MECHANISM OF DNA SCRUNCHING DURING INITIAL TRANSCRIPTION

by

ADAM HASEMEYER

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

Written under the direction of

Prof. Richard H. Ebright

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2018

ABSTRACT OF THE DISSERTATION

Mechanism of DNA scrunching during initial transcription

By ADAM HASEMEYER

Dissertation Director:

Richard H. Ebright

In bacteria, initial transcription occurs through a scrunching mechanism, where RNA Polymerase (RNAP) remains stationary on promoter DNA, unwinding and pulling downstream DNA into itself and past its active center. The incorporation of additional nucleotides into the active center cleft leads to expansion of the single-stranded transcription bubble, with the accumulated DNA on each strand proposed to be accommodated as single-stranded bulges in the unwound region. The location of these single-strand DNA bulges as they proceed to increase in size during initial transcription has not been fully elucidated.

In this work, we have used single-molecule fluorescence resonance energy transfer (FRET) to detect and analyze the path of nontemplate strand (NT) DNA at multiple increased levels of scrunched DNA. We have also defined the positions of scrunched NT DNA during scrunching using FRET-derived distance restraint docking onto structural models of RNAP.

We have prepared static initial transcription complexes (ITCs) with iteratively increasing scrunched states in the NT strand: RP_o, 2 nucleotides (nt) scrunched, 4 nt scrunched, 6 nt

scrunched, and 8 nt scrunched. The complexes were shown to be properly formed and fully functional in transcription.

In this work, we have demonstrated the specific incorporation of fluorescent probes at intended labeling sites in both RNAP and NT strand DNA, and have demonstrated the resulting DNA:RNAP derivative complexes are functionally functional. We have determined probe-probe single-molecule FRET distances for each labeled DNA:RNAP combination as part of each static scrunched NT strand complex, totaling 160 unique combinations.

By combining the FRET results with distance restraint docking methodology, we have established models of NT strand DNA in context of RP_o, 2 nt scrunched, 4 nt scrunched, 6 nt scrunched, and 8 nt scrunched. The RP_o models were compared to solved RP_o crystal structures to validate our methodology. Additionally, these models were formed for scrunched complexes containing either a consensus (DSC) discriminator sequence or anticonsensus (aDSC) discriminator sequence. Models were analyzed to both determine the path of NT strand DNA as it proceeds through higher levels of scrunching and to compare the different pathways seen between DSC and aDSC complexes.

Our work showed that the models fit well with crystals structures of RP_o. It also showed that scrunched NT strand DNA can be accommodated within the active center cleft up to at least 6-8 nt of scrunched DNA. Comparison of structures demonstrated that aDSC complexes have more flexibility in their possible locations within the active center cleft during scrunching and they may exit from the active center cleft at an earlier scrunched state.

ACKNOWLEDGMENT

I would like to express my thanks to my advisor, Dr. Richard Ebright, for his guidance, support, and patience during this work. I encountered many obstacles over the course of my time in the lab, some small, some tremendous, and his insightful ideas and helpful advice were always able to steer me in the right direction. For that I will always be grateful.

I would like to thank my committee members, Dr. Bryce Nickels, Dr. Eddy Arnold, and Dr. Smita Patel for their time and support in the development and refinement of my project.

I would like to thank Dr. Anirban Chakraborty for his strong guidance when I first arrived at lab. I relied heavily on his support and knowledge early on, and he was always there to lend an ear or a hand.

I would like to thank Dr. Yu Zhang for his collaborative work with this project. His crystallographic studies provided a crucial comparative framework for my project.

I would like to thank Dr. Abhishek Mazumder for his suggestions and thoughtful discussions during my time in the lab.

I would like to thank all of the members of the Ebright lab with whom I shared my time, especially Dr. Yon Ebright, Dr. Yu Feng, Dr. Juan Shen, Dr. Sergey Druzhinin, Dr. David Degen, Dr. Qumiao Xu, Ms. Shuang (Carol) Liu, and Dr. Miaoxin Lin.

I would like to thank my wife, Sonia Hasemeyer, for her lasting love and support through this time. She is my shelter through the storm.

I would like to thank my mother for her consistent love and all that she has sacrificed to raise me into the person I am today.

Finally, I would like to thank my son for all the joy he brings my life. There is no greater force to lift my spirits than the power of his hug.

DEDICATION

Dedicated to my wife Sonia

Table of contents

List of tables

List of illustrations

1. Introduction

Transcription, the process of DNA directed RNA synthesis, is the initial step in gene expression and is required for survival by all living organisms. Regulation of gene expression occurs most frequently during transcription. Since transcription is performed by RNA Polymerase (RNAP), this enzyme is a primary target for transcriptional regulators. The five core proteins that make up RNAP are conserved in bacteria (β, β′, α^{I}, α^{II}, and ω), archaea, (A, B, D, L, and K) and eukarya (RPB1, RPB2, RPB3, RPB11, and RPB6 for eukaryotic RNAP II), with additional subunits found in the higher order organisms (Cramer, 2002; Cramer et al., 2008; Darst, 2001; Ebright, 2000; Werner, 2007; Werner and Grohmann, 2011). The individual subunits and the entire bacterial RNAP enzyme have been extensively characterized, providing a solid foundation for its use as a model organism in further transcriptional studies.

1.1. Bacterial RNA polymerase core

The core bacterial RNAP enzyme consists of five subunits (α^{I}, α^{II}, β, β′, and ω), has a molecular mass of ~380 kDa, and encompasses dimensions of ~150 Å x ~100 Å x ~100 Å. The core enzyme is able to carry out all steps of transcription—initiation of transcription, elongation of the nascent RNA, termination of RNA synthesis—but is unable to identify sequence specific promoter sites in DNA (Burgess et al., 1969; Travers and Burgessrr, 1969). The structure of RNAP core was originally determined in *Thermus aquaticus* at a resolution of 3.3 Å (Darst et al., 2002; Zhang et al., 1999). Since then,

several additional structures have provided further insights into the enzyme's structure and function (Darst, 2001; Darst et al., 2002; Murakami and Darst, 2003; Vassylyev et al., 2002; Zhang et al., 2012). Bacterial RNAP resembles a crab claw in shape with two distinct pincers formed by the β and β′ subunits (Figure 1A). β, the second largest subunit, forms the other pincer. The two pincers define a channel ~ 27 Å wide, which forms the active center cleft of the enzyme. At the base of the cleft is the active site, where a Mg^{2+} ion crucial to catalytic activity remains chelated by three conserved aspartate residues. The active site is the location where RNA synthesis from ribonucleoside triphosphate (NTP) occurs.

Catalytic nucleotide addition during RNA synthesis in RNAP occurs by way of a two-metal-ion reaction mechanism, leading to phosphodiester bond formation. The first metal ion is responsible for reducing the affinity of the 3'-OH for its hydrogen atom, promoting nucleophilic attack on the α-phosphate. The role of the second metal is assisting the removal of the released pyrophosphate from the active site. While the first Mg^{2+} ion remains at the active site, the second Mg^{2+} ion is introduced to the active site by the incoming NTP (Steitz, 1998).

There are two other distinct channels in RNAP core that extend from the active center to the surface of the enzyme (Figure 1B). One is the secondary channel, or NTP-entrance channel, which mediates access of NTP substrates, inhibitory molecules, and regulatory factors, such as Gre factors, to the active center cleft (Mukhopadhyay et al., 2004; Nickels and Hochschild, 2004). The other is the RNA-exit channel, which mediates the departure of nascent RNA from the active center cleft. While the active center can hold a

DNA/RNA hybrid up to 9 nucleotides (nt) in length, any further elongation of the nascent RNA strand necessitates entry into the RNA exit channel, which is encapsulated by the β flap domain (Komissarova and Kashlev, 1998; Vassylyev et al., 2002). Up to 5 nt of RNA can be protected with the RNA exit channel before it emerges from the enzyme into the surrounding solution (Vassylyev et al., 2007).

The largest subunit of RNAP, β′ (~155 kDa), has multiple functions during transcription. It forms on of the pincers, designated the "clamp". The clamp is able to open and close during different points during transcription, permitting DNA into the active-center cleft and maintaining its position during transcription elongation (Chakraborty et al., 2012). An additional domain, the β′ jaw forms stabilizing interactions with double-stranded DNA downstream of the promoter location (+10 to +20) during initial transcription (Mekler et al., 2011). The β′ subunit comprises a majority of the active center and thus holds much of the determinants for catalysis.

The second largest subunit of RNAP is β, which forms the other pincer and contains sequence determinants for interactions with both DNA and RNA during transcription. It forms the remaining portion of the active center of RNAP and contains a domain labeled the β lobe, which is involved in DNA unwinding to initiate transcription (Brodolin et al., 2005). Both the β and β′ subunits together form the switch region, found at the base of the β′ pincer, which mediates opening and closing of the active center cleft (Mukhopadhyay et al., 2008; Srivastava et al., 2011).

Bacterial RNAP core contains two copies of α subunits: α^I and α^I, which are identical in sequence, but differ in their interactions with the rest of RNAP. α^I interacts with the β

subunit, while α^{II} interacts with the β′ subunit (Zhang et al., 1999). Each α subunit forms two distinct domains, the N-terminal domain (αNTD) and the C-terminal domain (αCTD). These two independently-folded domains are interconnected by a flexible linker of ~20 amino acids (Ebright and Busby, 1995). αNTD dimerizes and the α^{I}NTD-α^{II}NTD dimer acts as a scaffold for assembly of the β and β′ subunit in RNAP core (Darst et al., 1998). αCTD is a DNA-binding module that interacts with upstream promoter elements, termed UP elements, and transcriptional regulators that bind DNA (Lawson et al., 2004; Ross et al., 1993).

The smallest subunit of RNAP, ω, is not required for function of RNAP or bacterial growth under normal conditions. However, it has been shown to act as a molecular chaperone to promote the assembly of the RNA polymerase by fastening the N- and C-terminal regions of unfolded β′, facilitating its folding (Minakhin et al., 2001).

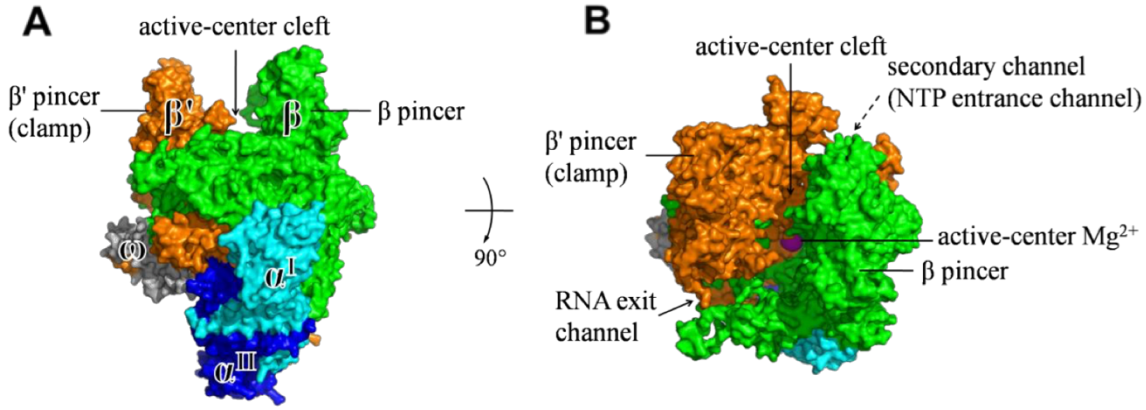

Figure 1. RNAP Core
Crystal structure of bacterial RNAP Core (Zhang et al., 1999). β′ is shown in orange; β is shown in green; α^{I} is shown in light blue; α^{II} is shown in dark blue; ω is shown in gray. (A) Upstream face (B) Top face, rotated from **A** 90° about the x-axis with a view into the active-center cleft

1.2. Bacterial RNA polymerase holoenzyme

RNAP core is able to perform non-specific transcription initiation, but is unable to discriminate between non-promoter and promoter-specific sites. When core RNAP is combined with another subunit, σ, the resulting holoenzyme is able to locate and bind to specific promoter sequences and initiate transcription at these sites. The determinants for the sequence-specific promoter interactions are contained within σ (Busby and Ebright, 1994; Gross et al., 1998). In addition to promoter sequence recognition, σ plays a significant role in promoter unwinding, promoter escape, and transcriptional pausing (Busby and Ebright, 1994; Gross et al., 1998; Helmann and Chamberlin, 1988). While most bacterial species have multiple σ factors, one principal factor is responsible for the majority of transcription in the cell. For *E.coli,* the primary σ factor is a 70 kDa protein called σ^{70}.

Promoter recognition is the primary function of σ factors and bacterial promoters contain two core sequence elements: the -10 element and the -35 element, each named for their location relative to the transcription start site (Campbell et al., 2002; Gross et al., 1998; Harley and Reynolds, 1987; Malhotra et al., 1996; Severinova et al., 1996). The -10 element has a consensus sequence 5'-TATAAT-3' and the -35 element has a consensus sequence 5'-TTGACA-3'. The similarity of a promoter's sequence to the consensus -35 and -10 element sequences plays a large factor in the strength of the promoter. An additional promoter sequence element found in some promoters is the extended -10 element, which is located immediately upstream of the -10 element and has a consensus

sequence 5'-TGN-3' (Mitchell et al., 2003). Promoters with a consensus extended -10 element are able to initiate transcription productively without a recognizable -35 element.

