation, reversible NTP binding and catalysis. Catalysis, and thus the entire composite RNA extension step, is functionally irreversible in the absence of added

PPi. The kinetics of each step of hybrid extension are quantified by a composite 2nd order rate constant ki (a k_{cat}/K_m analog) that includes equilibrium constants for translocation and NTP binding and the catalytic rate constant.

For T(T) at 19 °C, k_i values divide into two groups which we designate as small (< 0.05 μM^{-1} s^{-1}) and large (> 0.05 μM^{-1} s^{-1}). Steps 3 – 5 and 7 – 8 have small k_i values, and steps 6 and 9 – 11 have large k_i values (Table 2). Small k_i values steps are between 0.01 μM^{-1} s^{-1} and 0.02 μM^{-1} s^{-1}, similar to the smallest k_i values reported for initiation at λP_R (4-mer synthesis, k_4 = 0.012 μM^{-1} s^{-1}). These small and large rate constants show no correlation with the identity of the base incorporated, as is the case for initiation at L(L) and are reasonably interpreted in terms of the difficulty of translocation at each extension step, as for L(L). Steps with small k_i values therefore have more translocation stress. We propose that the main contribution to differences in k_i values for different initiation steps is from disruption of RNAP-promoter contacts required for RNAP escape. Other possible contributions to translocation stress are from opening a downstream base pair (if not compensated by closing an upstream base pair), scrunching of discriminator DNA or the hybrid, and steric clashes as the hybrid translocates into the cleft.

Because these translocation stresses are largely absent from elongation, and because the end of the abortive product distribution from nonproductive complexes that are incapable of escape is at 7-mer, we propose that the larger k_i values for steps synthesizing 9-mer, 10-mer and 11-mer RNA indicate that RNAP has broken its specific contacts with the promoter prior to these steps. Steps synthesizing 7-mer and 8-mer RNA have small k_i values indicating these are pre-escape steps and that RNAP escapes from

the T(T) promoter after 8-mer synthesis. This escape length is 1 bp longer than previously deduced based on the absence of nonproductive complexes longer than 7-mer.

All RNAP-promoter contacts must be broken before or during promoter escape. At the λP_R promoter the three cycles of large and small rate constants indicates that specific RNAP-promoter contacts with three regions are broken sequentially over the course of initiation. Rate constants for 3-mer, 6-mer and 10-mer synthesis are larger than those for synthesis of 4-mer and 5-mer, for synthesis of 7-mer to 9-mer, and for synthesis of 11-mer. These larger rate constants indicate steps with less translocation stress, which we interpret to mean that no contacts are broken in these steps. We infer that the only significant translocation stress in these steps is scrunching of the discriminator strands, building up stress toward breaking the next set of RNAP-promoter contacts.

The two-cycle pattern of k_i values for T(T) is intriguingly different from that of L(L). The two have in common the large k_i value for 6-mer synthesis, though this appears coincidental as discussed below. For T(T) the small 3-mer k_i value indicates that the first translocation step (before 3-mer synthesis) is difficult. This may be because the discriminator region of T(T) is prescrunched and because in-cleft contacts of RNAP with this discriminator (upstream end sequence TA) are weaker than for L(L) (upstream end sequence GG). Because of prescrunching and weak in-cleft interactions for T(T), it seems likely that disruption of -10 contacts begins in the step making 3-mer and continues in synthesis of 4-mer and 5-mer, in contrast to LPR where -10 contacts are disrupted in synthesis of 7-mer to 9-mer. For T(T) we deduce that contacts with the -35 region and UP element are disrupted in synthesis of 7-mer and 8-mer, leading to escape of RNAP and larger k_i values for 9-mer and subsequent steps. For T(T), disruption of these upstream

contacts with the promoter duplex occurs in two steps while for L(L) only one step is involved. This is probably because the T7A1 UP element is much closer to consensus than for LPR, making contacts of the UP element with the alpha CTD more difficult to break for T(T) than for L(L).

Stepwise Initiation Rate Constants at the T(T) Promoter at Higher Temperatures

Values of k_i for steps of initiation at the T(T) promoter construct at 25 °C are less well determined than at 19 °C. The time-courses of the most significant detected transients at both NTP condition are well-replicated by the fit. Less significant transients of length 8, 9, and 10 are not as well replicated, although the total amount of these RNA is small. Values of k_i for steps 3, 4, 7, and 8 increase compared to their 19 °C counterparts, indicating positive activation energies. Values of k_i for steps 6 and 9 are the two largest rate constants, although k_6 is smaller than at 19 °C.

The rate constant k_5 is surprisingly larger at 25 °C than at 19 °C, and similar in magnitude to k_6 which is large at both temperatures. One possible explanation for this is release of in-cleft and -10 element interactions prior to 5-mer synthesis at 25 °C. This would indicate that processes (like contact disruption) in some steps in the mechanism T(T) promoter are not the same at the two temperatures investigated. This is not observed at the λP_R promoter, where the pattern of large and small k_i^+ is consistent at 19 °C, 25 °C, and 37 °C despite an overall increase in k_i^+ at most steps.

Curiously, k_{10} and k_{11} are the same within uncertainty at 25 °C and 19 °C. If these rate constants are well determined, this either indicates temperature-independent elongation steps, or the effect of processes that become more favorable with increasing

temperature and processes that become less favorable with increasing temperature (such as bubble collapse, although this is a hallmark of initiation and not elongation) coincidentally cancel each other between 19 °C and 25 °C.

Caution should be taken when interpreting the 25 °C k_i values. The parameters reported here do not accurately reproduce length 8 bp - 10 bp transient RNA intermediates and the unexpected temperature dependence of k_5, k_{10}, and k_{11} indicate these parameters may not fully describe initial transcription at 25 °C for T(T). Experiments at additional NTP conditions could be used to determine k_i at 25 °C with better confidence. Nonetheless, these parameters provide a useful starting point for the investigation of temperature effects on initiation at the T(T) model promoter.

No transient short RNAs are detected at 37 °C at the T(T) promoter, preventing any k_i analysis. It is likely at 37 °C that this is the result of a rate-limiting step at the beginning of the initiation mechanism at 37 °C, which precludes transient detection. Consistent with this, values of k_{8+} for synthesis of the long (post-escape) RNA at 37 °C are smaller than expected from extrapolation of 19 °C and 25 °C results for the high U condition and especially for the low UTP condition. At the λP_R promoter this reduction in rate of synthesis of the long (post-escape) RNA at 37 °C results from the thermodynamically unfavorable conformational change of the 37 °C stable OC (RP_O) to an initiation competent complex (IC, thought to be I_3-like) prior to, and necessary for, binding of the first NTP. With a reduction in temperature, a progressively greater fraction of the population of binary OC are IC, reducing the fraction of the OC population that need to convert to IC to initiate.

At both 19 °C and 25 °C a different initial conformational change described by equilibrium constant $K_{eq,\ 1a}$ must be introduced to accurately fit the T(T) transient short RNA. For the T(T) promoter at 19 °C this initial equilibrium is heavily biased towards a complex that cannot initiate ($K_{eq,\ 1a}$ = 0.33) but at 25 °C has shifted towards the initiation competent OC ($K_{eq,\ 1a}$ = 1.5). This temperature dependence of $K_{eq,\ 1a}$ for T(T) initiation is in the opposite direction to that discussed in chapter 2 for L(L). For T(T) this initial conformation change with a van't Hoff enthalpy change of approximately 57 kcal is almost certainly the opening of the initiation bubble to convert a closed promoter complex to the initiation-competent OC. The enthalpy of base pair opening in solution at 25 °C is 5-6 kcal/mol [27], so the measured van't Hoff enthalpy indicates cooperative opening of approximately 10-12 bp, the size of the initiation bubble. Previous studies found this opening transition for T7A1 occurs near 20 °C, and that the entire population of RNAP-T7A1 promoter complexes is OC above 30 °C [28, 29]. Hence this initial conformational change cannot be involved in the behavior of T(T) in initiation at 37 °C.

Stepwise Initiation Rate Constants k_i at the L(T) Hybrid Promoter are Not Well Determined

Although significant RNA transients are detected at 19 °C and 25 °C in RQF experiments at the L(T) hybrid promoter (Figures 11-16), no set of k_i parameters has yet been determined at either temperature that satisfactorily reproduce these transients. Removing one or more transients from the analysis did not improve the fits as would be expected if inaccurate quantification of that transient was the reason for this inability to accurately fit the data. Further analysis of these extensive L(T) data sets is planned. Further experiments, including additional NTP concentrations, may also be needed to determine the minimal mechanism for L(T).

Conclusions and Remaining Questions

The significant amount of short discriminators predicted in the *E. coli* σ^{70} genome offers exciting insight into an important initiation regulatory mechanism, and many questions as to the nature and purpose of promoters with short discriminators remain. The effect of these discriminators on OC stability is of particular interest, due to the large apparent consequences of OC stability on initiation characteristics like escape length. However, these promoters must be tested in their native context, as the discriminator alone appears insufficient to determine TSS selection. As the effects of the discriminator cannot therefore be isolated by studying them in the context of a known ITR and upstream promoter, multiple promoters should be tested to determine if OC instability is a defining characteristic of short-discriminator promoters.

Promoter-specific effects on the kinetics of initial transcription are clear from comparison of the time courses of short RNA transients at λP_R, T(T), and L(T) promoter constructs. Analysis of these time courses for the T(T) promoter reveal the promoter escape occurs after 8-mer RNA synthesis, and that RNAP-promoter contacts are broken in two phases. It also revealed that at 19 °C a CC to OC conversion step is necessary for productive initiation at T(T). This conversion cannot explain the dramatic reduction in k_{8+} observed for initiation from both the T(T) and L(T) promoter constructs at the low UTP condition.

Methods

Prediction of Discriminator Length via Consensus -10 Element Sequence

The total number of each occurrence of the -10 element consensus sequence (TATAAT) and all sequences with one deviation from the -10 consensus sequence (e.g.

CATAAT) within the RegulonDB σ^{70} TSS database [13] was first determined. For each position, there are four possible sequences (e.g. at position 1 the possible sequences are AATAAT, CATAAT, GATAAT, and TATAAT). The number of occurrences of one of these sequences divided by the sum total of all four of these sequences is considered the probability of the corresponding nucleotide occurring at that position in the -10 hexamer. The position weight matrix (PWM) for each position in the -10 hexamer was generated in this way.

For each position at each promoter in the RegulonDB database, the -10 element score was generated as the product the probability of each NTP each position in the hexamer originating at this position. For example, if a given position is position 1 in the sequence GATCTA, the -10 element score will be the PWM probability of G at position 1, multiplied by the PWM probability of position 2, etc. The hexamer within either position -1 and -20, or -10 and -16 that has the highest -10 element score determines the -10 element position chosen for the purposes of determining the discriminator length.

Reagents, Buffers, and Gels

Reagents for buffers and stock solutions were the highest available grade and were used as received. All solutions were prepared using 18 MΩ deionized water from a Barnstead EPure system. NTPs and dNTPs (Boston Bioproducts, Thermo Fisher, New England Biolabs) used in transcription assays and PCR reactions were 99% pure and used as received. Enzymes for PCR reactions were purchased from NEB and used according to the manufacturer's protocols.

Storage buffer (SB) for core RNAP, σ^{70} and RNAP holoenzyme is 50% v/v glycerol, 0.01 M Tris (pH 8.0), 0.1 M NaCl, 0.1 mM EDTA, and 0.1 mM DTT, 0.05 mg/mL BSA.

Transcription buffer (TB) is 40 mM Tris, 5 mM MgCl2, 60 mM KCl, 1 mM DTT, and 0.05 mg/mL BSA, adjusted to pH 8.0 at the experimental temperature (19 °C, 25 °C, 37 °C).

2X initiation solution (IS) for rapid quench flow (RQF) transcription assays at 19 °C, 25 °C, and 37 °C assayed with α-^{32}P-GTP ("High UTP") is 400 μM ATP and UTP, 50 μM GTP, and 35 nM α-^{32}P-GTP, 0.1 mg/mL heparin in TB. IS with α-^{32}P-UTP ("Low UTP") is 400 μM ATP and GTP, 35 nM UTP, and 35 nM α-^{32}P-UTP, 0.1 mg/mL heparin in TB. 5X initiation solution for primer extension assays (IS_{PE}) is 1 mM all NTPs and 0.1 mg/mL heparin in TB.

Quench Solution (QS) for manual mixing transcription assays is 8 M urea and 15 mM EDTA in TB. QS with added dyes (QSD) for polyacrylamide gel electrophoresis (PAGE) has 0.05% xylene cyanol and 0.05 bromphenol blue in QS. TBE buffer for PAGE is 90 mM Tris-Borate (pH 8.3) and 2 mM Na_2EDTA. All gels for initial transcription assays are 20% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics). All gels for primer extension assays are 12% acrylamide-(bis)acrylamide (19:1), and were made using the UreaGel system (National Diagnostics).