The principal factor σ^{70} in *E.coli* (σ^{A} in both *T. thermophilus and T.aquaticus*) contains five conserved regions: σ-region-1.1 (σR1.1), σ-region-2 (σR2), σ-region-3 (σR3), σ-region-3.2 (σR3.2 or σR3/σR4 linker), and σ-region-4 (σR4) (Gross et al., 1998; Young et al., 2002). σR2, σR3, and σR4 are structurally ordered domains that contain determinants for sequence-specific interactions with the promoter -10 element, the promoter extended -10 element, and the promoter -35 element, respectively (Campbell et al., 2002; Gross et al., 1998; Malhotra et al., 1996; Severinova et al., 1996; Young et al., 2002). The other conserved regions, σR1.1 and the σR3/σR4 linker are unstructured, flexible, and highly negatively charged (Campbell et al., 2002).

Although σ holds the elements required for promoter DNA interaction, it is incapable of binding with promoter DNA without RNAP core. Association with RNAP core produces a conformation change in σ and exposes the DNA-binding elements (Callaci et al., 1998; Callaci and Heyduk, 1998; Dombroski et al., 1993; Dombroski et al., 1992). Upon formation of holoenzyme, all the promoter-recognition determinants in σ are solvent exposed and spaced consistent with the predicted separation of their target promoter elements (Murakami, 2013; Murakami and Darst, 2003; Murakami et al., 2002a; Murakami et al., 2002b).

The structures of *T. thermophilus*, *T. aquaticus*, and *E. coli* RNAP holoenzyme were determined at 2.6 Å resolution, 4.0 Å resolution, and 3.7 Å resolution, respectively (Murakami, 2013; Murakami et al., 2002a; Murakami et al., 2002b; Vassylyev et al.,

2002). In addition, the positions of σR1.1, σR2, σR3, σR3.2, and σR4 relative to RNAP core in *E. coli* RNAP holoenzyme were determined using systemic FRET (Mekler et al., 2002). The structural organization of core subunits in RNAP core and RNAP holoenzyme are nearly identical, except for rotation of the β′ clamp depending on its closed or open conformational state (Chakraborty et al., 2012).

The interactions between the core subunits and σ in RNAP holoenzyme are extensive (Figure 2) (Mekler et al., 2002; Murakami et al., 2002a; Murakami et al., 2002b; Sharp et al., 1999; Vassylyev et al., 2002; Young et al., 2002). The globular domains of σ run across the upstream face of the pincer structure. σR2 interacts with the β′ pincer within and above the RNAP active-center cleft and is made up of three α-helices. One of the α-helices forms contacts with the β′ subunit, while another of the α-helices is involved in DNA melting and -10 element promoter recognition (Vassylyev et al., 2002). σR3 interacts with the base of the β flap and is made up of three α-helices, one of which makes contacts with upstream DNA. σR4 interacts with the tip of the β flap and is made up of two α-helices, one of which makes contacts with promoter DNA between the -30 and -38 positions. In addition, the flexible, negatively charged σ regions, σR1.1 and σR3.2, form interactions with core subunits. Both serve as molecular mimics for DNA and RNA: σR1.1 is located in the RNAP active-center cleft and must be displaced to permit access of promoter DNA to the active-center cleft; and σR3.2 is located in the RNA-exit channel and must be displaced to permit access of nascent RNA to the RNA exit channel (Figure 2B).

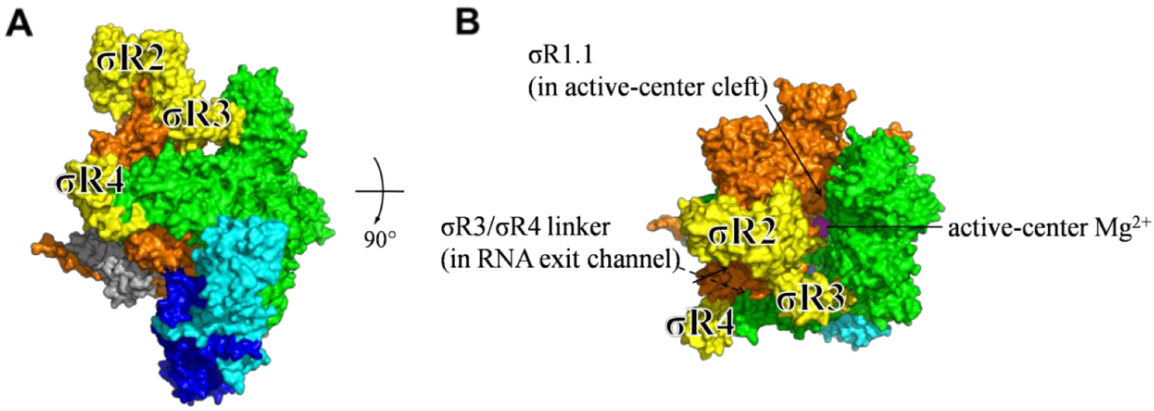

Figure 2. RNAP Holoenzyme
Crystal structure of bacterial RNAP Holoenzyme (Vassylyev et al., 2002). β′ is shown in orange; β is shown in green; α^{I} is shown in light blue; α^{II} is shown in dark blue; ω is shown in gray; σ^{70} is shown in yellow. (A) Upstream face (B) Top face, rotated from **A** 90° about the x-axis with a view into the active-center cleft

1.3. Transcription cycle

There are three steps to the transcription cycle: initiation, elongation, and termination. RNAP holoenzyme and promoter DNA must undergo subsequent reactions to move through transcription initiation, into transcription elongation, and finally transcription termination:

i. **RP_c (RNAP-promoter closed complex) –** RNAP holoenzyme binds DNA at the promoter site recognized by σ; the enzyme makes initial interactions with double-stranded DNA (dsDNA), which remains outside the active center cleft (Bai et al., 2006; deHaseth et al., 1998).

ii. **RP_o (RNAP-promoter open complex)--** approximately 13 base pairs of DNA are unwound at the transcription start site (-11 to +2 positions with respect to the transcription start site), forming a single-stranded DNA region, known as the transcription bubble, within the RNAP active center cleft (Bai et al., 2006; deHaseth et al., 1998).

iii. **RP_{itc} (RNAP-promoter initial transcribing complex) --** RNAP begins cycles of transcribing abortive RNA products 2-9 nucleotides in length while remaining bound to the promoter region of the DNA, a process known as abortive synthesis. As the RNA lengthens, more DNA is pulled into the active center cleft through a scrunching mechanism that creates a stressed intermediate (Hsu, 2002; Kapanidis et al., 2006; Mooney et al., 2005; Revyakin et al., 2006).

iv. **RD_e (RNAP-DNA elongation complex) –** once RNAP has formed an RNA product greater than ~ 9-11 nucleotides in length, enough energy has been built up to break from the interactions with the promoter DNA. Many of the interactions between core RNAP and σ are weakened as well, allowing for their eventual dissociation as the elongation complex proceeds. The enzyme then continues to transcribe RNA by moving along the DNA template in a stepwise fashion, in which RNAP translocates relative to DNA, one base pair at a time, for each nucleotide addition cycle. Elongation complexes are highly stable and highly processive (Korzheva and Mustaev, 2001).

v. Upon encountering a termination signal, RNAP, DNA, and RNA all dissociate from each other. The core RNAP is then able to reform holoenzyme and begin transcription again at the promoter region.

1.4. DNA scrunching during initial transcription

Whereas RNAP moves along the DNA stepwise during transcription elongation, initial transcription occurs by a different mechanism. During initial transcription, DNA undergoes several cycles of synthesis and release of short abortive RNA products while still maintaining contact with promoter DNA. RNAP is unable to form a stable transcription elongation complex (TEC) until it is able to release itself from the RNAP-promoter contacts (Murakami et al., 2002a; Murakami et al., 2002b; Revyakin et al., 2006; Werner and Grohmann, 2011). DNA footprinting of RNAP-promoter initial transcribing complexes showed that the upstream promoter boundary protected by RNAP does not change from promoter binding to initial transcribing complexes producing abortive transcripts (Carpousis and Gralla, 1985; Krummel and Chamberlin, 1989; Straney and Crothers, 1987). While RNAP remains bound to the upstream promoter region, its active center is able to move along the DNA, transcribing RNA transcripts up to 9-11 nucleotides in length.

Using single molecule FRET and DNA nanomanipulation experiments, it was determined that initial transcription occurs by means of a DNA scrunching mechanism, where RNAP remains stationary on promoter DNA and unwinds and pulls downstream DNA into itself and past its active center (Kapanidis et al., 2006; Revyakin et al., 2006). The scrunching

action leads to a stressed intermediate with two additional DNA nucleotides, one for each unwound strand, pulled into RNAP per each RNA nucleotide addition (Figure 3). The incorporation of additional nucleotides into the active center cleft leads to expansion of the single-stranded transcription bubble, with the accumulated DNA on each strand proposed to be accommodated as single-stranded bulges in the unwound region (Kapanidis et al., 2006).

Until recently little was known about where the bulges were accommodated within or around RNAP. It was postulated that the scrunched DNA strands would protrude out into solvent surrounding the active center cleft near position -5 to -6 of the nontemplate strand and -9 to -10 of the template strand (Kapanidis et al., 2006), or that the template strand was accommodated within the template strand tunnel (Zuo and Steitz, 2015). A recent study using crosslink mapping techniques was able to identify interactions between both the nontemplate and template strands and positions in RNAP holoenzyme in initial transcription complexes with 3 nucleotides scrunched (Winkelman et al., 2015). The template strand crosslinked to locations at the base of the RNA exit channel, suggesting its role in expulsion of σR3.2 from the channel. The nontemplate strand at the downstream region of its single-stranded DNA (ssDNA) (-3 to +6 in RP_{itc5}) crosslinked to a position β62, located at the upper boundary of the active center cleft in an accessible solvent gap between β domains 1 and 2, suggesting that it may protrude out into solvent during further scrunching states.

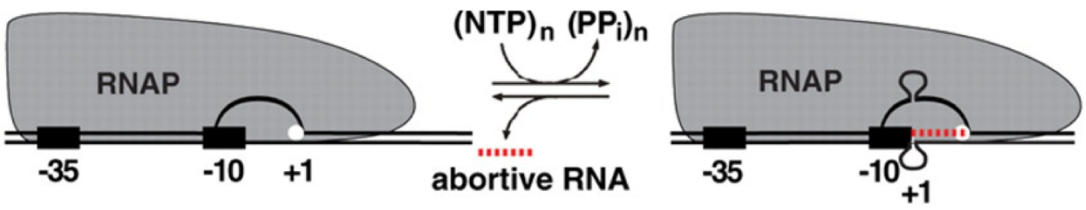

Figure 3. Scrunching mechanism of initial transcription.
Representation of scrunching mechanism leading to stressed intermediates (right), with one additional nucleotide per each unwound DNA strand pulled within RNAP per nucleotide addition after the first synthesis dinucleotide.

1.5. Discriminator element

The discriminator element is a sequence-dependent element in the nontemplate strand of promoter DNA immediately downstream of the -10 element. Originally, it was identified as a high G-C content location downstream of the -10 element that acted as a regulator for transcription initiation by increasing the difficulty of strand separation at that promoter site (Lamond and Travers, 1985; Pemberton et al., 2000; Travers, 1980). More recently it was discovered that the discriminator element, with consensus sequence of -GGGA- from -6 to -3 of the nontemplate strand relative to the transcription start site, can extend promoter lifetimes by interacting with σR1.2, stabilizing initial transcription, but also inhibiting promoter escape (Feklistov et al., 2006; Haugen et al., 2006; Haugen et al., 2008).

Crystal structure of *Thermus thermophilus* RNAP in an open complex with a consensus discriminator element elucidated many of the interfaces between the enzyme and the discriminator element (Figure 4) (Zhang et al., 2012). Contacts between σ and the discriminator region of the nontemplate strand occur along a shallow groove in σR1.2,

with positions -6 to -4 forming interactions with amino acids in σ (Figure 4A). The most upstream guanosine nucleotide (G-6) has been flipped out of its base stack with adjacent nucleotides and placed in a pocket in σR1.2, and that binding is sequence determinant (Figure 4B). Alanine substitution of multiple σ sites that make contact with discriminator element bases was shown to reduce transcription activity of the enzyme (Zhang et al., 2012). Binding of the discriminator into the pocket in σR1.2 has also been shown to play a role in transcription start site selection, in which a consensus discriminator element stabilizes the nontemplate strand DNA in the transcription bubble through σR1.2 interactions (Winkelman et al., 2016).