RNA Polymerase and Promoter DNA

The RNAP holoenzyme is overexpressed (as core RNAP and σ^{70} separately), purified, and reconstructed as described previously [16]. Filter binding activity assays performed on preparations of RNAP holoenzyme used here show that 50% ± 10% of RNAP molecules form a stable open complex with the λP_R promoter at 37 °C. All RNAP concentrations reported here refer to this active fraction. Sequences of PCR primers and promoter DNA fragments used to prepare the T(T), L(T), GcdP1, λP_R G3T (GcdP1), λP_R

(GcdP1), and λP_R chimeric promoters used to determine TSS selection of short discriminators (Figure 3) are reported in Table 3 with annotations for their use.

TSS Determination

RNAP (25 nM) was incubated with promoter DNA (12.5 nM) for 1 h at 37 °C. 5X IS_{PE} is manually mixed 1:4 with the RNAP-promoter DNA solution to obtain a final promoter concentration of 10 nM. The reaction is quenched with QS 300 s after IS_{PE} addition. RNA product length was determined via primer extension with Avian Myeloblastosis Virus (AMV) Reverse Transcriptase (NEB) primed by a 5'-end-labelled 23 base primer that anneals 9 bp downstream of the predicted TSS (Table 3). Primer extension products were separated via PAGE and imaged via phosphorimager as previously described [1]. Primer extension products were quantified via ImageQuant software.

Single-Round Transcription Assays

Rapidly quenched initiation kinetics assays were performed at 19 °C, 25 °C, and 37 °C and analyzed as previously described [16]. Briefly, 100 nM DNA is preincubated with 200 nM RNAP for 1 h at the experimental temperature. 2X IS is mixed 1:1 with this solution at time zero using a KinTek Corp Rapid Quench Flow (RQF) apparatus to obtain final promoter concentration of 50 nM. Reactions are quenched at the given time points with QS and RNA is separated via PAGE. Transcription gels are imaged and quantified as previously described [16].

Determination of Stepwise RNA Synthesis Rate Constants at the T(T) Promoter

To determine forward rate constants k_i for the T(T) promoter at 19 °C and 25 °C, high UTP and low UTP were simultaneously fit to the stepwise promoter escape

mechanism reported in [16] (Chapter 2). To obtain the time courses of transient productive RNA to use in this fitting, nonproductive RNA synthesis was modeled (via fitting the plateau value of total RNA to a single-exponential approach to a plateau value) and subtracted from the total amount of RNA synthesis as previously described [16] (Chapter 2). The result of this analysis is indicated on figures 5 – 16 where applicable.

Figures

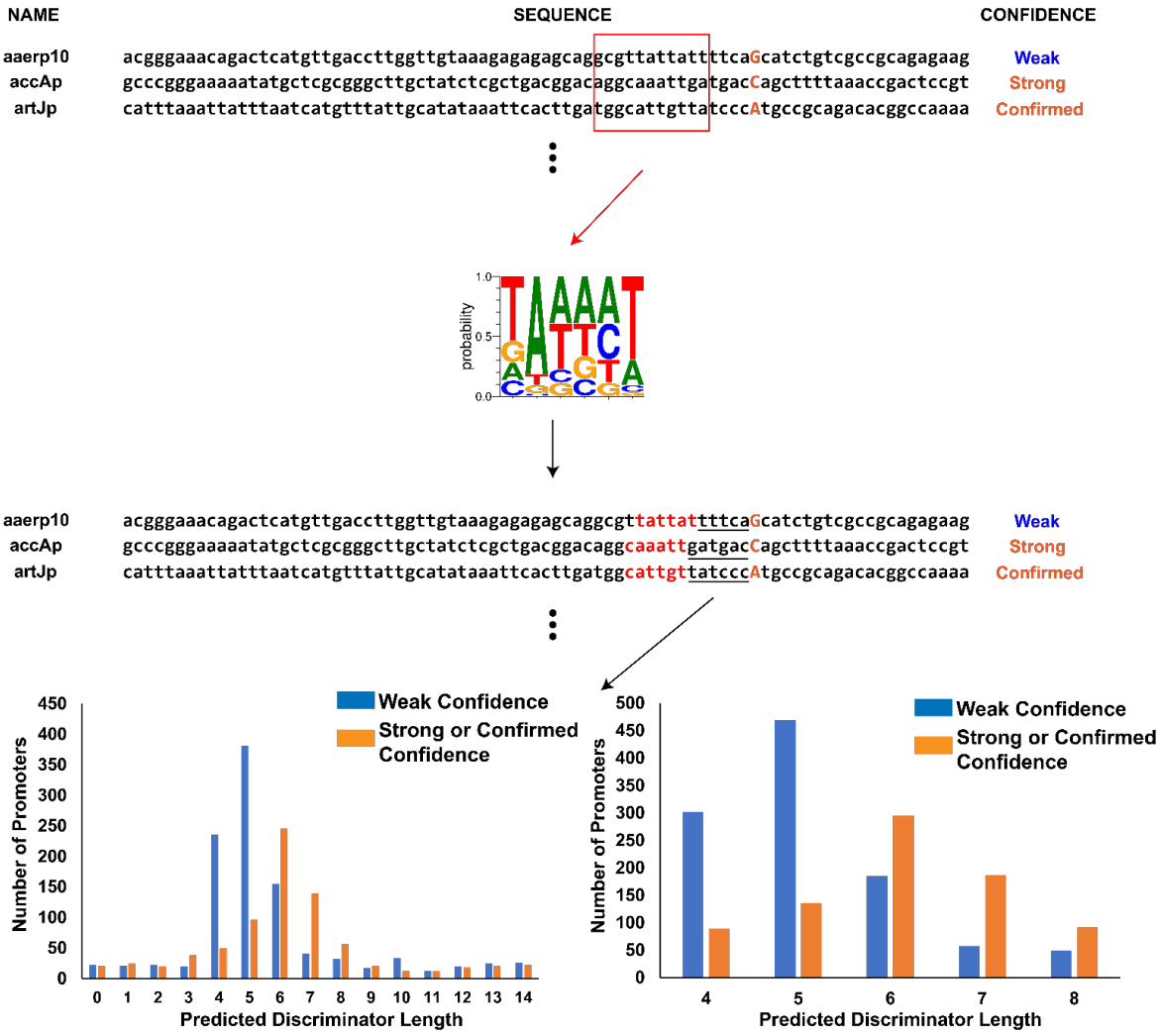


Figure 1: Prediction of Discriminator Length via Consensus-Sequence Based Motif Prediction: Panel A: Three example promoter entries in the RegulonDB σ^{70} TSS database. Promoter name, full tabulated sequence, and confidence designation are shown (Weak confidence: Blue, Strong or confirmed confidence: Orange). The transcription start site position is indicated by a capital letter, and highlighted orange. The 10 bp region predicted to contain the majority of -10 sequences is boxed in red. **Panel B:** The predicted probability of occurrence each nucleotide in the -10 hexamer is represented

by relative size of the corresponding letter. Probabilities were determined by the ratio of the number of occurrences of a single sequence with 1 bp deviations from the consensus -10 sequence (e.g. 5'-GATAAT-3') to the total number of occurrences of such sequences (5'-NATAAT-3'). **Panel C:** Three example promoter entries in the RegulonDB TSS database. Most likely -10 hexamers, predicted using the base probabilities in panel B, are shown at each promoter in red. The discriminator region is underlined. TSS position and confidence designation are indicated as in panel A. **Panel D:** Distribution of discriminator length in the entire set of *E. coli* σ^{70} promoters in the RegulonDB database. Promoters with weak confidence designation are blue, promoters with strong and confirmed confidence designations are orange. X-axis range for each distribution represents the imposed constraints for the distance of the predicted -10 element from promoter DNA.

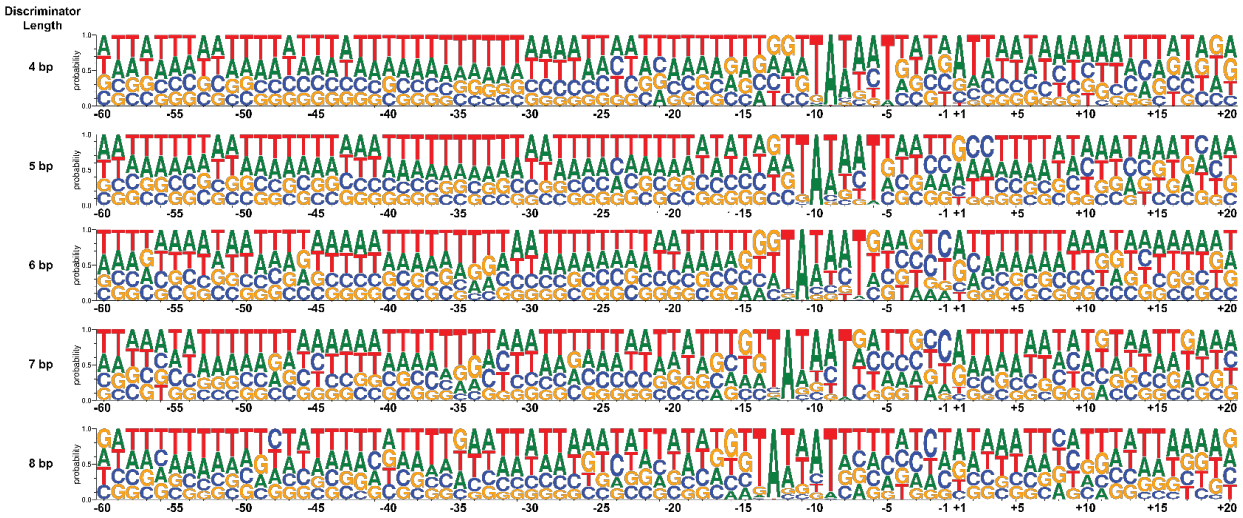


Figure 2: Nucleotide Probability at Each Promoter and ITR Position for Each Predicted Discriminator Length. The probability of each nucleotide at each position for all strong and confirmed confidence σ^{70} promoters predicted to have a discriminator length from 4 to 8 bp are shown here. Discriminator lengths are indicated on the left. Sequence positions are numbered relative to the transcription start site (+1). Logo was generated using WebLogo v. 3.7.4. Probability was determined by the fraction of sequences with a given predicted discriminator length that had each NTP at a given position.

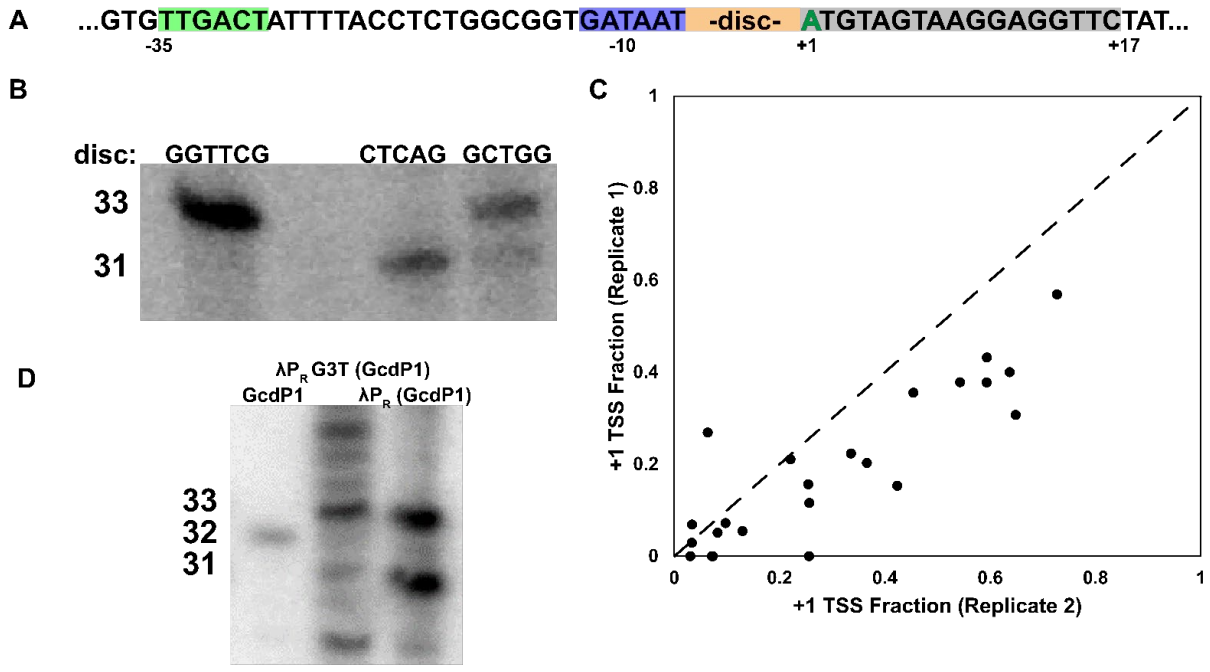


Figure 3: Discriminator Effects on TSS Selection at Hybrid Promoters. Panel A: The nontemplate-strand sequence of the core region of the modified λP_R promoter (in a 120 or 121 bp (-80 to -42) promoter fragment) used in this study of TSS selection. The region labelled -disc- was substituted with predicted 4 bp and 5 bp discriminators selected from those designated strong or confirmed confidence in the RegulonDB TSS database. Upstream promoter elements are numbered assuming a 6 bp discriminator. **Panel B:** Primer extension products indicating TSS selection by hybrid promoters containing the indicated example 5 bp discriminator sequences. Length of product (bp) are indicated on the left. Products starting at the predicted +1 position are 33 bp long, products starting at the predicted +3 position are 31 bp long. The leftmost lane shows the product from the native 6 bp λP_R discriminator. **Panel C:** Comparison of fraction of primer extension products starting from the predicted +1 position for two primer extension replicates. All promoter constructs shown had a predicted 5 bp discriminator. Dashed line indicates a

1:1 correlation between the two replicates. **Panel D:** Primer extension products for the GcdP1 WT promoter (5 bp, Gcpd1 core promoter, discriminator, and ITR), λP_R G3T (GcdP1) hybrid promoter (λP_R core promoter, GcdP1 discriminator, λP_R ITR with a G3T mutation), and λP_R (GcdP1) hybrid promoter (λP_R core promoter, GcdP1 discriminator, λP_R ITR) are shown. Length of product (bp) are indicated on the left. For the GcdP1 ITR, products which start at the predicted +1 site are 32 bp long. For both the λP_R and λP_R G3T ITRs, products starting at the predicted +1 position are 33 bp long, products

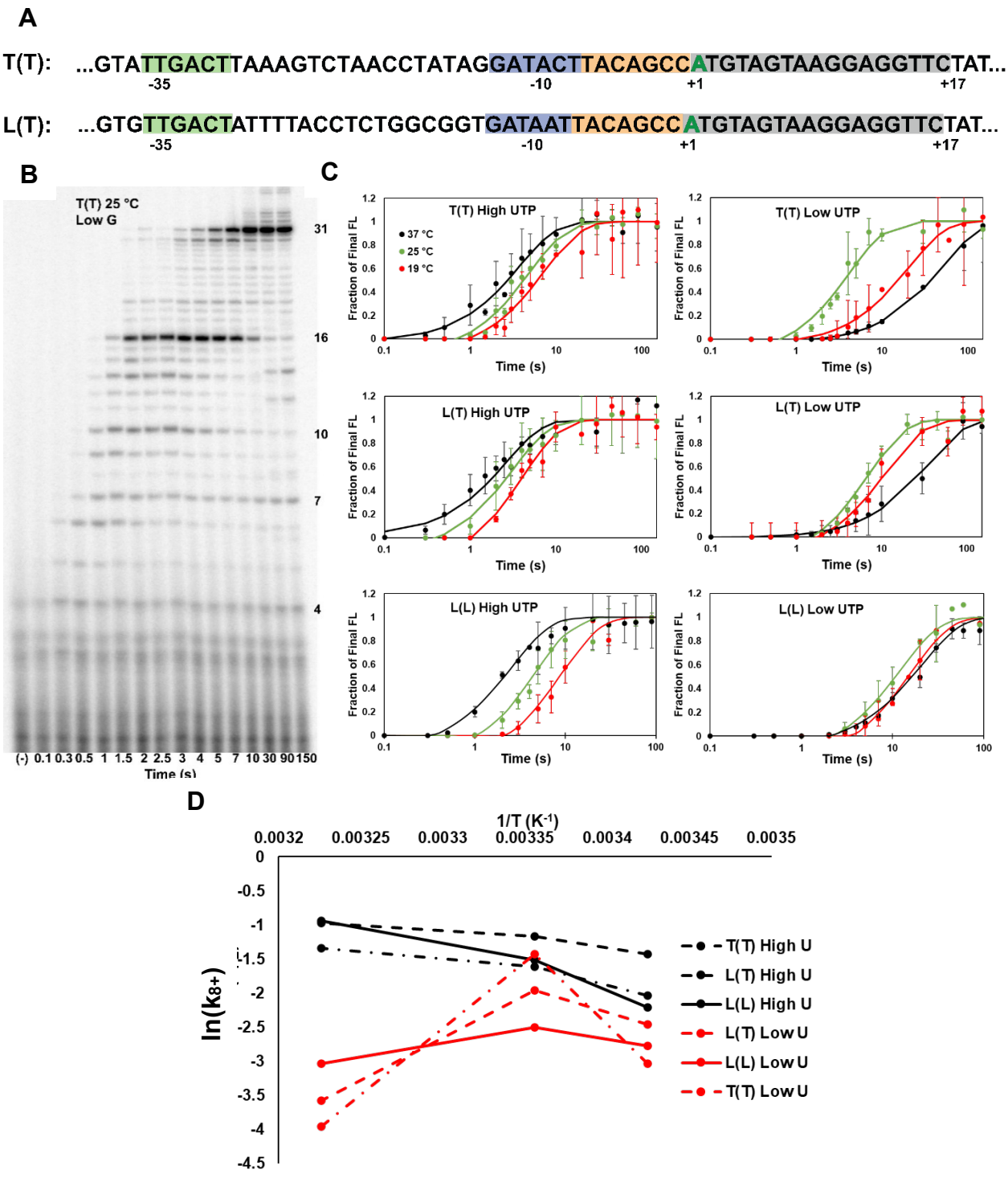


Figure 4: Synthesis of Long (post-escape) RNA at T(T) and L(T) Promoters; Comparison with L(L). Panel A: The nontemplate strand sequence of the core promoter and ITR region of the T(T) and L(T) promoter constructs used in this study of stepwise

initiation kinetics. **Panel B:** Representative gel showing single-nucleotide resolution PAGE separation of ^{32}P-labelled RNA products from RQF initiation from the T(T) promoter at 200 μM ATP and UTP, 10 μM GTP at 25 °C. **Panel C:** Time courses (log scale) of long RNA (>7-mer for T(T) and L(T), > 10-mer for L(L)) synthesis at the T(T), L(T), and λP_R (L(L)) promoters at 19 °C (red), 25 °C (green), and 37 °C (black) and at high UTP (left) and low UTP (right). L(L) graphs are from Chapter 2 and are included for comparison. Panel D: Natural logarithms of first order rate constants (k_{8+}) for synthesis of long (post-escape) RNA (k_{8+}) are plotted vs inverse T (K^{-1}) for T(T) (dashed), L(T) (dashed/dotted), and L(L) (solid) for High UTP (black) and Low UTP (red) conditions. At both NTP conditions, but especially low U, the 37 °C rate constant is significantly smaller than predicted assuming a linear Arrhenius relationship defined by the 19 °C and 25 °C rate constants. L(L) values are from Chapter 2 and are included for comparison.

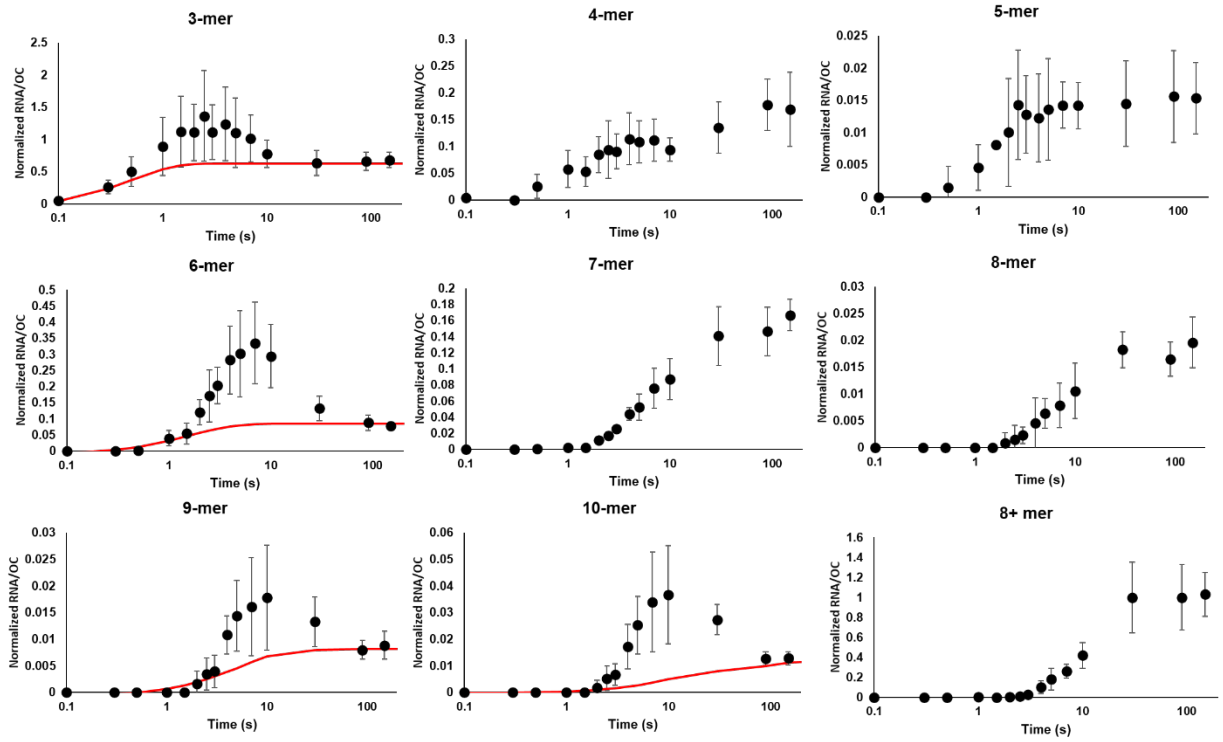


Figure 5: Average amount of RNA vs. Time at T(T), 19 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 5 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. For each RNA length where a productive transient peak is visible and substantial nonproductive RNA synthesis is observed, the predicted contribution from the initial round of synthesis by nonproductive complexes is shown (red). For 9-mer and 10-mer, contributions from hypothetical nonproductive complexes are also modeled, although T(T) is predicted to escape prior to 9-mer synthesis and therefore would not be expected to produce nonproductive RNA. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA.

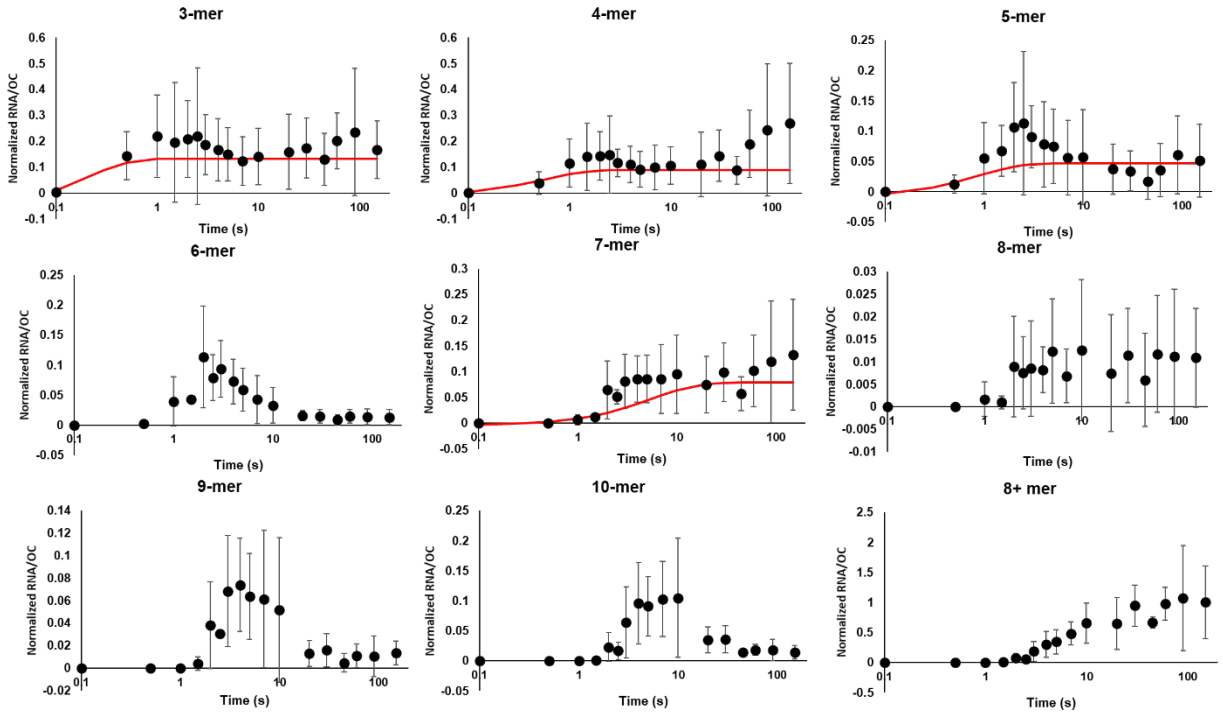


Figure 6: Average amount of RNA vs. Time at T(T), 19 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3-5 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. For each RNA length where a productive transient peak is visible and substantial nonproductive RNA synthesis is observed, the predicted contribution from the initial round of synthesis by nonproductive complexes is shown (red). All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA.