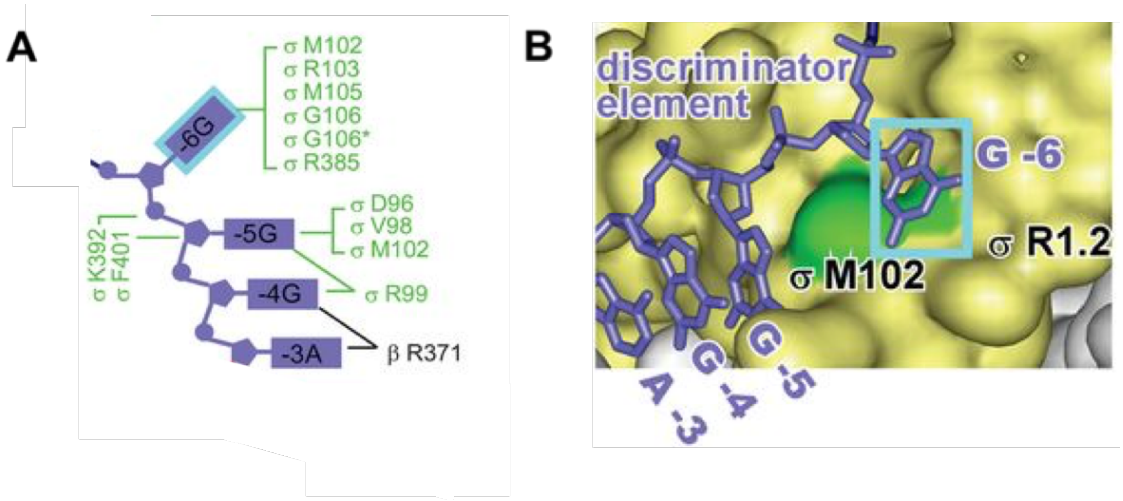

Figure 4. RNAP-discriminator element interactions in RP_o (adapted from Zhang, et al., 2012)
(A) Schematic summary of protein–nucleic-acid interactions between RNAP holoenzyme and consensus discriminator. Discriminator element of DNA nontemplate strand in blue. Black residue numbers and lines are interactions by RNAP; green residue numbers and lines are interactions by σ; asterisks are water-mediated interactions; cyan outline around G-6 base that is unstacked and inserted into pocket of σR1.2. Residues are numbered as in E. coli RNAP and σ.
(B) Interactions between σR1.2 and nontemplate-strand ssDNA of discriminator element. σ is in yellow; discriminator element is in blue; residue that crosslinks with position G–5 is in green; cyan outline around G-6 base that is unstacked and inserted into pocket of σR1.2.

1.6. Single-molecule fluorescence resonance energy transfer

Fluorescence resonance energy transfer, or FRET, is a physical phenomenon where there is a non-radiative energy transfer between two chromophores, one donor and one acceptor. FRET occurs in a system containing a fluorescent donor-acceptor pair, where there is significant overlap between the emission spectrum of the donor and the excitation spectrum of the acceptor. The steps of this system occur as follows: *(1)* upon excitation of the donor at its excitation wavelength, *(2)* energy can be non-radiatively transferred from the donor to the acceptor, *(3)* resulting in excitation of the acceptor and *(4)* emission at the acceptor's emission wavelength. This occurs through coupling of the emission transition dipole moment of the donor and absorption transition dipole moment of the acceptor when a donor fluorophore is activated to its excited electronic state. The efficiency of FRET and the effective range over which it can occur are determined by the spectral properties of a given donor-acceptor pair (Clegg, 1992).

FRET can act as a spectroscopic ruler, providing accurate measurement of distances between ~20 Å to ~ 100 Å, equivalent to approximately one-half the length of a transcription complex (Lilley and Wilson, 2000; Murakami and Darst, 2003; Selvin, 2000; Stryer and Haugland, 1967). FRET has been employed previously to define both structure and mechanisms in transcription (Chakraborty et al., 2012; Mekler et al., 2002; Mukhopadhyay et al., 2001; Mukhopadhyay et al., 2003).

The efficiency of energy transfer, *E*, is proportional to the inverse sixth power of the distance (*R*) between the donor and the acceptor, given by the equation (Clegg, 1992):

$$E = 1/[1 + (R/R_o)^6] \quad (1)$$

where R_o, the Förster distance for a given donor-acceptor pair, is the distance at which the energy transfer efficiency for a donor-acceptor pair is half maximal, or $R=R_o$ when FRET efficiency is 50%. The value of R_o depends on the spectral properties of each fluorescent probe in a donor-acceptor pair and the relative orientation of their dipole moments and is calculated by the following equation:

$$R_o = 9780(n^{-4}\kappa^2 Q_D J)^{1/6} \text{ Å} \quad (2)$$

where n is the refractive index of the medium, κ^2 is the orientation factor between the donor emission and acceptor excitation dipoles, Q_D is the fluorescent quantum yield of the donor in the absence of acceptor, and J is the integral of the spectral overlap between donor emission and acceptor absorbance spectra.

The efficiency of FRET, *E*, can be determined experimentally through multiple approaches, including measuring the decrease in fluorescent quantum yield of the donor, measuring the increase in fluorescent emission from the acceptor and decrease fluorescent emission of the donor, and measuring the decreased fluorescent lifetime of the donor (Clegg, 1992). R_o can be determined by experimentally as well and is independent of energy transfer for each donor-acceptor pair. Therefore, the donor-acceptor distance *R* can be derived from Equation 1 with the values of *E* and R_o.

FRET analysis was traditionally carried out at ensemble level, which provided mean values for the system studied, but was unable to taken into account heterogeneity within the sample. In contrast to ensemble FRET analysis, single-molecule FRET (smFRET) analysis provides measurements at the molecular level, permitting the identification and analysis of heterogeneous populations and dynamic processes within an observable FRET species (Ha et al., 1996). Individual, fluorescently-labeled molecules can be studied through two distinct methodologies: surface immobilization or free diffusion within solution (Deniz et al., 1999; Ha, 2001). Single-molecule analysis using surface immobilized molecules is particularly helpful in the detection of dynamic events or time-dependent conformational changes, since it focuses one individual molecule over a period of time. However, the immobilization itself can lead to potential changes within the natural order of the system, such as the structure and/or function of the biological complex being studied or the photophysical properties of the dyes.

Alternatively, in-solution single-molecule limits these unnatural perturbations FRET by using molecules freely diffusing in solution by virtue of Brownian motion. The donor-acceptor doubly-labeled molecules are examined at sub-nanomolar concentration using a setup containing a laser beam, at the donor excitation wavelength, focused onto a femtoliter-scale observation volume using confocal optics (Deniz et al., 1999; Schuler and Eaton, 2008). As a single molecule travels through the observation volume and is illuminated, it reaches an excited state, and subsequently releases fluorescence photon bursts corresponding to the excitation. If an acceptor is close enough in proximity to the donor, energy is partially transferred from donor to acceptor, and the acceptor also releases fluorescent photon bursts. The emitted photons from both donor and acceptor are

counted in their respective donor and acceptor detector channels. The FRET efficiency, *E*, for each molecule with a donor-acceptor pair is calculated by the ratio between the number of photons detected in the acceptor channel and the number of photons detected in both the donor and acceptor channels. The data from a large collection of events within a sample are plotted in one-dimensional FRET histograms, providing statistical insight into the static and dynamic heterogeneity of the population, as well as the presence of conformational changes and molecular interactions. However, smFRET measurements have been confined to FRET efficiencies greater than ~40% due to sub-stoichiometric complexes (donor-only or acceptor-only) and the presence of chemically or photophysically induced species that obscure FRET measurements at close range.

The use of alternating-laser excitation (ALEX) with smFRET can overcome these issues and provide a full FRET efficiency range (0-100%) (Kapanidis et al., 2005; Kapanidis et al., 2004; Lee et al., 2005) . in addition to the laser that excites the donor at donor excitation wavelength, ALEX confocal microscopy uses an additional laser that excites the acceptor at acceptor excitation wavelength. The lasers are tightly aligned, alternated at a microsecond time scale, and focused into a femtoliter-scale observation volume containing freely diffusing fluorescent single-molecules. This provides distinct donor and acceptor emission signatures, consisting of finite bursts of photons, compiled in the donor or acceptor emission detectors, for each single molecule observed.

Since ALEX allows acceptor emission to be measured independent of donor excitation, it enables the calculation of a stoichiometry parameter, *S*, which reports on the relative stoichiometry of donor and acceptor probes for each molecule observed. Incorporating the additional parameter into the results provides a two-dimensional FRET histogram, an

E/*S* plot, in which *S*, with values ranging between 0 and 1, can be used to virtually sort species that contain both donor and acceptor dyes (clustered around *S* values of 0.3 to 0.8) from donor-only (clustered around *S* values of ~1.0) species and acceptor only species (clustered around *S* values of 0 to 0.2). The results from doubly-labeled species can be isolated from donor-only and acceptor-only species, and the corresponding *E* distributions can be plotted and fitted with Gaussians. Each Gaussian represents a sperate subpopulation and the mean of each Gaussian defines the mean *E* value of that subpopulation.

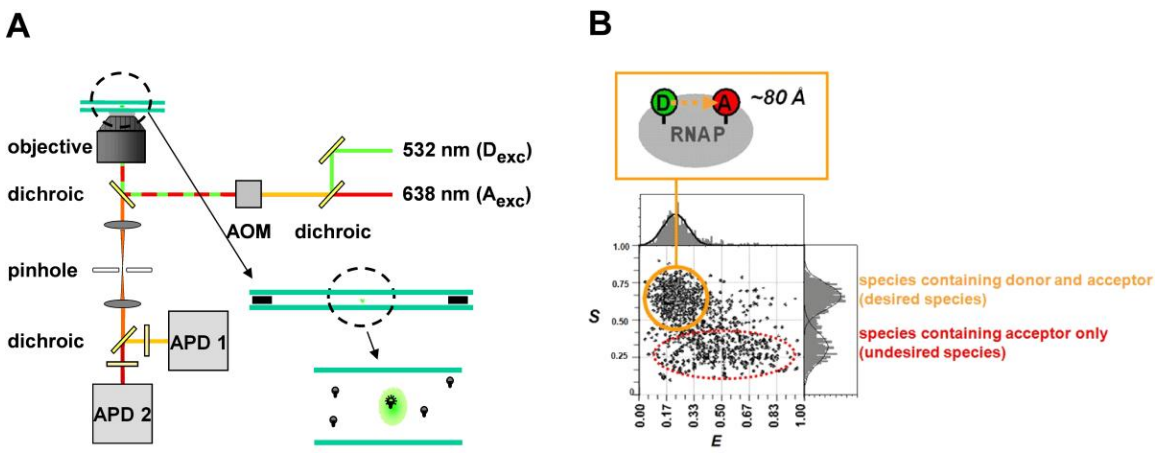

Figure 5. Measurement of smFRET: two-color ALEX confocal microscopy.

(A) Experimental set-up. The system contains two lasers, one providing excitation at the donor excitation wavelength (532 nm) and the other providing excitation at the acceptor excitation wavelength (638 nm), which are alternated using an acousto-optical modulator (AOM) at a microsecond time-scale. Excitation laser beams are coupled through optical fiber, directed onto an objective by means of a dichroic mirror, and tightly focused to a femtoliter-scale observation volume in a sample chamber (inset at right center). Freely diffusing fluorescently-labeled single molecules are excited by the laser beams (inset at lower right), and the generated fluorescence emission travels down though the objective. Unfocused light is filtered out by the pinhole and the fluorescence emission signals from donor and acceptor are directed into respective donor and acceptor detection channels, counted by avalanche photodiodes (APD1 and APD2)

(B) Molecular sorting using ALEX confocal microscopy. Each single molecule transiting the femtoliter-scale observation volume yields two parameters: a donor-acceptor stoichiometry parameter (*S*) and a donor-acceptor smFRET efficiency (*E*), resulting in in a two-dimensional *E*/*S* plot. The distribution of observations on the *S*-axis (histograms at right) enables distinction between species containing both the donor and the acceptor (desired species) and species containing only the donor or only the acceptor (undesired species, comprising incompletely labeled complexes, incompletely assembled complexes, or complexes in photophysical dark states). Upon selection of doubly-labeled species, the distribution of *E* values on the *E*-axis (histogram at top) defines mean *E* and enabling calculation of mean donor-acceptor distance, *R* (image at top). The number of peaks in the distribution determines the number of distinguishable subpopulations and ,for each distinguishable subpopulation, mean *E* and mean *R* can be defined.

2. Experimental Strategy

The goal of this work is to define the location of scrunched nontemplate single-stranded DNA during initial transcription using single molecule fluorescence resonance energy transfer and distance-restrained rigid body docking. Accomplishment of this goal will be done through three separate aims:

(1) Formation of fluorescently-labeled RNAP-DNA pairs in static scrunched initial transcribing complexes (ITCs) of lengths ranging from 0 nucleotides scrunched (RP_o) to 8 nucleotides scrunched.

(2) Determination of the path of scrunched nontemplate single-stranded DNA in fluorescently-labeled static ITCs using smFRET derived distance restrained docking.

(3) Comparison of smFRET derived distance restrained docking results for scrunched complexes with either consensus or anti-consensus discriminator elements.