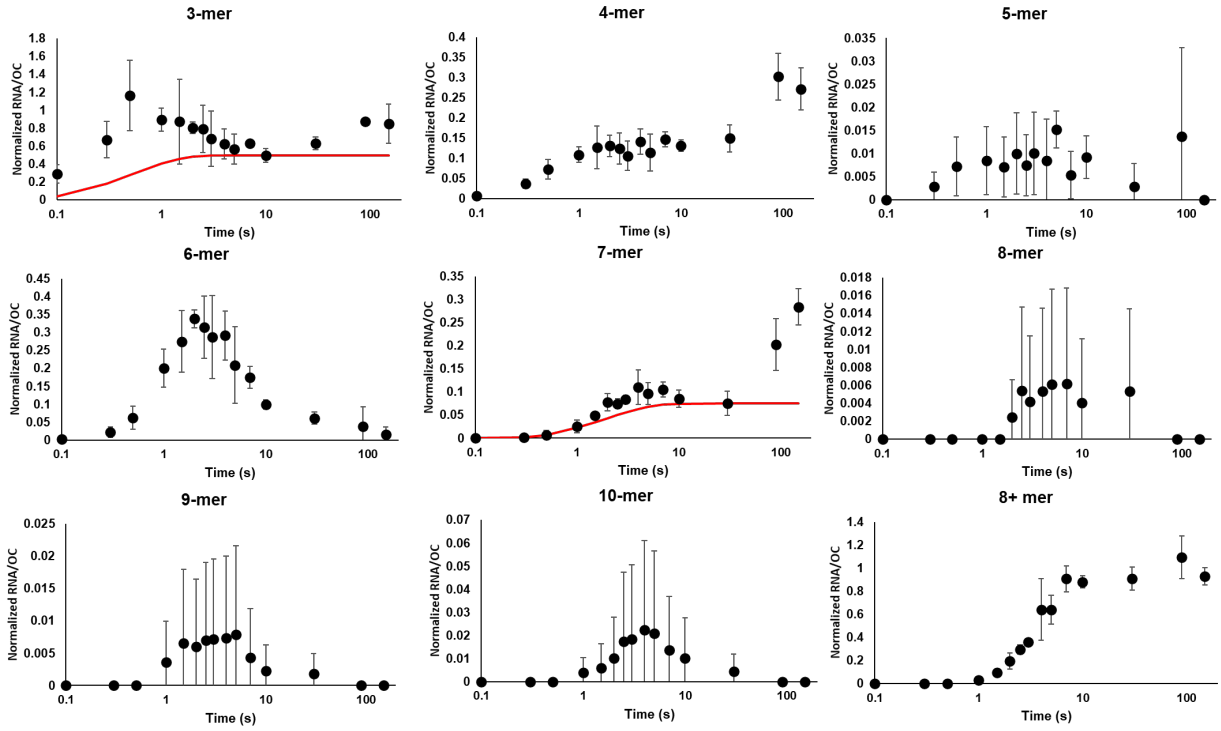


Figure 7: Average amount of RNA vs. Time at T(T), 25 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. For each RNA length where a productive transient peak is visible and substantial nonproductive RNA synthesis is observed, the predicted contribution from the initial round of synthesis by nonproductive complexes is shown (red). All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA.

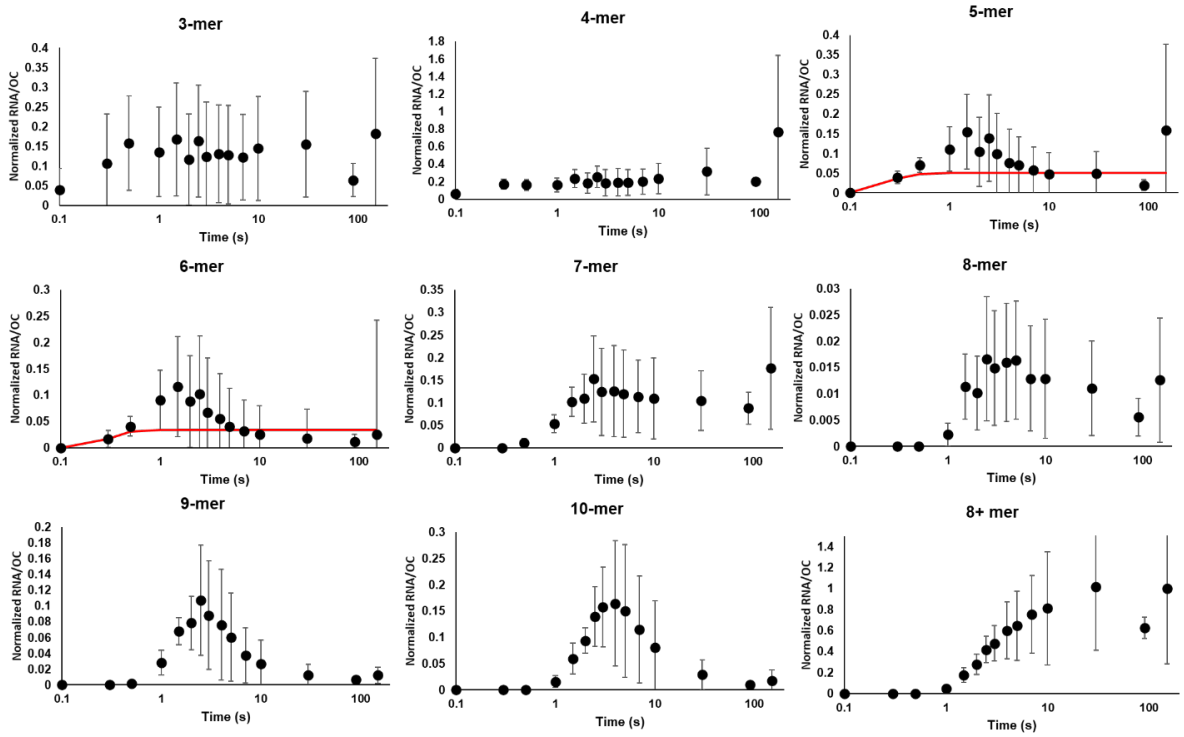


Figure 8: Average amount of RNA vs. Time at T(T), 25 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. For each RNA length where a productive transient peak is visible and substantial nonproductive RNA synthesis is observed, the predicted contribution from the initial round of synthesis by nonproductive complexes is shown (red). All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA.

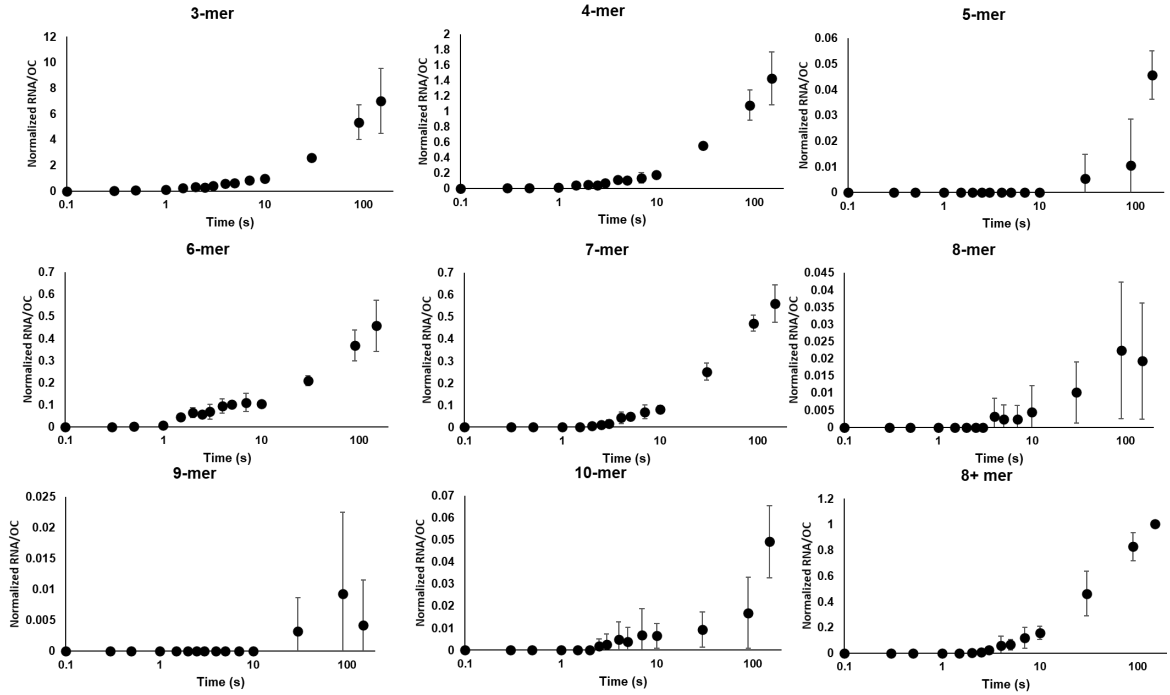


Figure 9: Average amount of RNA vs. Time at T(T), 37 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 4 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. All lengths are normalized to the maximum amount of productive OC detected at this condition, determined by the maximum value of 8+ mer RNA.

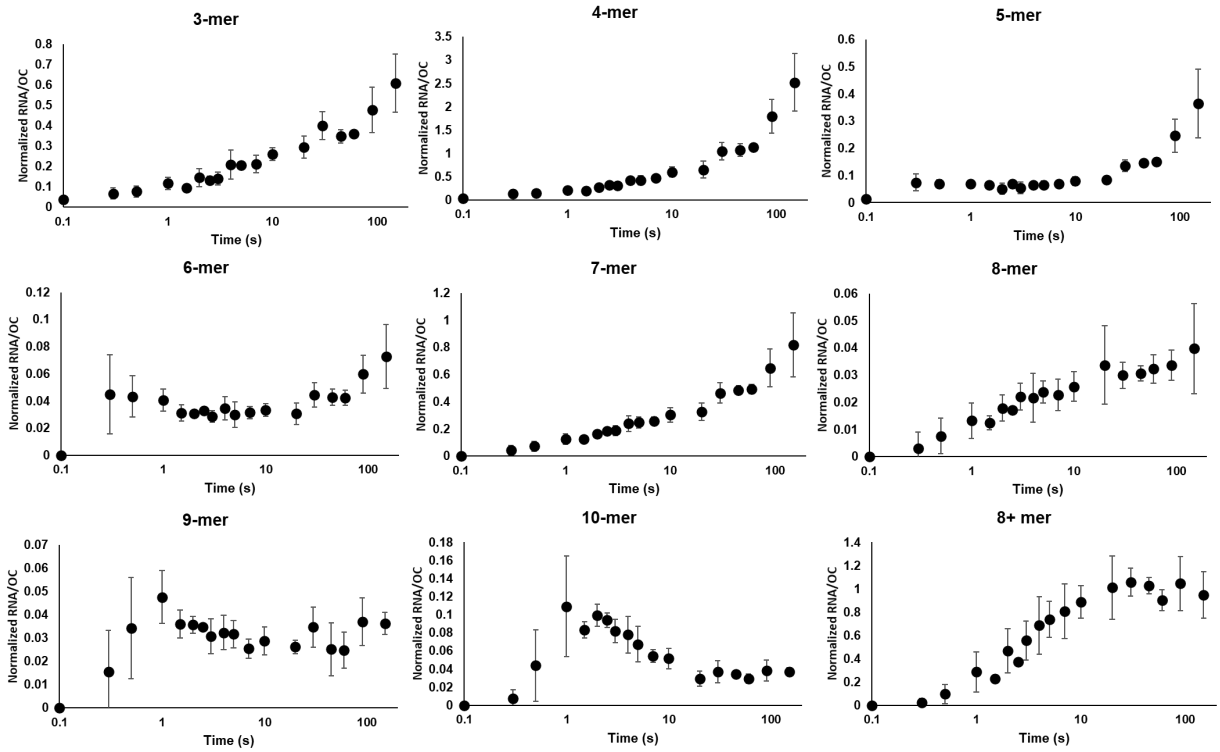


Figure 10: Average amount of RNA vs. Time at T(T), 37 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 4 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA.

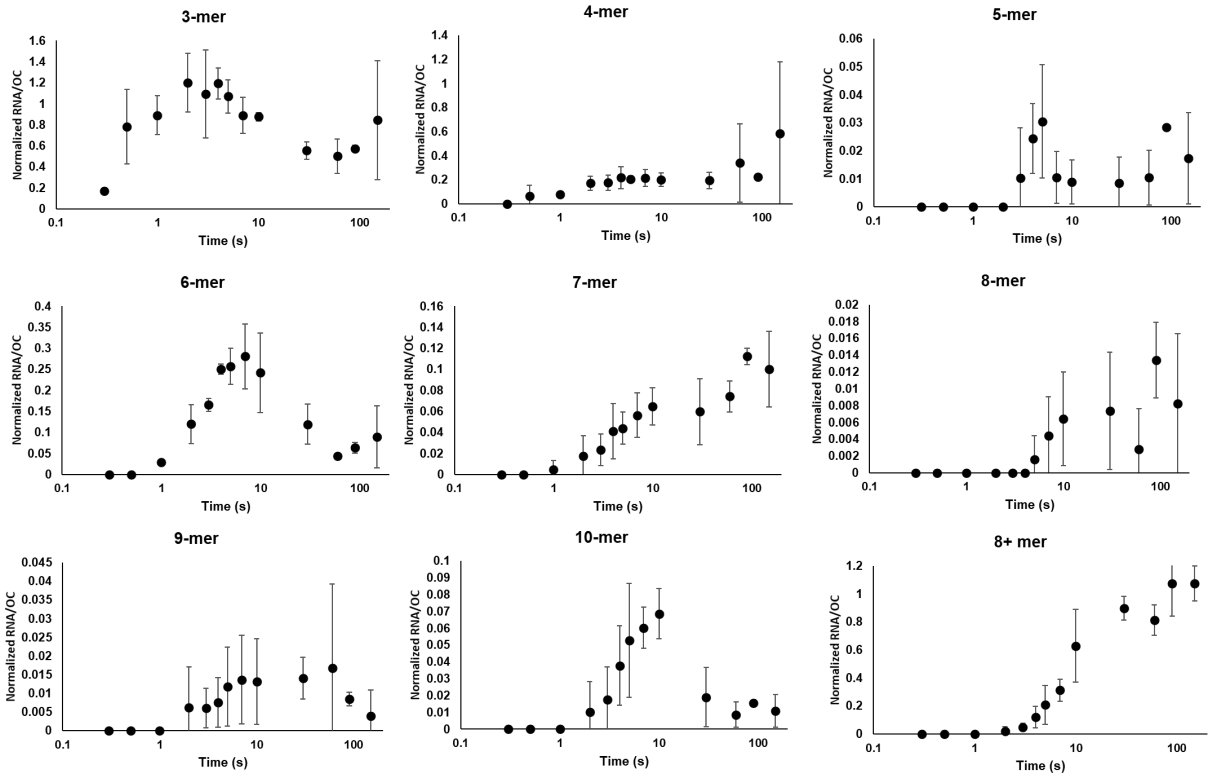


Figure 11: Average amount of RNA vs. Time at L(T), 19 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA

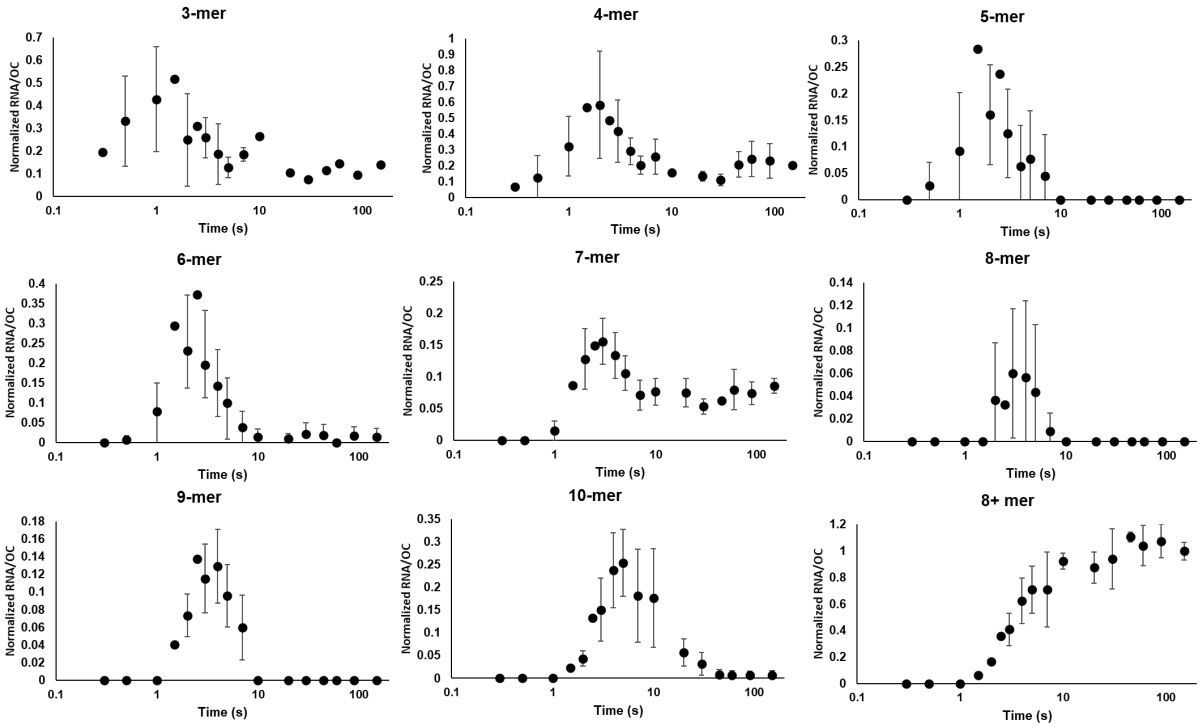


Figure 12: Average amount of RNA vs. Time at L(T), 19 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA

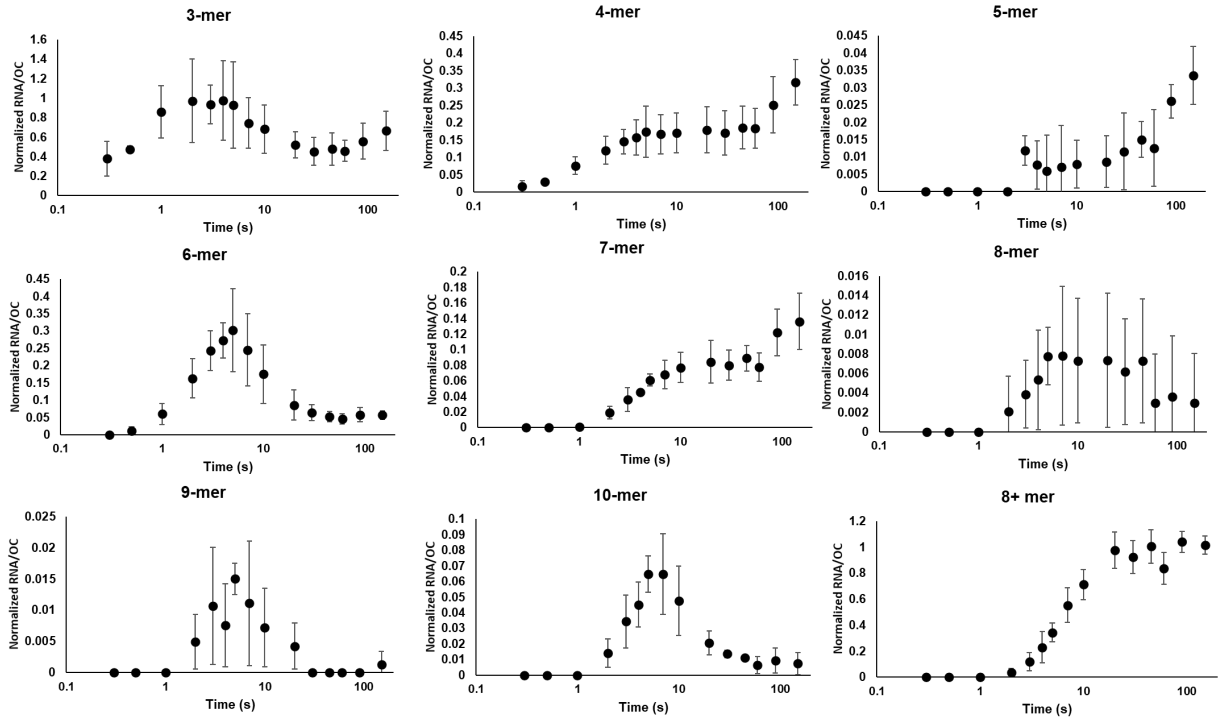


Figure 13: Average amount of RNA vs. Time at L(T), 25 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA

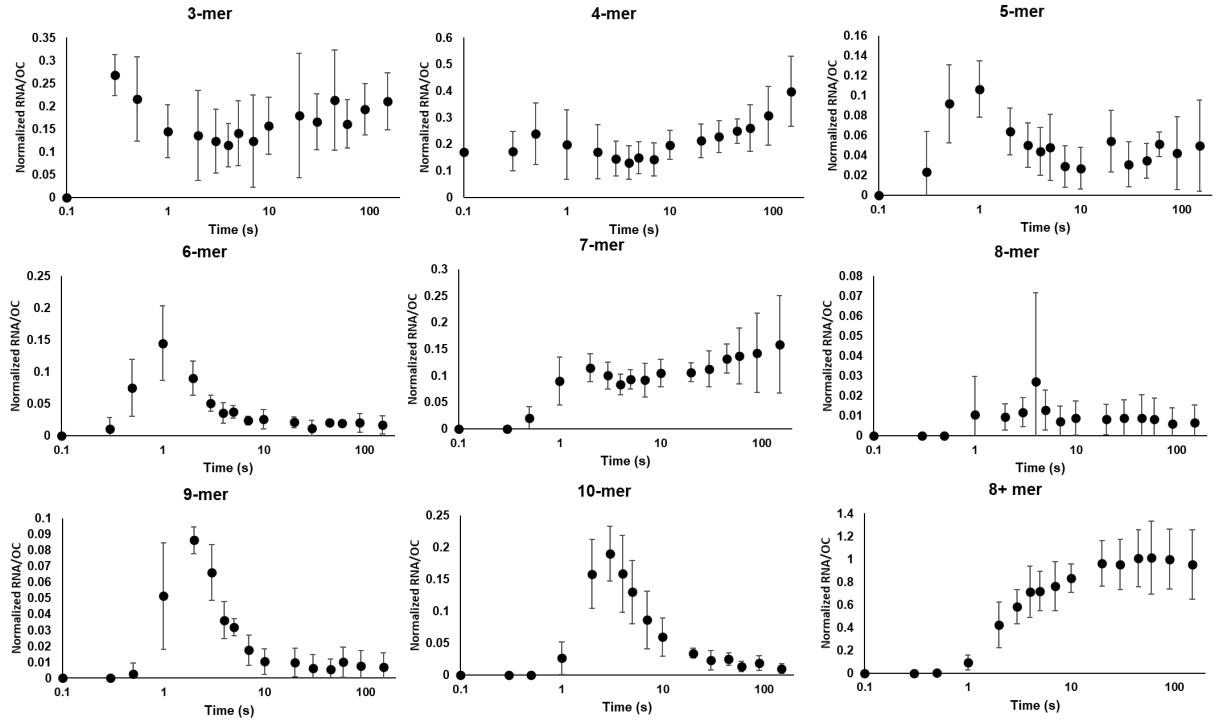


Figure 14: Average amount of RNA vs. Time at L(T), 25 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 3 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. All lengths are normalized to the total amount of productive OC produced at this condition, determined by the plateau amount of 8+ mer RNA

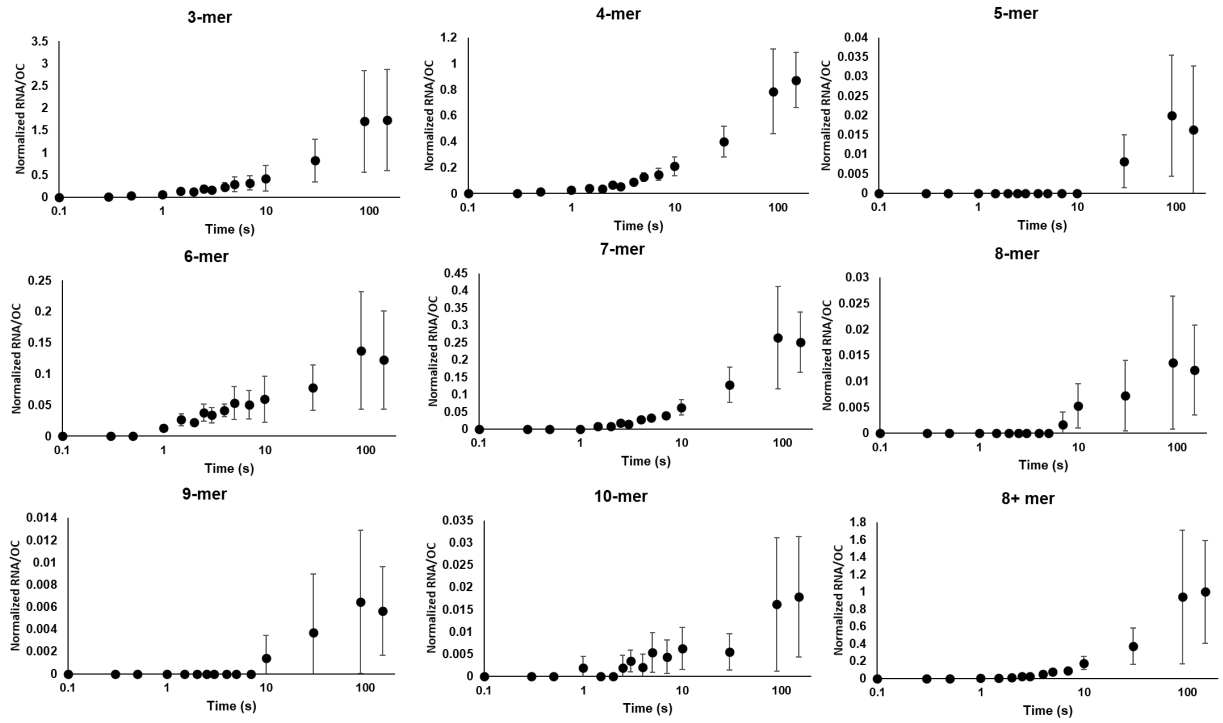


Figure 15: Average amount of RNA vs. Time at L(T), 37 °C, Low UTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 4 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and GTP, 10 μM UTP. All lengths are normalized to the maximum amount of productive OC detected at this condition, determined by the maximum value of 8+ mer RNA.