2.1. Formation of fluorescently-labeled static scrunched complexes

In order to analyze the location of scrunched nontemplate (NT) single-stranded DNA (ssDNA) during initial transcription using smFRET, ITCs of varying lengths were formed, each containing a fluorescent donor probe in RNAP and a fluorescent acceptor probe in the scrunched NT ssDNA. These complexes needed to be static, that is, they must remain at the desired scrunched length throughout the smFRET analysis. This was achieved by forming five ITCs with scrunched NT ssDNA de novo, including 0 nucleotides (nt) scrunched (RP_o), 2-nt scrunched (S2), 4-nt scrunched (S4), 6-nt

scrunched (S6), and 8-nt scrunched (S8). To facilitate the formation of the scrunched complexes, synthetic nucleic acid scaffolds corresponding to the transcription bubble and downstream double-stranded DNA (dsDNA) were used. The NT strand, in the case of RP_o, consisted of a 14-nt single-stranded DNA tail located from -12 to +2 in relation to the +1 transcription start site, with a consensus promoter -10 element (TATAAT) and consensus discriminator element (GGG), followed by an 18-nt duplex-forming segment (+3 to +20). The template (T) strand consisted of a noncomplementary 8-nt ssDNA tail, followed by an 18-nt duplex-forming segment. A 7-nt RNA oligonucleotide complementary to the T strand (-6 to +1) made up the final component of the nucleic acid scaffold. Previous studies have shown that these scaffolds retain all functional properties of the transcription bubble and downstream dsDNA in RP_o (Mekler et al., 2011; Zhang et al., 2012). For scrunched complexes, additional nucleotides were incorporated into the NT ssDNA between positions 4-nt and 5-nt downstream of the downstream edge of the -10 element (Figure 6). The number of nucleotides added was equivalent to the magnitude of scrunching, i.e., two nucleotides were incorporated for the 2-nt-scrunched complex. The structure and function of these probe-labeled scrunched ITC's were verified using radiochemical and transcription assays.

Figure 6. Nucleic acid scaffolds with target probe sites in nontemplate strand DNA Nucleic acid scaffolds, representing the labeling sites for nontemplate strand DNA, are shown with the ssDNA transcription bubble and downstream dsDNA duplex illustrated. Each scaffold contains nontemplate strand DNA (top), template strand DNA (middle) and RNA (bottom, red). Acceptor probe (A647N) sites are indicated by green boxes. The -10 element and discriminator element of the NT strand are shown in blue and cyan, respectively. The single-stranded portion of the template strand and the RNA strand are shown in lowercase as they complementary to each other, but not complementary to the nontemplate strand.

2.1.1. Generation of fluorescently-labeled RNAP derivatives

For this work, I incorporated the reference probe Dylight 550 (D550) at each of four sites on RNAP core: residues 106, 222, and 937 of β and residue 284 of β' (Figure 7). Each of the probe sites has been previously studied using smFRET: β106 (Chakraborty et al., 2012) (Xu, 2013), β222 and β937 (Xu, 2013), and β'284 (Chakraborty et al., 2012). The four RNAP probe sites were well separated and radially distributed around the active center cleft on the periphery of RNAP core to provide accurate nucleotide distances using FRET (~20-100 Å). These four sites were at non-conserved residues and have been shown to not affect the structural integrity or functionality of RNAP (Chakraborty et al., 2012; Xu and Rutgers University. Graduate School--New Brunswick., 2013).

Each D550 RNAP derivative was prepared using unnatural amino acid mutagenesis (Chin et al., 2002) (Wang, 2008), Staudinger ligation (Chakraborty et al., 2010; Saxon and Bertozzi, 2000), and RNAP reconstitution methods (Tang et al., 1995). The procedure (Figure 8) involved (i) incorporation of nonsense amber (TAG) in the genes expressing β and β′ subunits at the residues of interest; (ii) formation of β and β′ subunits containing 4-azidophenylalanine at the sites of interest using cells with an engineered suppressor-transfer RNA (tRNA)/aminoacyltRNA-synthase pair in media supplemented with 4-azidophenylalanine; (iii) and incorporation of the fluorescent probe D550 into β and β′ subunits at the proposed sites via azide-specific chemical modification, achieved by Staudinger ligation using D550-phosphine derivatives.

RNAP holoenzyme derivatives were reconstituted from each of the D550 derivatives, forming four distinct RNAP holoenzyme derivatives. Transcriptional activities of RNAP derivatives were comparable to wild-type RNAP, thus suitable for smFRET analysis.

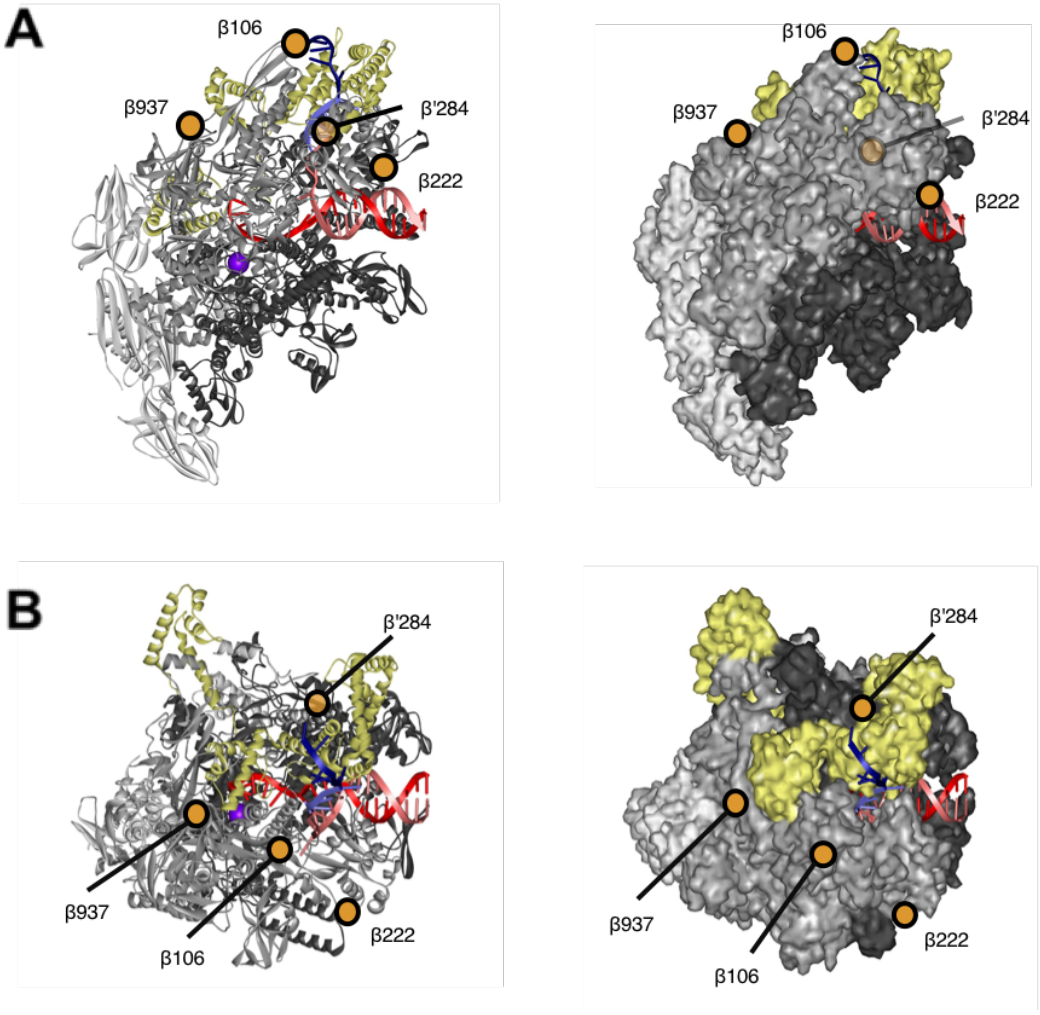

Figure 7. Labeling sites on RNAP Holoenzyme in context of RP_o.
Crystal structure of RP_o (two orthogonal views in ribbon representations at left and surface representation at right, showing reference probe sites in RNAP as orange circles. The DNA template strand and DNA nontemplate strand are in red and pink, respectively, with the discriminator portion of the nontemplate strand in light blue and the -10 element of the nontemplate strand in blue. RNAP β′ and β, α, and ω are in black, dark gray, light gray, and light gray, respectively. σ^{70} is in yellow. The active-center Mg^{2+} is shown as a violet sphere. The RNAP β′ nonconserved domain has been omitted for clarity.
(A) Upstream face
(B) Top face (view into active-center cleft; -90°rotation about x-axis relative to A)

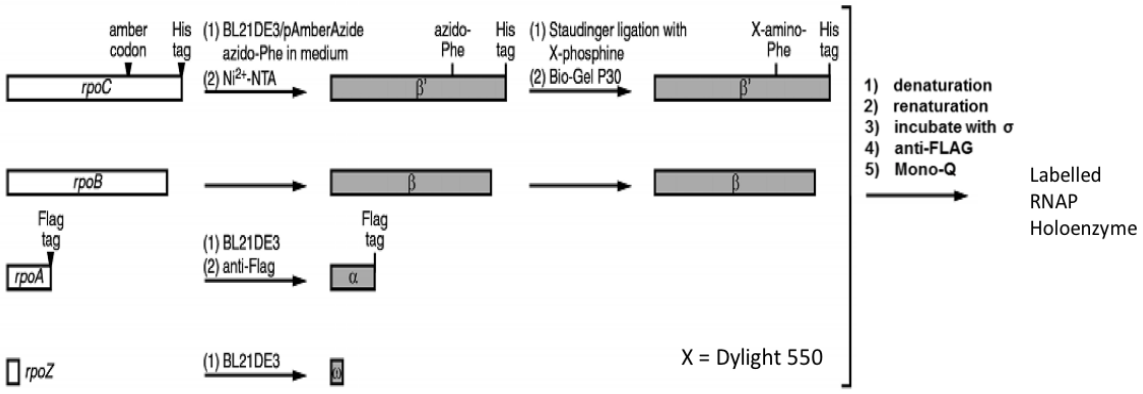

Figure 8. Preparation of RNAP holoenzyme derivative with a donor fluorophore (D550) on β subunit
Conjugation of fluorescent probes at β by unnatural amino acid mutagenesis to incorporate 4-azidophenylalanine at sites of interest in β, followed by Staudinger ligation to incorporate fluorescent probes at 4-azidophenylalanine in β, followed by in vitro reconstitution of RNAP from D550-labeled β (or β′) and unlabeled β′ (or β), α* (covalently linked α-N-terminal-domain dimer) and ω (see Materials and Methods). RNAP holoenzyme derivative was prepared by incubating RNAP core derivative with labeled σ, purified by affinity chromatography and ion-exchange chromatography. Plasmids, genes, and proteins are shown as ovals, open bars, and closed bars, respectively.

2.1.2. Generation of fluorescently-labeled DNA

A fluorescent acceptor probe was incorporated at each of twenty positions on the nontemplate strand. For RP_o, the probe was incorporated at each position 3 nt (termed +3 for the study) or 5 nt (+5) downstream of the downstream edge of the -10 element in the nontemplate strand. For each scrunched complex, the probe was placed at the same positions as RP_o, in addition to each of every other position within the added NT ssDNA segment, as seen in Figure 6. That is, S2 would be labeled at +3, +5, and +7; S4 at +3, +5, +7, and +9; S6 at +3, +5, +7, +9, and +11; and S8 at +3, +5, +7, +9, +11, and +13.

These positions were chosen because they allow us to examine the large majority of the NT ssDNA without perturbing the necessary interactions of the -10 element and the downstream dsDNA duplex in maintaining the transcription bubble.

The acceptor probe used for this study was Atto647N (A647N). Its excitation spectrum suitably overlaps with the emission spectrum of D550. Atto647N was covalently attached to the targeted nucleotide base of the NT through a six-carbon flexible linker strand DNA, allowing for flexibility and room for free orientation of the probe. Labeling specificity of ~100% and labeling efficiency of ~100% were ensured by use of a single reactive modified base incorporated in the synthetic oligos and by purification of the labeled oligos with HPLC equipped with a fluorescence detector.

2.2. Determination of scrunched nontemplate strand DNA path

Functional initial transcribing complexes labeled with a donor probe on RNAP and an acceptor probe on nontemplate single-stranded DNA were formed and FRET measurements were obtained and analyzed. The calculated FRET distances were then used to map the positions of the scrunched NT ssDNA during each examined state of scrunching using FRET-derived distance-restrained docking.

2.2.1. smFRET measurements

In-solution smFRET data acquisition was performed using ALEX confocal microscopy (Figure 5A) for each of the five ITC's studied (RP_o, S2, S4, S6, S8).

Each smFRET measurement produced a 2-dimensional *E*/*S* plot, from which the doubly labeled species were isolated and their *E* distributions were plotted. The equilibrium distribution of *E* values was fitted with Gaussians, allowing distinction of subpopulations. For those complexes that reported more than one population, each population was further examined for functionality and the mean *E* values for populations containing a transcriptionally-competent ITC were defined.

For each RNAP-DNA probe pair, the Förster parameter R_o was determined in each scrunched state. This allowed the mean probe-probe distances for each RNAP derivatives to be derived from Equation (1) using values of *E* and R_o. In total, 60 probe-probe distances were employed in the subsequent computational modeling, including 8 for RP_o, 12 for S2, 16 for S4, 20 for S6, and 24 for S8 complexes.

2.2.2. Rigid body docking

Positions of NT ssDNA in each of the scrunched complexes were mapped by use of FRET-derived distance-restrained docking. A comprehensive toolkit for FRET-restrained structural modeling was recently developed by Seidel et al., termed FRET-restrained positioning and screening (FPS), which is able to carry out FRET-derived rigid body docking (Kalinin et al., 2012). The high accuracy of their method was demonstrated by

docking a DNA primer template to HIV-1 reverse transcriptase and comparing the derived model with the known crystal structure to great effect. FPS was applied to map the positions of NT ssDNA in each of scrunched state using the experimentally-determined systematic smFRET measurements. In FPS, spatial distributions of the dyes were modeled by a geometric accessible volume (AV) algorithm (Cai et al., 2007; Muschielok et al., 2008) based on the dye dimensions, the linker length and width, and the local structure of the biomolecules where the dyes were attached. The dye molecules are assumed to be able to freely occupy every position within its AV and spend an equal amount of time in each position, therefore its position can be represented by a mean position. The mean positions were used to calculate the modeled donor-acceptor distances. To find the position of labeled NT ssDNA with respect to each scrunched state, rigid body docking through FPS was accomplished by minimizing the weighted data-model deviation (χ_E^2):

$$\chi_{\mathrm{E}}^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{\left(\mathrm{R_{DA(i)}} - \mathrm{R_{model(i)}}\right)^2}{(\Delta \mathrm{R_{DA}})^2} \tag{3}$$

where R_{DA} is the experimentally measured donor-acceptor distance, ΔR_{DA} is the uncertainty of R_{DA}, R_{model} is the modeled donor-acceptor distance, and n is the number of restraints. The subscript i refers to the i^{th} pair of donor-acceptor.

To further refine the method, steric clashes between docking bodies were prohibited by introducing strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii, as defined by:

$$\chi_{\mathrm{clash}}^2 = \sum_{i,j} \begin{cases} 0 & , r_{ij} \geq r_{wi} + r_{wj} \\ \frac{\left(r_{wi} + r_{wj} - r_{ij}\right)^2}{r_{\mathrm{ctol}}^2} & , r_{ij} < r_{wi} + r_{wj} \end{cases} \tag{4}$$

where r_{ij} is the distance between atoms i and j which belong to different subunits, r_{wi} and r_{wj} are their van der Waals radii, and r_{ctol} is the pre-defined clash tolerance. Therefore, positioning of NT ssDNA in relation to the scrunched ITC was guided by minimizing an overall quality parameter, reduced χ_r^2, that takes into account both violation of FRET restraints and of van der Waals radii, as defined by:

$$\chi_r^2 = \frac{\chi_E^2 + \chi_{clash}^2}{n-p} \rightarrow \min \quad (5)$$

where n is the number of distance restraints and p is the number of degrees of freedom, which is equal to 6 × (number of bodies – 1).

2.3. Analysis of discriminator element affects in scrunching

To assess the effects of the discriminator element in the nontemplate strand on the location of scrunched NT ssDNA in the ITC's studied in this work, a the consensus discriminator element (GGG), immediately downstream of the -10 element, was substituted with weak discriminator element (CCC) in the DNA scaffolds. For future referencing, antidiscriminator complexes will be labeled aDSC and consensus discriminator complexes will be labeled DSC. Subsequent labeling of DNA with probes, formation of static ITC's with donor/acceptor probe pairs, validation of ITC structure and function, smFRET data acquisition and analysis, and FRET-derived rigid body docking were all done as before. The resulting aDSC structures were compared with the DSC structures to analyze differences.

3. Materials and Methods

3.1. Nucleic acid fragments

Unlabeled DNA oligos used for the template strand and nontemplate strand were purchased from Integrated DNA Technologies (IDT). DNA with six-carbon linkers incorporated at the labeling sites in the nontemplate strand were purchased from Biolink Technologies Inc (Figure 9). RNA oligos were purchased from Biolink Technologies Inc.

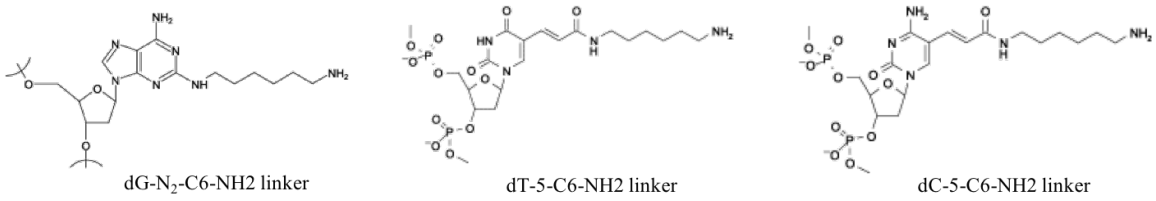

Figure 9. DNA nucleotide derivatives used for fluorescent probe attachment
Nucleotides used for attachment of Atto647N NHS ester to the amine group at the end of a six carbon linker. The linker was incorporated at positions within the nitrogenous base of nucleotide (C2 nitrogen in guanosine, C5 in thymidine and cytidine).

3.2. Plasmids

Plasmids used in this work are summarized in Table 1. pET28a-NF-α^{I}NTD-α^{II}NTD encodes the N-terminally Flag-tagged *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{I}), followed by the linker GlySerGlyGlySerGly, followed by a second *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{II}) (Wang, 2008). Plasmids pMKSe2, pT7β′, pT7ω, and pGEMD encode *E.coli* RNAP β, β′, ω, and σ^{70} subunits, respectively (Igarashi and Ishihama, 1991; Naryshkin et al., 2001; Severinov et al., 1993).

Plasmid pET21d rpoB-CH6 encodes *E.coli* β subunit with a C-terminal-tagged hexahistidine. Plasmids pET21d-rpoBxxxTAG-CH6 (xxx = 106, 222, 937) were constructed by replacing the corresponding rpoB codon in template pET21d-rpoB-CH6 with an amber codon (TAG) using site-directed mutagenesis (QuikChange II; Agilent Technologies, Inc.) (Wang, 2008; Xu, 2013). Plasmid pET21a rpoC-CH6 encodes *E.coli* β′ subunit with a C-terminal-tagged hexahistidine. Plasmid pET21a-rpoC284TAG-CH6 was constructed from plasmid pET21a-rpoC-CH6 using the same method described above for the rpoB amber codon substitutions (Wang, 2008).

pEVOL-pAzF encodes a mutant *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tyrosyl amber suppressor *Mj*$tRNA^{Tyr}_{CUA}$ pair (Chin et al., 2002; Young et al., 2010), which specifically charges for the unnatural amino acid 4-azidophenylalanine. The plasmid was kindly provided by Peter Schultz, The Scripps Research Institute, La Jolla CA.

Table 1. Plasmids

Plasmids	Characteristics	Source
pET28a NF α^{I}NTD-α^{II}NTD	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoA(1-235)-rpoA(1-235)(NFLAG)	Chakraborty et al., 2012
pMKSe2	Ap^R; ori-pBR322; P_{lacUV5}-rpoB	Severinov et al., 1993
pT7β'	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoC	Naryshkin et al., 2001
pT7ω	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoZ	Naryshkin et al., 2001
pGEMD	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD	Igarashi and Ishihama, 1991
pET21d rpoB-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB(CH6)	Chakraborty et al., 2012
pET21d rpoB106TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB106amber(CH6)	Chakraborty et al., 2012
pET21d rpoB222TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB222amber(CH6)	Wang, 2008
pET21d rpoB937TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB937amber(CH6)	Xu, 2013
pET21a rpoC-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoC(CH6)	Chakraborty et al., 2012
pET21a rpoC284TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoC284amber(CH6)	Chakraborty et al., 2012
pEVOL-*p*AzF	Cm^R; ori-p15A; P_{araBAD}-aaRS/$tRNA^{Tyr}_{CUA}$	Young et al., 2010

3.3. Fluorescent probes

Fluorescent donor probe Dylight 550-phosphine was purchased from ThermoFisher (catalog # 88910). Fluorescent acceptor probe Atto647N-NHS Ester was purchased from Atto-Tec (catalog # AD647N-31).

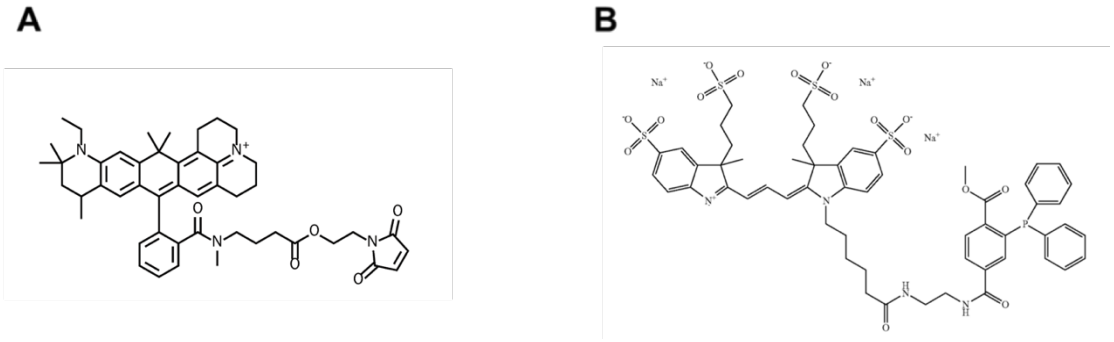

Figure 10. Fluorescent donor and acceptor probes used in this study

(A) Atto647N NHS-ester; excitation maximum: 644 nm, emission maximum: 679 nm
(B) Dylight 550 phosphine; excitation maximum: 526 nm, emission maximum: 576 nm

3.4. RNAP holoenzyme formation

3.4.1. Preparation of wild-type β, β′, and ω

Inclusion bodies containing unlabeled E.coli RNAP β subunits, inclusion bodies containing unlabeled E.coli RNAP β′ subunits, and inclusion bodies containing E.coli RNAP ω subunits were prepared as in Naryshkin et al., 2001.

3.4.2. Preparation of Flag-αNTD^{I}-GSGGSG-αNTD^{II}

Preparation of the N-terminally Flag-tagged α fusion protein (Flag-αNTD^{I}-GSGGSG-αNTD^{II}) containing the first *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{I} and second *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{II}) with a) connected by a linker of GlySerGlyGlySerGly went accordingly: *E. coli* strain BL21(DE3) (Invitrogen, Inc.) was transformed with plasmid pET28a-NF-αNTD^{I}-αNTD^{II}. Inoculated a single transformed colony in 50 ml LB medium containing 40 μg/ml kanamycin and incubated for 16 h at 37 ºC while shaking. 1 L LB medium containing 40 μg/ml kanamycin was inoculated with 10 ml of the culture, incubated at 37 ºC while shaking until OD_{600} = 0.6. Protein production was induced by addition of 1 mM IPTG and the culture were incubated for an additional 3 h at 37ºC. Cells were collected by centrifugation (4,500xg; 20 min at 4 ºC) and then resuspended in 25 ml of lysis buffer [20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10 mM EDTA, and one protease inhibitor cocktail tablet (Roche, Inc)]. The resuspended cells were lysed using an Avestin

EmulsiFlex-C5 cell disrupter (Avestin, Inc.). Lysates were centrifuged (20,000xg; 20 min at 4 ºC), supernatants was collected, and ammonium sulfate was added to precipate the protein (35 g per 100 ml supernatant) The pellets were centrifuged (20,000xg; 10 min at 4 ºC), dissolved in 10 ml TBS (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 5% glycerol) and loaded onto two 5 ml columns packed with anti-FLAG M2 affinity gel (Sigma Aldrich, #A2220) pre-equilibrated in TBS. 50 ml of TBS was added to wash the columns and then 25 ml TBS containing 0.1 mg/ml FLAG peptide (Sigma Aldrich, #F3290) was added to elute 25- 1 ml fractions. Fractions containg αNTD^{I}-αNTD^{II} were combined, and precipitated using ammonium sulfate and the pellets were stored at -80 ºC. Typical yields of Flag-αNTDI-GSGGSG-αNTDII were 50 mg/L.