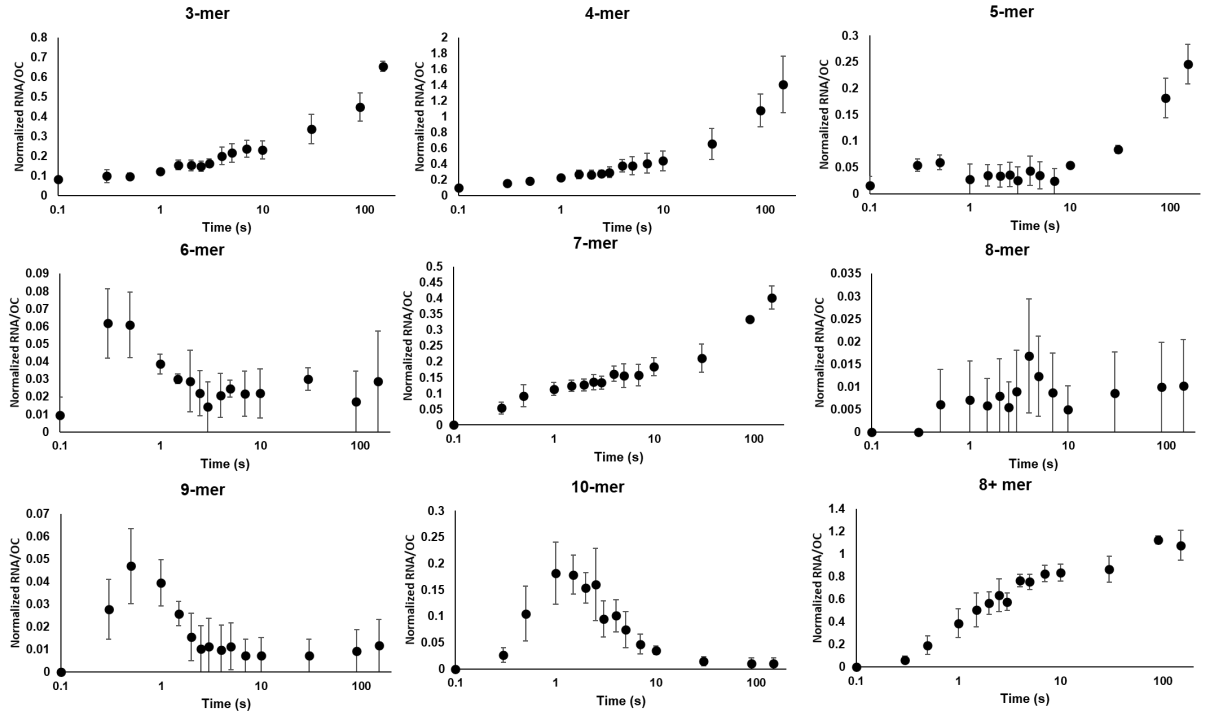


Figure 16: Average amount of RNA vs. Time at L(T), 37 °C, Low GTP. Amounts of 3-mer to 10-mer RNA per OC (averages of 4 independent experiments) are shown with one standard deviation error bars (black). NTP concentrations are 200 μM ATP and UTP, 10 μM GTP. All lengths are normalized to the maximum amount of productive OC detected at this condition, determined by the maximum value of 8+ mer RNA.

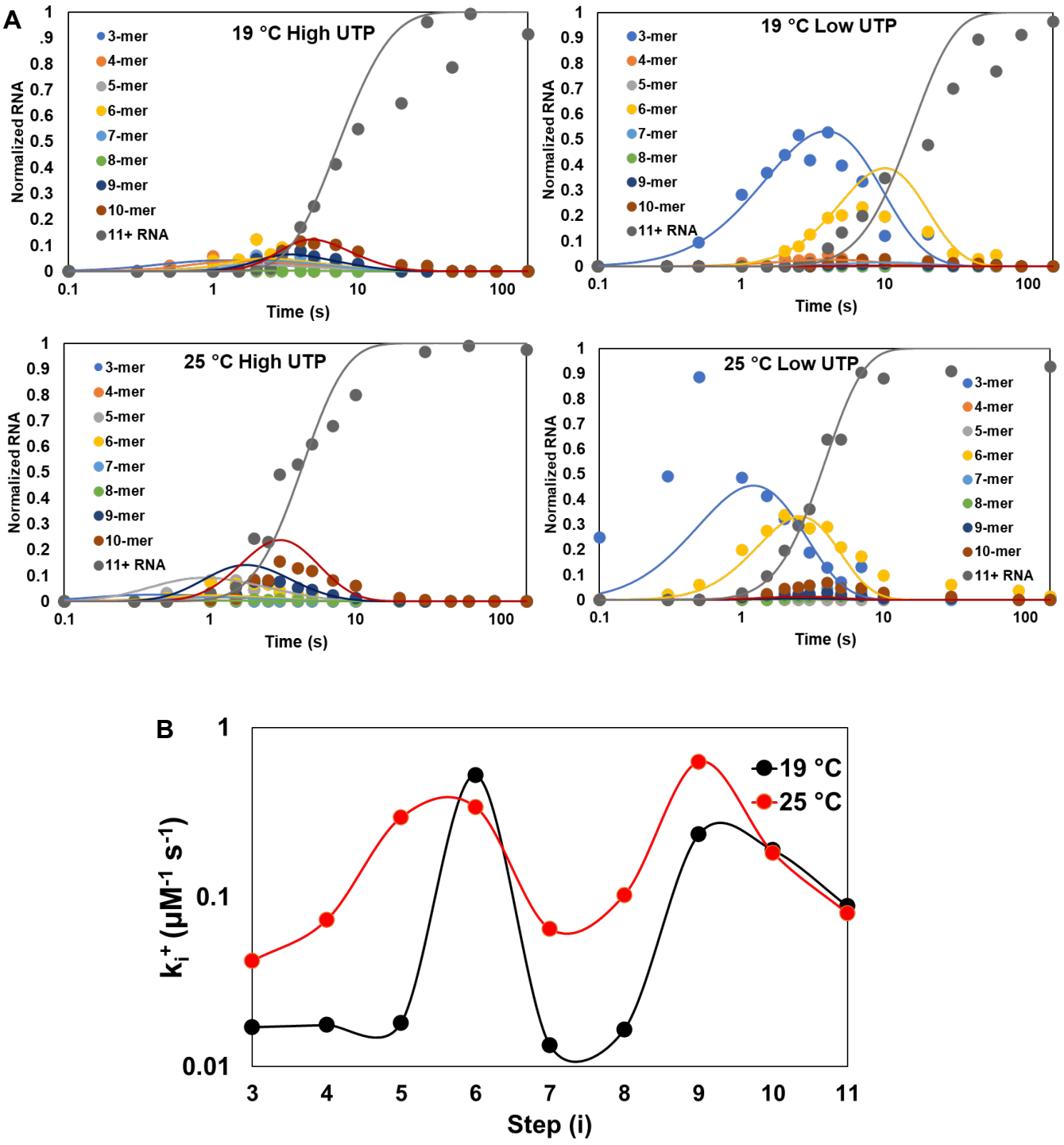


Figure 17: T(T) Stepwise Kinetic Parameters at 19 °C and 25 °C. Panel A: Comparison of productive initiation data with fits (solid curves) to the stepwise mechanism reported in chapter 2 at 19 °C (top) and 25 °C (bottom). **Panel B:** Composite 2nd order rate constants

k_i^+ for NTP addition steps 3-11 at 19 °C (black) and 25 °C (red) determined from the fits in panel A, plotted on a log scale.

Table 1: $t_{1/2}$ Comparison Between L(L), T(T), and L(T)

	L(L)[a]	T(T)	L(T)
High UTP 37 °C	1.8 s	3.5 s	1.8 s
Low UTP 37 °C	14 s	36.5 s	24.8 s
High UTP 25 °C	2.2 s	3.5 s	2.2 s
Low UTP 25 °C	8.5 s	2.9 s	5.0 s
High UTP 19 °C	6.3 s	5.3 s	5.3 s
Low UTP 19 °C	11.2 s	14.4 s	8.2 s

Table 2: Stepwise RNA extension parameters for T(T)

Step	Base Incorporated	19 °C $K_{eq, 1a}$	25 °C[b] $K_{eq, 1a}$
1a		0.33	1.5
		k_i ($\mu M^{-1} s^{-1}$)	**k_i ($\mu M^{-1} s^{-1}$)**
3	G	0.017 ± 0.001	0.042
4	U	0.018 ± 0.001	0.073
5	A	0.018 ± 0.008	0.30
6	G	0.52 ± 0.45	0.34
7	U	0.013 ± 0.001	0.0651
8	A	0.016 ± 0.01	0.10
9	A	0.23 [a]	0.63
10	G	0.19 ± 0.07	0.18
11	G	0.09 ± 0.02	0.08

[a] The number reported is a lower bound

[b] Parameters are not well-enough determined to have reportable error. Values reported therefore reflect a set of k_i^+ that accurately replicate some, but not all, features of the experimental data.

Table 2: Primers used for construction and primer extension of T(T), L(T), Gcdp1 wt, λP_R (gcdp1), λP_R (gcdp1) G3T, and λP_R chimeric (short discriminator) template preparation

DMP_T7A1_F	CTCCACGAATTCAATTTAAAATTTATCAAAAA GAGTATTGACTTAAAGTCTAACCTATAG	forward oligo for T(T)
DMP_T7A1_R	TTTCAGATTGGATATCCTATGAATGTCGGTA CATCATTCCTCCAAGATACTTCCAAAACA	reverse oligo for T(T)
DMP_LPR_F	CCACGAATTCGGATAAATATCTAACACCGTG CGTGTTGACTATTTTACCTCTGGCGGTG	forward oligo for L(T), λP_R (Gcdp1), λP_R (Gcdp1) G3T, λP_R chimeric promoters
DMP_LPR(T7A1) _TXN_R	ACAAAACCTTCATAGAACCTCCTTACTACAT GGCTGTAATTATCACCGCCAGAGGT	reverse oligo for L(T)
HBOT	CACCTGCACCGACAAAACCTT	PCR reverse primer for T(T), L(T), λP_R (Gcdp1), λP_R (Gcdp1) G3T, λP_R chimeric promoters
HTOP	CCAGCATTCCTCCACGAATTC	PCR forward primer for T(T), L(T), λP_R (Gcdp1), λP_R (Gcdp1) G3T, λP_R chimeric promoters
DMP_gcdp1_for	ATTCGCGGAATCTGTTAATAAAGCTCGAATT GTGATGACGATCACACATGTTAAAACCT	forward oligo for Gcdp1 WT promoter
DMP_gcdp1_gcdp 1-lpr_rev	ACAAAAGCTTCAGGGAACCTCCTTACTACAT AATTAATTATAGGTTTTAACATGT	reverse oligo for λP_R (Gcdp1) promoter
DMP_gcdp1_gcdp 1-lpr3T_rev	ACAAAAGCTTCAGGGAACCTCCTTACTAAAT AATTAATTATAGGTTTTAACATGT	reverse oligo for λP_R (Gcdp1) G3T promoter
DMP_gcdp1wild_r ev	CACCATTTCAATTCATTAATATTTTAGTAGCA ATTAATTATAGGTTTTAACATGT	reverse oligo for gcdp1 wt promoter
DMP_gcdp1wild_ HBOT	CCATAAAGAGACACCATTTC	PCR reverse primer for gcdp1 wt promoter
DMP_LPR_Discm od_1_R	ACAAAACCTTCATAGAACCTCCTTACTACATt tgcATTATCACCGCCAGAGGT	This an subsequent primers are reverse oligo for chimeric λP_R promoters (discriminators in lower case).
DMP_LPR_Discm od_2_R	ACAAAACCTTCATAGAACCTCCTTACTACATc atcATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_3_R	ACAAAACCTTCATAGAACCTCCTTACTACAT gttgATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_4_R	ACAAAACCTTCATAGAACCTCCTTACTACAT ggcaATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_5_R	ACAAAACCTTCATAGAACCTCCTTACTACAT aattATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_6_R	ACAAAACCTTCATAGAACCTCCTTACTACATc ctgATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_7_R	ACAAAACCTTCATAGAACCTCCTTACTACAT gcttATTATCACCGCCAGAGGT	
DMP_LPR_Discm od_8_R	ACAAAACCTTCATAGAACCTCCTTACTACATc agcATTATCACCGCCAGAGGT	

DMP_LPR_Discmod_9_R	ACAAAACCTTCATAGAACCTCCTTACTACATatggATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_10_R	ACAAAACCTTCATAGAACCTCCTTACTACATgtatATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_11_R	ACAAAACCTTCATAGAACCTCCTTACTACATgctcATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_12_R	ACAAAACCTTCATAGAACCTCCTTACTACATcaaaATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_13_R	ACAAAACCTTCATAGAACCTCCTTACTACATgacgATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_14_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**cggg**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_15_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**t****ggt**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_16_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**cgctc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_17_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gttgc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_18_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gattt**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_19_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**aaact**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_20_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**t****ggct**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_21_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**aggta**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_22_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ccgcc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_23_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gctcg**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_24_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**t****tgtt**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_25_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**cgtta**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_26_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ccagc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_27_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gcgct**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_28_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**atcag**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_29_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ggttt**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_30_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**t****atga**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_31_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**t****tttc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_32_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ggggg**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_33_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**aatac**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_34_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gacgg**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_35_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**acaag**ATTATCACCGCCAGAGGT	

DMP_LPR_Discmod_36_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**tgttc**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_37_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ggttt**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_38_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**agcta**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_39_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**ggtcg**ATTATCACCGCCAGAGGT	
DMP_LPR_Discmod_40_R	ACAAAACCTTCATAGAACCTCCTTACTACAT**gactc**ATTATCACCGCCAGAGGT	
DMP_PE23OL_LPRHBOT	CGACAAAAGCTTCAGGGAACCTC	Primer extension primer for chimeric λP_R promoters
DMP_PE23gcdp1	ACACCATTTCAATTCATTAATAT	Primer extension primer for gcdp1 wt promoter

References

[1] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[2] Libing Y, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. Elife. 2017;6.