3.4.3. Preparation β′ and β derivatives

4-azidophenylalanine-labeled β and β′ derivatives were prepared using the following procedure (Chin et al., 2002): *E. coli* competent cell strain BL21(DE3) (Stratagene, Inc) was transformed with plasmid pEVOL-*p*AzF, which encodes the aaRS/tRNA pair that incorporates 4-azidophenylalanine, and was then transformed with either plasmid pET21d-rpoBxxxTAG-CH6 (xxx = 106, 222, 937) or plasmid pET21a-rpoC284TAG-CH6, depending on which derivative was being formed. Individual colonies from the transformants were inoculated into 50 ml of LB medium containing 100 μg/ml ampicillin and 35 μg/ml chloramphenicol, and cultures were incubated in the dark for 16 h at 37°C while shaking. The cells were harvested by centrifugation at 4000xg for 10min at 4°C, resuspended in 10 ml of M9+ medium (prepared from M9 minimal salts (Sigma Aldrich), supplemented with 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 mM $FeSO_4$, 3 nM

$(NH_4)_6Mo_7O_{24}$, 400 nM H_3BO_3, 30 nM $CoCl_2$, 10 nM $CuSO_4$, 80 nM $MnCl_2$, 20nM $ZnSO_4$, 0.4 μg/ml choline chloride, 0.5 μg/ml folic acid, 0.5 μg/ml nicotinamide, 1 μg/ml myo-inositol, 1 μg/ml pyridoxal HCl, 2 μg/ml thiamine HCl, 0.05 μg/ml riboflavin, 1 μg/ml biotin), and added to 1 L M9+ medium containing 100 μg/ml ampicillin, 35 μg/ml chloramphenicol, and 1 mM 4-azido-L-phenylalanine (Chem-Impex International, Inc., Catalog #06162), and the cells were grown in the dark at 37°C while shaking. When the OD_{600} = 0.4, arabinose was added to the culture at a final concentration of 0.02%; when the OD_{600} = 0.6, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and post-induction incubation continued for an additional 3 h in the dark at 37°C while shaking. Cells were harvested by centrifugation at 4000xg for 30 min at 4℃, were lysed, and inclusion bodies containing 4-azidophenylalanine-labeled β and β′ derivatives were prepared and stored as in Naryshkin et al. (2001), with the following modifications: cell pellet from 1 L of culture was suspended in 25 ml of lysis buffer (40 mM Tris-HCl, pH 7.9, 300 mM KCl, 10 mM EDTA, 0.2% (w/v) sodium deoxycholate (Sigma Aldrich)]. Cells were lysed using an Avestin EmulsiFlex-C5 (Avestin Inc.)). Cell lysis was centrifuged at 24,000xg for 30 min and the pellet containing the inclusion bodies was collected. The pellet was suspended in 20 ml of lysis buffer with 0.2% Octyl-β-D- glucopyranoside (Sigma Aldrich) and 0.02% lysozyme by sonication for 2 minutes at 40% maximum sonication output, followed by centrifugation for 20 min at 24,000xg at 4°C. The washing procedure was repeated with 0.5% Triton X-100 (Sigma Aldrich) added to the lysis buffer instead during this round. The pellet was then resuspended in 8 ml of storage buffer (40mM Tris-HCl, pH 7.9, 300mM KCl, 10mM

EDTA, 10% glycerol), divided into 1-ml aliquots, and stored at -80°C. Typical inclusion body yields were ~100 mg proteins for 1 L culture.

β and β′ derivatives containing 4-azidophenylalanine were labeled with Dylight 550-phosphine using a Staudinger-Bertozzi ligation (Chakraborty et al., 2010). 20 μM 4-azidophenylalanine-labeled β or β′ derivative and 200 μM Dylight 550 phosphine were mixed in 3 ml of denaturation buffer [50 mM Tris-HCl, pH 7.9, 6 M guanidine-HCl, and 5% glycerol], and incubated in the dark for 15 h at 37℃. The mixture was then added to 2- 10 ml Bio-Gel P30 columns (BioRad, Inc.) pre-equilibrated in denaturation buffer. The columns were washed with denaturation buffer, and then the Dylight 550-labeled β and β′ derivatives were eluted with the same buffer. Labeling efficiencies and specificities were determined as per Chakraborty et al. (2010). Typical labeling efficiencies and labeling specificities were >90%.

3.4.4. Preparation of σ^{70}

E.Coli strain BL21(DE3) cells were transformed with plasmid pGEMD, plated on LB plates containing 100 μg/ml ampicillin, and grown overnight at 37°C. A single colony was inoculated into 5 ml of LB containing 100 μg/ml ampicillin and grown overnight at 37°C. The overnight culture was inoculated into 1L of LB media containing 100 μg/ml ampicillin and grown until OD_{600} = 0.6. Protein expression was induced with 1mM IPTG and cells were harvested after three hours by centrifugation at 4000xg for 30 min at 4℃. The cell pellet was suspended in 150ml of lysis buffer and the cells were lysed using

an Avestin EmulsiFlex-C5. Sample was centrifuged at 4000xg for 30 min and the pellet containing the inclusion bodies of σ^{70} was collected. The pellet was washed twice with the lysis buffer plus 0.5% of Triton X-100 (first wash) and the lysis buffer plus 0.5% Triton X-100 plus 10mM DTT (second wash). The pellet was dissolved in 50 ml of denaturation buffer (50mM Tris HCl pH 7.9, 6M Guanidine HCl, 10mM $MgCl_2$, 10uM $ZnCl_2$, 1mM EDTA, 10mM DTT, 10% glycerol) and dialyzed against 2 L of the TGEB buffer (20mM Tris HCl pH7.9, 0.1mM EDTA, 10mM β -mercaptoethanol, 5% glycerol) plus 0.2 M NaCl for 36 hours with three changes of the buffer. Dialyzed sample was centrifuged at 15,000xg for 30min and the supernatant was loaded onto MonoQ HR 10/10 (GE Healthcare) anion exchange column equilibrated in TGEB buffer plus 0.2 M NaCl. Protein was eluted with a linear gradient of NaCl (from 0.2M to 0.5M). Fraction containing σ^{70} were combined, concentrated, and stored in storage buffer (25mM Tris HCl pH 7.9, 50% glycerol, 0.1M NaCl, 0.1mM EDTA, 1mM β-mercaptoethanol) at -80°C. Typical yields were ~40 mg.

3.4.5. RNAP holoenzyme derivative reconstitution

Reconstitution of RNAP derivatives was done as per Naryshkin et al. (2001) under denaturing conditions using D550-labeled β or β′ derivatives, wild-type β or β′ inclusion bodies , ω inclusion bodies, and purified flag-αNTD^{I}-GSGGSG-αNTD^{II}. Reconstitutions were done with following amounts: 4.3 mg (80 nmol) Flag-αNTD^{I}-GSGGSG-αNTD^{II}, 3.0 mg (20 nmol) Dylight 550-labeled or unlabeled β derivative, 7.8 mg (50 nmol) Dylight 550-lebelled or unlabeled β′ derivative, and 2.0 mg (200 nmol) ω in 60 ml

denaturation buffer containing 6 M guanidine-HCl, 50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM EDTA, 10 mM DTT, and 10% glycerol. Denatured reconstitution mixtures were transferred into a dialysis membrane (Spectra/Por Dialysis Membrane, MWCO 3500 Da, #132724) and dialyzed against 3000 L reconstitution buffer containing 50 mM Tris–HCl, pH 7.9, 200 mM KCl, 10 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 20% glycerol and 5 mM β-mercaptoethanol for 36 hours at 4 ºC with two buffer changes (final buffer changed did not contain β-mercaptoethanol in the buffer). The dialyzed sample was centrifuged at 20,000xg for 30mins at 4 ºC to separate out particulates, and the remaining supernatant containing RNAP core derivatives was incubated with 2.8 mg (40 nmol) of σ^{70} for 45 mins at 30°C to form RNAP holoenzyme derivatives.

Initial purification was done by anti-flag affinity chromatography, loading the protein sample onto 3-ml columns packed with anti-FLAG M2 affinity gel pre-equilibrated with buffer A, containing 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1 mM EDTA, 5% glycerol. The column was washed with 50 ml of buffer A and the bound protein was subsequently eluted with 15 ml buffer A containing 100 μg/ml FLAG peptide. The eluted sample was further purified using anion-exchange chromatography, using a Mono-Q HR 10/10 column (GE healthcare life sciences) pre-equilibrated in TGED buffer (20 mM Tris–HCl, pH 7.9, 1 mM EDTA, 5% glycerol, 1 mM DTT) following procedures from Minakhin et al. (2001). Purified RNAP holoenzyme derivatives were aliquotted and stored in 25 mM Tris-HCl, pH 7.9, 0.1 M NaCl, 0.1 mM EDTA, 1 mM β-mercaptoethanol, and 50% glycerol at -80°C.

3.5. Transcription Assay

Transcriptional activities of wild type RNAP holoenzyme and RNAP holoenzyme derivatives were measured using Ribogreen transcription assays. RNAP holoenzyme and RNAP holoenzyme derivatives were incubated with DNA fragment N25 (Revyakin et al., 2006) for 20 min at 37°C to form RP_o. Each reaction contained 75 nM RNAP holoenzyme, 20 nM DNA in transcription buffer [50 mM Tris-HCl, pH8.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 100 μg/ml bovine serum albumin, or BSA]. The negative control omitted RNAP in the reaction. NTPs were added to the mixtures at a final concentration of 100 μM each and transcription reactions were allowed to proceed for 1 h at 37°C. Transcription was quenched by addition of 1 μl of 5 mM $CaCl_2$ and 2 U of DNase I (Thermo Fisher Scientific Inc.) in each reaction, which were incubated for 90 min at 37°C. Reaction mixtures were supplemented with 1/500 diluted Ribogreen RNA reagent (Invitrogen; #R11491) in 100 μl TE (10 mM Tris-Cl, pH8.0, 1 mM EDTA) and incubated for 10 min at room temperature. Product RNA was transferred to a Corning 3686 assay plate and quantified by measuring fluorescence intensity in a Tecan GENios Pro plate reader (λex = 485 nm; λem = 535 nm). The ribogreen assay readout is in terms of fluorescence intensity (FI). Wild-type RNAP transcription activity is quantified by the ratio of FI for the sample to FI for the negative control (without RNAP). RNAP transcription activity of each labeled RNAP derivative was quantified by percentage of FI for the RNAP derivative over FI for the wild-type RNAP.

3.6. Formation of static scrunched complexes

The static scrunched ITCs for this study were formed in a multi-step process, involving (i) annealing of the NT and T strands, (ii) incubation with D550-labeled RNAP holoenzyme derivative, and (iii) subsequent incubation of a 7 mer RNA. A647N-labeled nontemplate strand DNA and unlabeled template strand DNA were annealed by mixing equimolar amounts in a buffer containing, 20 mM Tris, pH 7.9, 100 mM NaCl, and 0.1 mM EDTA heating for 2 min at 95°C, and slowly cooling down to 25°C. The DNA duplex was then incubated with an equimolar amount of D550-labeled RNAP holoenzyme derivative in transcription buffer for 15 min at room temperature. Doubly-labeled RNAP/DNA complexes were then incubated with a three-fold higher concentration of 7-mer RNA and incubated for 10 min at room temperature. The higher concentration of RNA is needed to "push" its inclusion into the static scrunched complex.

3.7. Radiochemical assays to assess components of fluorescently-labeled static scrunched complexes

Radiochemical assays were performed to verify the presence of all components of the static scrunched complexes. For each probe-labeled DNA/RNAP combination, four separate reactions were performed. The first two reactions used a 5'-[^{32}P]-labeled-A647N-labeled NT DNA fragment. The third reaction used a 5'-[^{32}P]-labeled T DNA fragment. The fourth reaction used a 5'-[^{32}P]-labeled RNA 7-mer oligonucleotide. Each 5'-[^{32}P]-labeled-nucleic acid was formed by incubation of 10mM nucleic acid, 50mM [γ-^{32}P]-ATP (>5000 Ci/mmol), and 10U bacteriophage T4 polynucleotide kinase in

bacteriophage T4 polynucleotide kinase (20 μl total reaction volume) (Sigma Aldrich) for 60 min at 37°C, followed by heating at 68°C for 10 min to inactivate the kinase. The static scrunched complexes were formed according to 3.6, with the following changes: 50 nM of each DNA anneal reaction was incubated with 100 nM probe-labeled RNAP (2-fold excess). The first reaction with 5'-[^{32}P]-labeled-A647N-labeled NT strand DNA fragment was incubated with RNAP core derivative, while the second reaction with 5'-[^{32}P]-labeled-A647N-labeled NT strand DNA fragment was incubated with RNAP holo derivative. After incubation of the RNAP/DNA complexes, 150 nM 7-mer RNA oligo was added (3-fold excess). Products were run on a 4-20% TBE polyacrylamide gel for two hours at 4°C and analyzed by storage-phosphor imaging.

3.8. RNA extension transcription assays

Radiochemical transcription assays were performed to verify the formation of functionally competent probe-labeled static scrunched complexes that can extend the 7-mer RNA within the complex. The static scrunched complexes were formed according to 3.6, with the following changes: 50 nM of each anneal reaction was incubated with 100 nM probe-labeled RNAP (2-fold excess) to assure that all of the DNA duplex was part of a scrunched complex. The DNA/RNAP complexes were then incubated with 150 nM of 5'-[^{32}P]-labeled RNA oligonucleotide. The radiolabeled static scrunched complexes (18 μl) were then brought to 37°C and 1 μl of either a CTP + UTP mix (1 mM), complete NTP mix (all 4 NTPs, 1 mM), or transcription buffer (negative control) was added. The reaction was allowed to proceed for 5 min at 37°C, and was stopped by

addition of 10 µl 80% formamide, 10 mM EDTA, 0.04% bromophenol blue, and 0.04% xylene cyanol, followed by incubation for 5 min at 90°C. Products were analyzed by PAGE on 20% TBE-urea sequencing gels followed by storage-phosphor imaging.

3.9. smFRET sample preparation

The static scrunched complexes were formed according to 3.6, with the following changes: the labeled-DNA duplexes and RNAP holo derivatives were incubated in a 1:1 molar ratio (20 nM, 19 µl reaction volume) in transcription buffer (filtered) for 15 min at room temperature. 1 µl of 1.2 µM 7-mer RNA (60 nM final concentration) was added to the mixtures and incubated for 10 min at room temperature. 0.2 µl aliquots from the scrunched complex reaction mixture were transferred to tubes containing 40 µl of KT buffer [40 mM HEPES-NaOH, pH 7.0, 100 mM potassium glutamate, 10 mM $MgCl_2$, 1 mM DTT, 100 µg/ml BSA, 2 mM (±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox; Sigma Aldrich, #238813) , 1 mM Cysteamine (Sigma Aldrich, #30070) and 5% glycerol; filtered].

10 µl of sample in KT placed between was added to a reservoir within a sealable plastic gasket (Sigma Aldrich, #S4435) between two No.1 cover slips (Fisher Scientific, #12-548-B). Cover slips were cleaned prior to usage by washing with acetone (high purity grade), methanol (high purity grade) and distilled water, and dried in air.

3.10. smFRET data acquisition and analysis

Confocal microscopy with alternating laser excitation (ALEX) was used for smFRET data acquisition and analysis, using the setup (Figure 5A) described in (Kapanidis et al., 2005; Kapanidis et al., 2004). A green laser was used for direct excitation of the donor (532 nm; Compass 215M-20; Coherent, Inc.), and a red laser was used for direct excitation of the acceptor (638 nm; Radius 635-25; Coherent, Inc.). Lasers were set at continuous-wave excitation intensities of 100-120 μW at 532 nm and 60-70 μW at 638 nm, aligned using dichroic mirrors, and alternated at 25 μs intervals using an acousto-optical modulator (Neos Technologies, Inc.). The excitation beams were coupled through a single-mode optical fiber, collimated, and directed to an Olympus IX71 inverted microscope (Olympus America, Inc.). A dichroic beam splitter reflected the alternating beams, focusing them into the sample through a 60x oil-immersion objective (20 μm form the bottom coverslip). Fluorescence emission from the sample was collected through the objective, filtered through a 100 μm pinhole, and spectrally split by a dichroic mirror into the donor emission and acceptor emission channels. The photons for each channel were passed through filters (Chroma 585BP70 for D550; Chroma 650LP for A547N) and were focused onto two avalanche photodiode detectors (APD; SPCM-AQR-15; Perkin-Elmer, Inc.), which recorded their arrival times. Data was acquired for each sample during a 30 min time frame.

Photons detected at the donor emission channel (D_{em}) and the acceptor emission channel (A_{em}) were assigned to donor excitation (D_{exc}) or acceptor excitation (A_{exc}) based on photon arrival times, generating fluorescent intensities $F_{D_{exc}}^{A_{em}}$, $F_{D_{exc}}^{D_{em}}$, $F_{A_{exc}}^{A_{em}}$, *and* $F_{A_{exc}}^{D_{em}}$.

The stoichiometry parameter (*S*) was calculated for each above-threshold photon burst using the following equation (Kapanidis et al., 2005; Kapanidis et al., 2004; Lee et al., 2005):

$$S = \left(\gamma F_{D_{exc}}^{D_{em}} + F_{D_{exc}}^{A_{em}}\right) / \left(\gamma F_{D_{exc}}^{D_{em}} + F_{D_{exc}}^{A_{em}} + F_{A_{exc}}^{A_{em}}\right) \quad (6)$$

where $F_{D_{exc}}^{D_{em}}$ is the fluorescence emission intensity for donor-excitation-based donor emission, $F_{D_{exc}}^{A_{em}}$ is the fluorescence emission intensity for donor-excitation-based acceptor emission, $F_{A_{exc}}^{A_{em}}$ is the fluorescence emission intensity for acceptor-excitation-based acceptor emission, $F_{D_{exc}}^{A_{em}}$ is corrected-for donor leakage and acceptor direct excitation, and γ is a detection-correction factor, 1 in this work (Lee et al., 2005).

The donor-acceptor smFRET-efficiency, *E*, was calculated as a ratio of donor-excitation-based acceptor emission intensity to donor-excitation-based donor and acceptor emission intensities:

$$E = \left(F_{D_{exc}}^{A_{em}}\right) / \left(F_{D_{exc}}^{A_{em}} + F_{D_{exc}}^{D_{em}}\right) \quad (7)$$

Diffusing species are identified and sorted using two-dimensional *E*-*S* plots. The parameter *S* distinguishes species containing only a donor probe (D only, *S*~1), containing only an acceptor probe (A only, $S < 0.3$), and containing both donor and acceptor (D-A, $S = 0.3$-0.8). One-dimensional *E* histograms were plotted and fitted with Gaussian curves for those species containing both donor and acceptor probes (D-A) (Figure 5B). The resulting histograms provide population distributions and the mean *E* value for each distinguishable subpopulation.

Mean donor-acceptor distances (R) were calculated from mean E values using the equation (Kapanidis et al., 2005; Kapanidis et al., 2004; Lee et al., 2005):

$$R = R_0[(1/E) - 1]^{1/6} \quad (1)$$

where R_0, the Főrster parameter, is the distance at which 50% of the energy is transferred, as determined by (Clegg, 1992):

$$Ro = 9780(n^{-4}\kappa^2 Q_D J)^{1/6} \text{ Å} \quad (2)$$

where n is the refractive index of the medium (1.4, Clegg, 1992), κ^2 is the orientation factor that depends on the relative orientation of the donor emission and acceptor excitation dipoles [approximated as 2/3 and justified by fluorescence anisotropy measurements demonstrating donor and acceptor probes reorient on the time scale of the donor excited-state life time, and also by the fact that, in some instances, $E < 0.5$ (Clegg, 1992; Wu and Brand, 1992), Q_D is quantum yield of the donor in the absence of the acceptor and J is the spectral overlap integral of the donor emission spectrum and acceptor absorption spectrum, determined using corrected spectra for donor-only and acceptor-only controls (Clegg, 1992).

3.11. Fluorescence quantum yield measurements

The quantum yield of a D550 incorporated into RNAP was determined by comparison to a reference fluorophore with a well-established quantum yield. For D550, Rhodamine 101 (Q = 1.0 in ethanol, (Karstens and Kobs, 1980)) was used as the reference fluorophore. The quantum yield (Q) for each sample was calculated using equation (Lakowicz, 2006):

$$Q = Q_R \frac{I}{I_R} \frac{OD_R}{OD} \frac{n^2}{n_R^2} \quad (8)$$

where I is the integrated fluorescence intensity, n is the refractive index of solvent, and OD is the optical density. The subscript R denotes the reference fluorophore with known quantum yield. Measurements were done in solution using a sub-micro fluorometer cuvette with 10 mm path length. OD was determined at absorption wavelength 500 nm using Lambda 25 UV/VIS spectrometer (PerkinElmer, Inc.). Fluorescence intensity curves were recorded upon excitation at 500 nm with QuantaMaster (Photon Technology International, Inc.).

3.12. Fluorescence anisotropy measurements

Steady-state fluorescence anisotropies were determined in solution containing donor-only or acceptor-only scrunched ITCs, using a T-format QuantaMaster equipped with excitation and emission polarizers (Photon Technology International, Inc). Excitation and emission wavelengths were 540 nm and 575 nm for D550, and 640 nm and 675 nm for A647N. Anisotropy (*r*) was calculated with equation (Chen and Bowman, 1965):

$$r = (\mathrm{I_{VV}} - \mathrm{GI_{VH}})/(\mathrm{I_{VV}} + 2\mathrm{GI_{VH}}) \quad (9)$$

where $\mathrm{I_{VV}}$ and $\mathrm{I_{VH}}$ are fluorescence intensities with the excitation polarizer at the vertical position and the emission polarizer at the vertical position and the horizontal position respectively. G is the grating correction factor, as calculated by equation:

$$\mathrm{G} = \mathrm{I_{HV}}/\mathrm{I_{HH}} \quad (10)$$

where $\mathrm{I_{HV}}$ and $\mathrm{I_{HH}}$ are fluorescence intensities with the excitation polarizer at the horizontal position and the emission polarizer at the vertical position and the horizontal position, respectively.

3.13. Distance-restrained rigid body docking

FRET-derived distance-restrained rigid body docking was performed for each of the static scrunched ITCs (RP_o, S2, S4, S6, and S8) using the FPS program (Kalinin et al., 2012). For each docking analysis performed, three components were required: structures for rigid body docking, labeling position data, and distance restraints.

Two rigid body docking structures were used for each labeled NT strand position in each static scrunched ITC: an ITC with same nucleic acid scaffold used for RP_o in this study, but with the majority of the NT ssDNA removed from the transcription bubble, and a single nucleotide representing the labeled DNA position. As the first docking partner, the ITC consists of RNAP holoenzyme and an RP_o nucleic-acid scaffold with a portion of the NT ssDNA removed to make space for potential positions of the second docking partner. The RNAP ITC was modeled from an *Ec* RNAP ITC, a complete transcription bubble, and a 4nt RNA oligo [PDB accession code 4YLN] (Zuo and Steitz, 2015). The DNA/RNA complex was replaced with the DNA/RNA scaffold for RP_o in this study that was solved in *Tth* RP_o (Zhang, unpublished data), with NT strand -5 to +2 removed. The second docking partner was a single nucleotide specific to the identity of the labeled DNA position; either 2'-deoxyguanosine-5'-monophosphate (dG), thymidine-5'-monophosphate (dT), or 2'-deoxycytidine-5'-monophosphate (dC). Each docking structure was prepared into a .pdb file.

All labeling positions in either RNAP or DNA were specified with the following data for use in the FPS software: labeling position name, molecule name, donor or acceptor, linker length, linker width, AV3, dye radius 1, dye radius 2, dye radius 3, and attachment

point in the corresponding molecule (atom ID, extracted from .pdb file). Dyes were approximated by a sphere with defined radii, which were estimated from the physical dimensions of the dye molecules using Chem3D software. The connecting linker was modeled as a flexible cylinder with length (L_{link}) measured from the attachment point (the C_β of the amino acid for RNAP labeling, and the nitrogenous base of the nucleotide, as in Figure 9, for DNA labeling) to the dye center and width (W_{link}) using typical value of 4.5 Å (Kalinin et al., 2012; Muschielok et al., 2008). For this study, the accessible volume simulation with three dye radii, AV3, was used (Table 2). The systematic FRET distance measurements between each donor-acceptor pair were applied for the distance restraints. The uncertainties of restraints (ΔR_{DA}, from equation 3) were estimated to 15% of the corresponding measured distance based on a benchmark study of FRET-derived distance restraint structural modeling (Knight, et al., 2005).

Model generation was performed in a three-step procedure: search, refinement, and assessment. In the search step, 200 model solutions were generated from random configurations of the docking bodies, excluding those with steric clashes based on a clash tolerance of 2 Å. For the refinement step, the top 20 solutions with the lowest χ_r^2 (from equation 5) were refined with a more stringent steric clash tolerance of 0.5 Å. During this step, AVs were recalculated based on the new environment surrounding the dyes and the attached biomolecule structure and the new mean dye positions were used to optimize the structure. In the assessment step, each of the top 20 refined solutions was analyzed to determine if its position in the structural model was functionally viable, that is, its distance from the known position of NT ssDNA in RP_o is within the range to still maintain a functional transcription initiation complex with intact interactions at the

upstream (-10 element) and downstream elements (dsDNA duplex) of the transcription bubble. The maximum distance that a modeled position could be from the original crystal structure position of the NT ssDNA nucleotide was 7 x N (in Å), where N is the number of nucleotides scrunched in that complex. Seven was chosen because it defines the approximate length of a nucleotide monophosphate. Model positions outside of the assessment parameter were not considered functionally viable, and were grayed out in the final model representations (Figures 13-18).

Structural models were made, producing 20 refined positions for each labeled NT nucleotide, in each scrunched state, in relation to RNAP.

Table 2. Dye and linker parameters for distance-restrained docking

Probe Location	Dye	L_{link} (Å)	W_{link} (Å)	$R_{dye(1)}$ (Å)	$R_{dye(2)}$ (Å)	$R_{dye(3)}$ (Å)
RNAP	D550	12.5	4.5	7.5	5.0	1.5
dG	A647N	19.5	4.5	7.2	4.5	1.5
dT	A647N	19.5	4.5	7.2	4.5	1.5
dC	A647N	19.5	4.5	7.2	4.5	1.5

4. Results

4.1. Site-specific incorporation of fluorescent probes into β and β′ subunits

Site-specific incorporate of probes at selected positions in β′ and β required an approach using unnatural amino acid mutagenesis and Staudinger ligation. The procedure involved (i) site-directed mutagenesis of genes expressing β′ and β subunits to incorporate nonsense amber (TAG) codons at the position of interest; (ii) incorporation of 4-azidophenylalanine into β and β′ subunits at the site of interest, accomplished by growing cells which contained an engineered suppressor-transfer RNA (tRNA)/aminoacyl-tRNA-synthetase pair, in medium supplemented with 4-azidopheylalanine; (iii) incorporation of fluorescent probe D550 into β′ and β subunits through azide-specific chemical modification, completed through Staudinger ligation using phosphine derivatives of the fluorescent probe. The resulting labeled proteins showed high labeling efficiencies (>90%) and high labeling specificities (>90%), mirroring results from work using the same positions and methods (Chakraborty, 2012; Xu, 2013)

4.2. Incorporation of fluorescent probes into functional RNAP

In order to incorporate fluorescent probes into RNAP, *in vitro* reconstitutions of RNAP holoenzyme were performed. Unlabeled and fluorescently labeled subunits of RNAP core ($\beta'/\beta/\alpha^{I}/\alpha^{II}/\omega$) were combined, denatured, renatured, incubated with σ^{70} and purified to prepare intact RNAP holoenzyme derivates labeled at β′or β.