[3] Winkelman JT, Chandrangsu P, Ross W, Gourse RL. Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. Proceedings of the National Academy of Sciences. 2016;113:1787-95.

[4] Vvedenskaya IO, Vahedian-Movahed H, Zhang Y, Taylor DM, Ebright RH, Nickels BE. Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. Proceedings of the National Academy of Sciences. 2016;113:E2899-E905.

[5] Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural Basis of Transcription Initiation. Science. 2012;338:1076-80.

[6] Saecker RM, Chen J, Chiu CE, Malone B, Sotiris J, Ebrahim M, et al. Structural origins of Escherichia coli RNA polymerase open promoter complex stability. Proceedings of the National Academy of Sciences. 2021;118:e2112877118.

[7] Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. Molecular Cell. 2015;58:534-40.

[8] Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, et al. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science. 2016;351:1090-3.

[9] Heyduk E, Heyduk T. Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. Biochemistry. 2014;53:282-92.

[10] Hsu LM, Cobb IM, Ozmore JR, Khoo M, Nahm G, Xia L, et al. Initial Transcribed Sequence Mutations Specifically Affect Promoter Escape Properties. Biochemistry. 2006;45:8841-54.

[11] Basu RS, Warner BA, Molodtsov V, Pupov D, Esyunina D, Fernandez-Tornero C, et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. Journal of Biological Chemistry. 2014;289:24549-59.

[12] Haugen SP, Ross W, Manrique M, Gourse RL. Fine structure of the promoter-sigma region 1.2 interaction. Proceedings of the National Academy of Sciences. 2008;105:3292-7.

[13] Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation inE. coliK-12. Nucleic Acids Research. 2019;47:D212-D20.

[14] Cassiano MHA, Silva-Rocha R, Fodor A, Su Z. Benchmarking Bacterial Promoter Prediction Tools: Potentialities and Limitations. mSystems. 2020;5:e00439-20.

[15] Hawley DK, McClure WR. Compilation and analysis ofEscherichia colipromoter DNA sequences. Nucleic Acids Research. 1983;11:2237-55.

[16] Plaskon DM, Henderson KL, Felth LC, Molzahn CM, Evensen C, Dyke S, et al. Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. Proceedings of the National Academy of Sciences. 2021;118:e2021941118.

[17] Henderson KL, Evensen CE, Molzahn CM, Felth LC, Dyke S, Liao G, et al. RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. Biochemistry. 2019;58:2339-52.

[18] Johnson KA, Simpson ZB, Blom T. Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. Analytical Biochemistry. 2009;387:20-9.

[19] Sclavi B, Zaychikov E, Rogozina A, Walther F, Buckle M, Heumann H. Real-time characterization of intermediates in the pathway to open complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:4706-11.

[20] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006;314:1144-7.

[21] Winkelman JT, Winkelman BT, Boyce J, Chen AY, Ross W, Gourse RL. Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. Molecular Cell. 2015;59:768-80.

[22] Forrest D, Warman EA, Erkelens AM, Dame RT, Grainger DC. Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity. Nature Communications. 2022;13.

[23] Skalenko KS, Li L, Zhang Y, Vvedenskaya IO, Winkelman JT, Cope AL, et al. Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in Escherichia coli. Proceedings of the National Academy of Sciences. 2021;118:e2106388118.

[24] Yu L, Winkelman JT, Pukhrambam C, Strick TR, Nickels BE, Ebright RH. The mechanism of variability in transcription start site selection. eLife. 2017;6:e32038.

[25] Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277:1453-62.

[26] Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record MT, Jr. *E. coli* polymerase determinants of open complex lifetime and structure. Journal of Molecular Biology. 2015;427:2435-50.

[27] Holbrook JA, Capp MW, Saecker RM, Record MT. Enthalpy and Heat Capacity Changes for Formation of an Oligomeric DNA Duplex: Interpretation in Terms of Coupled Processes of Formation and Association of Single-Stranded Helices†. Biochemistry. 1999;38:8409-22.

[28] Sclavi B, Zaychikov E, Rogozina A, Walther F, Buckle M, Heumann H. Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. Proceedings of the National Academy of Sciences. 2005;102:4706-11.

[29] Metzger W, Schickor P, Meier T, Werel W, Heumann H. Nucleation of RNA chain formation by *Escherichia coli* DNA-dependent RNA polymerase. Journal of Molecular Biology. 1993;232:35-49.

Chapter 6: Discussion

Overview

A more complete understanding of bacterial gene regulation enables breakthroughs in fields like antibiotic discovery and synthetic promoter design. Regulation during transcription initiation can have dramatic effects on gene expression levels and is underpinned by the complex interaction between promoter sequence and the RNA polymerase (RNAP) holoenzyme. In this thesis we determined a detailed mechanism of initiation from the kinetics of initial transcription, and identified correlations between the characteristics of the discriminator element of the promoter and escape kinetics.

In chapters 2 and 3 we determined the kinetics of each step of RNA extension starting from the stable open complex (OC) to the point of promoter escape in the forward (k_i^+, synthesis) and reverse (k_i^-, pyrophosphorolysis) directions. We determined the temperature dependence of k_i^+ for each extension step identified a structural rearrangement step prior to NTP binding that is required for productive RNA synthesis.

In chapter 4 we determined the kinetics of productive RNA synthesis from the model promoter constructs R(R) and L(R). We identified the transcription start site (TSS) of each of these model promoters. Utilizing dinucleotide primers we determined that the high dependence of the kinetics of R(R) promoter escape on the concentration of the 2nd initiating dinucleotide is overcome by priming.

Chapter 5 explored the distribution of discriminator lengths in the *E. coli* genome, and correlations between discriminator length and promoter sequence alongside the unexpectedly high amount of 4 bp and 5 bp discriminators. The role of the discriminator in stepwise escape kinetics was investigated using the 7 bp T7A1 discriminator in the T(T) and L(T) model promoter constructs. We determined that at the T(T) promoter the

pattern of large and small stepwise RNA extension rate constants k_i indicates that promoter escape occurs in two steps. The first of these steps occurs prior to 6-mer synthesis and the second occurs after 6-mer synthesis but before 9-mer synthesis.

These studies highlight the large variations in rates and mechanism between promoters and individual steps within initiation. These variations provide the framework for initiation regulation by additional factors within the cell.

The Kinetics and Mechanism of Productive Initial Transcription at the λP_R Promoter

In chapter 2 through our studies of the temperature-dependence of irreversible RNA synthesis we determined the consistency of the pattern of low and high k_i^+ throughout initial transcription at the λP_R model promoter. These k_i^+ reflect the difficulty of translocation for each extension step, and thus the level of translocation stress at each step due to the upstream promoter contacts remaining in place during initial transcription. The k_i^+ pattern at all temperatures therefore indicate stepwise, rather than concerted, breaking of upstream promoter contacts leading to promoter escape.

To obtain k_i^+ values for productive initiation that accurately reflect the detected transient RNA at 37 °C requires an initial, reversible NTP independent rearrangement step prior to initial NTP binding defined by the equilibrium constant $K_{eq,\,1a}$. At 19 °C and 25 °C, the value of $K_{eq,\,1a}$ indicates a significant amount of the initiation-competent form of the OC, while at 37 °C the equilibrium shifts to the initiation-incompetent form and must transition to the initiation competent form before productive initiation can occur. Additionally, the temperature dependence of k_i^+ at each position defines the activation enthalpy $\Delta H_i^{o\ddagger}$, and that intermediate extension steps 5-9 have much lower $\Delta H_i^{o\ddagger}$ than

expected. The values of $\Delta H_i^{o\ddagger}$ do not correlate with k_i^+ and so indicated an enthalpically favorable process occurring during these extension steps, offset by an entropic barrier at extension steps with low k_i^+.

In chapter 3 we leveraged our knowledge of k_i^+ determined in chapter 2 in the absence of added PPi to determine the stepwise rate constants for pyrophosphorolysis (k_i^-) at the λP_R promoter. From this analysis we determined that the rate of pyrophosphorolysis is faster in initiation than elongation, and that the same dependence on the identity of the 3' dinucleotide of the RNA exists in both transcription phases.

The level of pyrophosphate within *E. coli* is predicted to change dramatically with growth phase, up to low mM concentrations during exponential growth [1, 2]. At these levels, our rate constants predict significant levels of pyrophosphorolysis in initiation, slowing down productive transcription ~14-fold at 1 mM added PPi compared to the 0 mM added PPi condition.

Considerations and Future Directions

The model λP_R promoter studied here is well-suited to bulk kinetic studies of this kind, as it forms a very long-lived open complex at 37 °C [3, 4], and at lower temperatures and NTP concentrations demonstrates initiation kinetics that result in well-defined transients at nearly every RNA length to the point of promoter escape [3]. These properties make it an excellent choice for further studies of these effects to determine a general mechanism for initial transcription and promoter escape.

The effect of ITR sequence on stepwise promoter escape and initial transcription kinetics is an intriguing avenue for further study. Variations in the sequence of the λP_R

ITR are a logical next step in determining if ITR sequence changes the RNA length needed to break individual promoter contacts. It is also known that certain ITR sequences induce pausing during and initiation [5-7], and the full effect of these sequences on promoter escape kinetics can be best determined within the context of well-defined initial transcription kinetics such as those reported here.

During formation of the λP_R OC 13 bp of duplex DNA are opened to form the initiation bubble [8]. The upstream edge of this bubble is held open by -10 element contacts with σ^{70} [8]. An additional base pair of downstream DNA is melted for each of the 9 translocation steps that occur prior to or during promoter escape [9]. A total of 21 bp of duplex DNA are therefore melted during initiation. In elongation, the transcription bubble is consistently 9 – 10 bp [10]. Collapse of the upstream edge of the initiation bubble must therefore occur at some point in promoter escape. A concerted re-formation of ~11 bp of duplex DNA should have a relatively large negative enthalpy, and may reasonably be expected to provide a driving force for promoter escape. However, studies of promoter escape from heteroduplex promoters that do not reform a duplex indicate no such effect [11]. The most reasonable explanation for the small $\Delta H_i^{o\ddagger}$ observed for intermediate extension steps 5-9 in chapter 2 is incremental collapse of the initiation bubble during these initial transcription steps. A major question that remains in this analysis, however, is how collapse could start at a point in initial transcription where the -10 contacts are predicted to still be in place, presumably preventing duplex reformation at the upstream edge of the initiation bubble. We proposed in chapter 2 that bubble collapse starts at the upstream edge of the RNA-DNA duplex and propagates upstream as the RNA-DNA hybrid is translocated into the cleft. Fluorescence studies at specific regions of the

initiation bubble are a powerful tool that has already been used to measure initiation kinetics [12, 13]. Further application of these techniques can be used to probe the mechanism of bubble collapse in escape. Additionally, the continued generation of structures for intermediates of initiation by cryo-electron microscopy (cryo-EM) makes it more and more likely that these initial transcription intermediates will be captured, in the same way intermediates of OC formation with a partially open bubble were identified [14].

The *in vivo* regulatory consequences of initiation pyrophosphorolysis are not known, despite the large difference in $[PPi]_{cyto}$ between the log and stationary growth phases in *E. coli* [1, 2]. We have established the same pyrophosphorolysis sequence dependence in initiation as in elongation. To generalize these findings, the kinetics of pyrophosphorolysis for a diverse set of promoters and ITRs should be studied. Once the generalizability of the PPi effects are known, these parameters can be extended to the known set of ITR sequences in *E. coli* to determine promoter-specific PPi regulation.

For initial transcription in both the forward and reverse directions, the most significant outstanding question is how general the kinetics and underlying mechanisms reported here are. The best to determine this generalizability is by repeating these studies for a large, diverse set of promoters. However, the low-throughput nature of these studies could render this impractical, and a high-throughput method of studying stepwise kinetics of multiple sequences is preferable. Similar studies of DNA polymerase have been performed with the aid of zero-mode waveguide (ZMW) technology, which allows for the time-resolved detection of fluorescence from a single NTP-addition event at single-molecule resolution [15-17]. In an analogous method to single-molecule sequencing, each NTP can be fluorescently labelled with a different fluorophore, allowing for the

simultaneous collection of sequence and kinetic information. The potential scalability of this method is on par with similar sequencing methods, which are capable of detecting tens of thousands of sequences per read [18]. As a result, the sequence space coverage possible far outpaces what can be accomplished by the low-throughput methods reported here. This method would also theoretically offer much more precise kinetic information than the current methods used to obtain high-throughput kinetic information, which relies on reaction quenching, reverse transcription, and subsequent sequencing and therefore loses resolution and is better suited for detecting long-term pausing, such as that caused by the elementary pause sequence [5]. Potential complications (both practical and monetary) include attaining efficient OC formation and distribution on a ZMW chip, synthesis of the required fluorophore-labelled NTPs, and optimization of detection systems for an RNA extension reaction instead of DNA extension. Development of a high-throughput method to obtain stepwise initiation kinetic data would provide the most straightforward path to a full understanding of sequence and promoter effects on initial transcription kinetics.

The Discriminator Distribution and Effects on Productive Initiation kinetics

Chapters 4 and 5 primarily focus on the initiation kinetics of several promoter constructs with non-optimal discriminator lengths. Additionally, we determined the distribution of discriminator lengths *in vivo*.

In chapter 4, we constructed two hybrid promoters, R(R) and L(R), which have the 8 bp rrnbp1 discriminator and the same λP_R ITR as in chapters 2 and 3, and either the rrnbp1 or λP_R promoter elements (UP element, -35 and -10 hexamers, and spacer). When all 4 NTPs are present, transcription starts entirely from the expected +1 positions at the

R(R) promoter and is split between the predicted +1 and -3 positions at the L(R) promoter. This difference in TSS selection confirms that the discriminator alone does not determine the TSS for the rrnbp1 promoter. There are notable similarities between the length and sequence of the L(R) spacer and previously reported mutations to the rrnbp1 spacer that induces TSS selection at the -3 position and are the likely cause of the TSS discrepancy between R(R) and L(R) [19].

The kinetics of long (post-escape) RNA initiation at R(R) and L(R) could only be determined at 37 °C and with no added heparin due to the instability of the OC formed at these promoters. Because no runoff transcripts were observed, the experiments are still clearly single-round despite the omission of heparin. At R(R), the rate constant describing long RNA synthesis (k_{8+}) is highly dependent on the concentration of UTP, the second initiating nucleotide (iNTP). That same dependence is not observed for the L(R) promoter. Priming with the dinucleotide CpA which binds to positions -1 and +1 overcomes this iNTP dependence.

In chapter 5, the hybrid promoters T(T) and L(T), which have the 7 bp T7A1 discriminator and the same λP_R ITR as in chapters 2 and 3, and either the T7A1 or λP_R promoter elements (UP element, -35 and -10 hexamers, and spacer) were constructed. These promoter constructs form more stable OC than those made with the rrnbp1 discriminator [4], and so long RNA synthesis kinetics could be determined at 19 °C, 25 °C, and 37 °C. Both T(T) and L(T) showed a dramatic UTP dependence of long RNA synthesis kinetics at 37 °C compared to 25 °C and 19 °C, much larger in magnitude than the analogous effect at λP_R seen in chapter 2.

For T(T), composite forward rate constants k_i^+ could not be determined at 37 °C, and were not well-determined at 25 °C due to fewer detected short transient RNA detected than for analogous studies of λP_R in chapter 2. However, k_i^+ was well determined for each step up to 11-mer synthesis at 19 °C. The pattern of fast and slow rate constants k_i^+ indicate two phases of initial transcription where RNAP-promoter contacts are likely being broken. Additionally, interpreting small k_i^+ parameters as indicating the presence of translocation stress unique to initiation also indicates that escape at the T(T) promoter likely occurs after 8-mer synthesis, not 7-mer synthesis as previously predicted. Despite the presence of many significant detectable transient RNA at 19 °C for the L(T) promoter, stepwise kinetic parameters could not be determined with the minimal mechanism used to describe λP_R and T(T).

The T(T), R(R), and associated hybrid promoters L(T) and L(R) demonstrate large initiation kinetic difference between promoters with different discriminators. To determine if similar studies should be conducted with shorter than optimal discriminators, we utilized a basic motif-prediction algorithm to determine the most probably -10 element at each TSS in a database of >1800 *E. coli* σ^{70} promoters [20]. At promoters that had been both computationally predicted and experimentally detected, a roughly equivalent amount of short (4 bp and 5 bp) discriminators were predicted as long (7 bp and 8 bp). However, promoter constructs produced with these discriminators did not start at the predicted TSS except in their fully native context.

Considerations and Future Directions

Discriminator length influences key initiation parameters like open complex stability, and as seen at the T(T), L(T), R(R), and L(R) promoters this is also reflected in

initiation kinetics. However, just as for OC stability [4], the discriminator alone is insufficient to fully recapitulate the initiation behavior of the original promoter. These promoters offer an excellent opportunity for future studies of initiation kinetics at promoters with non-optimal discriminators. A major limitation of the studies presented here is that although large differences in initiation kinetics are seen, it is difficult to deconvolute the effects of discriminator length, sequence, and TSS selection on these difference between promoters. We have attempted to mitigate this by maintaining the same ITR between each promoter construct, thereby controlling for ITR sequence effects on initiation, but it is clear that promoter elements do not act fully independently of ITR sequence.

The immediate continuation of this research is to determine stepwise extension rate constants for the T(T), L(T), R(R), and L(R) promoters at multiple temperatures and NTP conditions. Although the standard initiation experiment conditions did not facilitate this, variations in NTP concentrations or other conditions can be used to induce additional transient short productive RNA, allowing for the determination of stepwise rate constants at these promoters. By determining the temperature effect on initiation kinetics for promoters that form unstable OC, we can begin to try to determine general effect of OC stability on the mechanisms that underlie promoter escape like the breaking of promoter contacts and bubble collapse.

The mechanism used in the determination of stepwise initiation parameters for λP_R is relatively straightforward. Each step clearly corresponds to an easily assigned event in initial transcription. However, although this mechanism describes the T(T) promoter well, several questions remain. The minimal mechanism developed to describe λP_R does not

describe the L(T) promoter. For T(T), like λP_R, an initial equilibrium step prior to NTP binding is required to accurately fit the data. Unlike λP_R, at 19 °C this step is biased towards the initiation-incompetent OC form. Although fits for T(T) at 25 °C are not well-determined towards longer lengths, at shorter lengths the fit is acceptable, and the pre-NTP binding equilibrium step is fit to close to a true equilibrium. If this trend in the bias of the initial equilibrium continues to 37 °C, the extremely low k_{FL} for T(T) at the low UTP condition could not be explained in the same way as at λP_R. Experiments at different temperatures between 25 °C and 37 °C at the T(T) promoter and low UTP condition will help determine at what temperature the overall rate of promoter escape begins to decrease, and help determine the origin of this effect.

For R(R) and L(R), stepwise kinetic information is difficult to obtain because these promoters do not initiate at 19 °C (under the conditions investigated here) where the most transient short RNA are detectable at other promoters. This limitation could potentially be circumvented by higher NTP concentrations, which would stabilize the R(R) and L(R) OC. However, the corresponding increase in transcription rate would need to be accounted for by taking faster timepoints with a higher concentration of radiolabelled NTP, which is within the capabilities of the RQF instruments used in these assays. These rate constants are necessary to confirm and extend the two-step mechanism used to describe rrnbp1 initiation in chapter 4.

It is unknown what portion of the short (4-5 bp) discriminators predicted here are active promoters *in vivo*, but this information is vital for determining the importance of short discriminators within the cell. As the majority of short discriminators detected were at promoters that had exclusively been annotated via computational prediction, there are

two major possibilities. The first is that these are promoter-like features of the *E. coli* genome that do not actively transcribe under normal conditions, but act as RNAP binding sites to globally regulate transcription in some way. Alternatively, these features may be promoters that are not significantly transcribed from without the presence of an activator that is not expressed under normal growth conditions. Additional high-sensitivity transcriptomic assays are often added to annotated *E. coli* TSS databases such as the RegulonDB database used here, and it is likely that as these experiments are performed at more and more diverse conditions, the function of many of these putative TSS with short discriminators will become apparent.

There is a significant amount of short discriminators that have confirmed TSS, and the development of model promoters with which to study the processive initiation kinetics is the next major step in understanding if, and how, this promoter feature is relevant to initiation regulation. Unlike 7 bp and 8 bp discriminators, hybrid promoters could not be readily constructed by simply replacing the λP_R discriminator with a short discriminator from a known promoter. This is most likely because the long discriminators studied here have CTP at the -1 (and -2, for the 8 bp rrnbp1 discriminator) position. This prevents productive initiation from shorter discriminator lengths when CTP is withheld from the reaction mixture and allows study of initiation kinetics from a single TSS. This same strategy cannot be used at short discriminators for which some portion of initiation occurs at a downstream TSS, as it would prevent all productive initiation from occurring at the TSS of interest. The kinetics of initiation from short discriminators must therefore be studied in the native context of that discriminator. The stability of the OC formed by promoters with short discriminators, which can be investigated with filter-binding assays,

should be studied concurrently with their escape kinetics to further our understanding of the link between OC stability and promoter escape.

Final Considerations

The cell devotes significant resources to regulating initiation through transcription factors, small molecules, NTP concentration, and promoter sequence and architecture. In this thesis we used detailed kinetic parameters to infer mechanistic details about promoter escape at the model λP_R promoter. We also applied those same techniques to other model promoters that form less stable OC, and investigated the distribution of discriminator length in the *E. coli* genome. The detailed kinetic parameters reported here are the first of their kind for a processively initiating complex and have revealed novel insights into the promoter escape mechanism. I eagerly anticipate further advances in this field that will continue to move us towards to a complete understanding of promoter sequence control of initiation kinetics.

References

[1] Kukko-Kalske E, Lintunen M, Inen MK, Lahti R, Heinonen J. Intracellular PPi concentration is not directly dependent on amount of inorganic pyrophosphatase in Escherichia coli K-12 cells. Journal of Bacteriology. 1989;171:4498-500.

[2] Kukko E, Heinonen J. The Intracellular Concentration of Pyrophosphate in the Batch Culture of Escherichia coli. European Journal of Biochemistry. 1982;127:347-9.

[3] Henderson KL, Evensen CE, Molzahn CM, Felth LC, Dyke S, Liao G, et al. RNA Polymerase: Step-by-Step Kinetics and Mechanism of Transcription Initiation. Biochemistry. 2019;58:2339-52.

[4] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2017;114:E3032-E40.

[5] Winkelman JT, Pukhrambam C, Vvedenskaya IO, Zhang Y, Taylor DM, Shah P, et al. XACT-Seq Comprehensively Defines the Promoter-Position and Promoter-Sequence Determinants for Initial-Transcription Pausing. Molecular Cell. 2020;79:797-811.e8.

[6] Lerner E, Chung S, Allen BL, Wang S, Lee J, Lu SW, et al. Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. Proceedings of the National Academy of Sciences. 2016;113:E6562-E71.

[7] Bauer DLV, Duchi D, Kapanidis AN. E. Coli RNA Polymerase Pauses during Initial Transcription. Biophysical Journal. 2016;110:21a.

[8] Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. Molecular Cell. 2015;58:534-40.

[9] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006;314:1144-7.

[10] Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. 2007;448:157-62.

[11] Ko J, Heyduk T. Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochemistry Journal. 2014;463:135-44.

[12] Jensen D, Manzano AR, Rammohan J, Stallings CL, Galburt EA. CarD and RbpA modify the kinetics of initial transcription and slow promoter escape of the Mycobacterium tuberculosis RNA polymerase. Nucleic Acids Research. 2019;47:6685-98.

[13] Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS, Belogurov GA. Active site opening and closure control translocation of multisubunit RNA polymerase. 2012;40:7442-51.

[14] Chen J, Chiu C, Gopalkrishnan S, Chen AY, Olinares PDB, Saecker RM, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. Molecular Cell. 2020;78:275-88.e6.

[15] Vilfan ID, Tsai Y-C, Clark TA, Wegener J, Dai Q, Yi C, et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. Journal of Nanobiotechnology. 2013;11:8.

[16] Schadt EE, Banerjee O, Fang G, Feng Z, Wong WH, Zhang X, et al. Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. Genome Research. 2013;23:129-41.

[17] Korlach J, Bibillo A, Wegener J, Peluso P, Pham TT, Park I, et al. Long, Processive Enzymatic DNA Synthesis Using 100% Dye-Labeled Terminal Phosphate-Linked Nucleotides. Nucleosides, Nucleotides and Nucleic Acids. 2008;27:1072-82.

[18] Weirather JL, De Cesare M, Wang Y, Piazza P, Sebastiano V, Wang X-J, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Research. 2017;6:100.

[19] Winkelman JT, Chandrangsu P, Ross W, Gourse RL. Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. Proceedings of the National Academy of Sciences. 2016;113:1787-95.

[20] Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation inE. coliK-12. Nucleic Acids Research. 2019;47:D212-D20.