Correct subunit stoichiometries and the presence of fluorescent probes were checked by denaturing polyacrylamide gel electrophoresis (SDS-PAGE) of labeled RNAP derivatives, followed by Coomassie staining and fluorescence scanning. Correct subunit stoichiometries and correct fluorescent emissions were observed in the RNAP holoenzyme derivatives and labeling specificities and efficiencies were typically > 90%.

4.3. Transcription activities of RNAP derivatives

The D550-labeled RNAP holoenzyme derivatives were tested to assure that incorporation of the probe into RNAP did not impede protein function. Ribogreen transcription assays were performed on all RNAP derivatives in comparison to wild-type RNAP (Materials and Methods, 3.5). Their relative transcriptional activities as a percentage to WT RNAP holo are shown in Table 3. All RNAP derivatives showed activities >70% and most were over 90%, indicating that probe incorporation into RNAP did not significantly reduce its function, thereby justifying their use in smFRET experiments.

Table 3. Relative transcriptional activities of RNAP derivatives

RNAP Probe Location	Transcription activity
β106	95%
β222	91%
β937	74%
β′284	90%

4.4. Formation of probe-labeled static scrunched complexes

In order to form static scrunched ITC complexes: (1) NT and T strand fragments must be annealed together, (2) RNAP holoenzyme derivatives must interact with the DNA duplexes to form scrunched ITC complexes, and (3) RNA oligos must anneal with the T strand DNA to further stabilize the scrunched complex. Both during and after its formation, it important that the static scrunched ITCs stably maintain all of their components in order to employ them for smFRET studies. Multiple concurrent radiochemical assays were performed with different 5'-[^{32}P]-labeled-nucleic acid components (T, NT, RNA) to confirm that the complexes contain and retain all of their components, including A647-labeled NT, T, RNA, D550-labeled RNAP core, and σ (see Materials and Methods, 3.7). The reaction mixtures with 5'-[^{32}P]-labeled-nucleic acid were run through gel electrophoresis and examined by phosphor imaging. The results show that the scrunched complexes contain RNAP, σ, NT, T, and RNA in 1:1:1:1:1 stoichiometry (Figure 11). The radiochemical stoichiometry assay was done for each combination of donor-acceptor probe pair in each level of scrunched ITC.

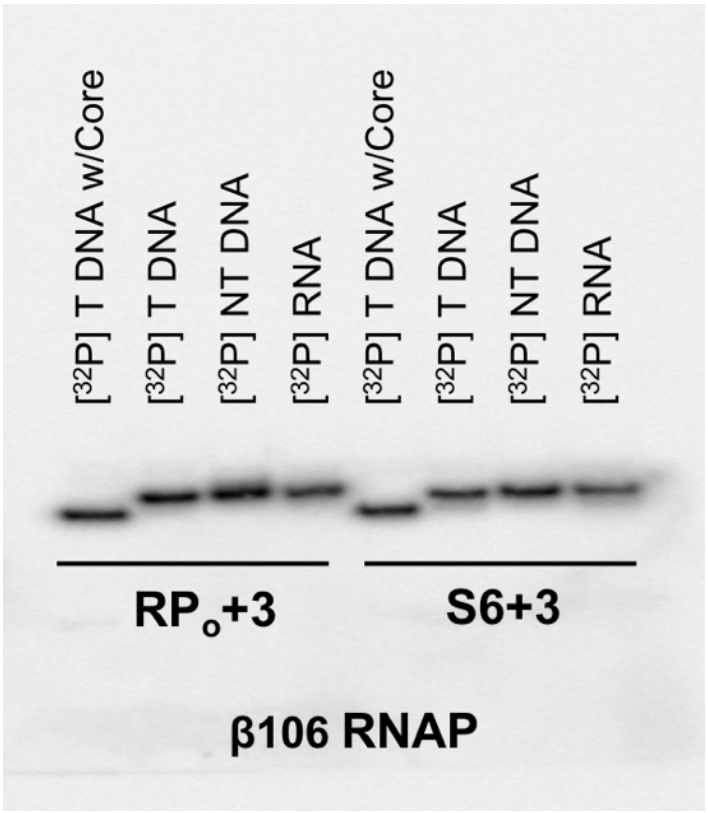

Figure 11. Radiochemical assay to confirm scrunched ITC formation
Complexes were formed with either [^{32}P]-labeled template-strand DNA, [^{32}P]-labeled nontemplate-strand DNA, or [^{32}P]-labeled RNA. Reactions were performed with RNAP holoenzyme derivative except for the left lane in each scrunched complex pair, which was incubated with RNAP Core (lacking σ) to serve as a negative control and size marker. The results show that the complexes contain RNAP, σ, NT, T, and RNA in 1:1:1:1:1 stoichiometry.

4.5. Functionality of fluorescent-labeled static scrunched complexes

Static scrunched complexes were further investigated to confirm that not only were all the necessary components present with the structure, but the that the resulting complexes were transcriptionally active. Radiochemical RNA extension assays were done to confirm that the ITC could competently extend the 7-mer RNA in the presence of NTPs. D550-A647N-labeled scrunched ITCs were formed with 5'-[^{32}P]-labeled RNA and then incubated with a subset of NTPs (no NTP, UTP+CTP, all NTP) to determine if the ITCs could extend the RNA and do so at the right lengths (See Materials and Methods, 3.8) RNA extension assays were done for each donor-acceptor probe pair at each ITC level. Results indicated that each donor-acceptor complex was able to extend the RNA at expected lengths (Figure 12). Since RNA is added in three-fold excess to the RNAP:DNA duplex in order to drive it into the scrunched complexes, there will be a significant band of unextended RNA. Quantification of band intensities for extended RNA as compared to unextended RNA shows that nearly all (>90%) of the complexes are able to competently perform transcription.

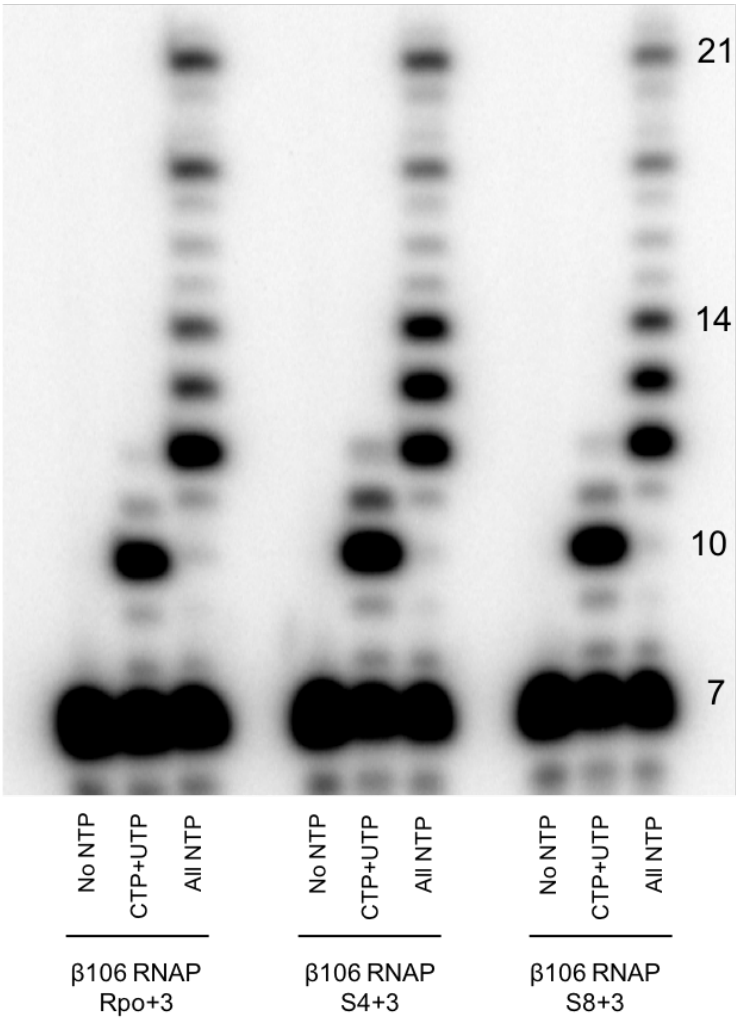

Figure 12. RNA extension assay of static scrunched ITCs
(D550-β106; A647N+3) labeled static ITCs (RP_o, S4, and S8) were formed with 5'-[^{32}P]-RNA (-6/+1) at 3-fold molar excess and were subsequently incubated with NTP subsets (no NTP, UTP+CTP, all NTP). The resulting reaction products were run on a 20% TBE-urea sequencing gel and analyzed using storage phosphor imaging. Results indicate that the ITCs are able to extend the RNA and do so at essentially the correct lengths.

4.6. Spectral properties of D550-A647N pairs in scrunched complexes

Spectral properties of fluorescent probes can be influenced by the environment, primarily due to fluorophore interactions with nearby molecules in its surrounding environment. Accordingly, measurements of donor quantum yield (Q_D), fluorescent anisotropy (*r*), spectral overlap integral (*J*), and Förster parameter R_o needed to be done for each donor-acceptor probe combination for each scrunched state, containing either a consensus (DSC) or anticonsensus (aDSC) discriminator sequence.

Donor quantum yields (Q_D) were measured for fluorescent probes at each of the RNAP labeling sites using donor-only proteins in ITC complexes with unlabeled DNA for each scrunched state and either DSC or aDSC sequence. The Q_D values at different labeling positions varied only slightly, ranging from 0.28-0.34 (Tables 4 & 5). Spectral overlap, *J*, between D550 emission and A647N excitation was determined for each D550-A647N pair in each scrunched state and with either DSC or aDSC sequence, with results ranging between 4.0E-13 to 6.3E-13 $M^{-1}cm^3$ (Tables 4 & 5).

Steady-state anisotropies of scrunched ITCs were measured in samples using donor-only and acceptor-only scrunched complexes. The resulting anisotropy values were between 0.21-0.31 for all complexes (Tables 6 & 7), which was consistent and low compared to calculated anisotropies of probes linked to molecules of ~0.5 MDa and having restricted local motion (Cantor, 1980). Because the probes reorient on the time scale of the probe life times, it substantiates the use of $\kappa^2 = 2/3$ in calculations of R_o. The anisotropy values for each scrunched state were relatively similar and consistent as well, as were the anisotropy values between DSC and aDSC complexes.

Förster parameter R_o was calculated for each donor-acceptor labeled scrunched ITC with either DSc or aDSC sequence from Q_D and donor emission-acceptor excitation spectral overlap integral (J) using equation (2) (Materials and Methods, 3.10) (Tables 4 & 5). The variation in R_o is minimal for each donor-acceptor probe pair in the scrunched ITCs with either DSC or aDSC sequences, suggesting that any changes in smFRET efficiencies were due to the changes in donor-acceptor distance.

Table 4. Spectral properties of fluorescent probes in DSC-scrunched complexes

DNA Probes		β106			β222		
		Q_D	J ($M^{-1}cm^{3)}$	R_o (Å)	Q_D	J ($M^{-1}cm^{3)}$	R_o (Å)
RP_o	**+3**	0.28	4.0E-13	51	0.34	5.2E-13	55
	+5	0.28	4.7E-13	52	0.34	5.4E-13	55
S2	**+3**	0.28	4.0E-13	51	0.30	5.2E-13	54
	+5	0.28	4.7E-13	52	0.30	5.4E-13	54
	+7	0.28	4.4E-13	52	0.30	5.3E-13	54
S4	**+3**	0.32	4.1E-13	52	0.31	5.2E-13	54
	+5	0.32	4.7E-13	53	0.31	5.5E-13	54
	+7	0.32	4.6E-13	53	0.31	5.3E-13	54
	+9	0.32	4.9E-13	54	0.31	5.5E-13	54
S6	**+3**	0.29	4.2E-13	51	0.33	5.3E-13	55
	+5	0.29	4.8E-13	53	0.33	5.5E-13	55
	+7	0.29	4.5E-13	52	0.33	5.3E-13	55
	+9	0.29	5.0E-13	53	0.33	5.7E-13	55
	+11	0.29	4.6E-13	52	0.33	5.3E-13	55
S8	**+3**	0.31	4.3E-13	52	0.30	5.3E-13	54
	+5	0.31	4.9E-13	53	0.30	5.7E-13	54
	+7	0.31	4.6E-13	53	0.30	5.4E-13	54
	+9	0.31	5.1E-13	54	0.30	6.0E-13	55
	+11	0.31	4.7E-13	53	0.30	5.5E-13	54
	+13	0.31	5.2E-13	54	0.30	6.3E-13	55

Table 4 (continued). Spectral properties of fluorescent probes in DSC-scrunched complexes

DNA Probes		β937			β′284		
		Q_D	J ($M^{-1}cm^{3)}$)	R_o (Å)	Q_D	J ($M^{-1}cm^{3)}$)	R_o (Å)
RP_o	**+3**	0.28	4.3E-13	51	0.32	4.4E-13	53
	+5	0.28	4.7E-13	52	0.32	4.7E-13	53
S2	**+3**	0.28	4.3E-13	51	0.29	4.4E-13	52
	+5	0.28	4.7E-13	52	0.29	4.7E-13	52
	+7	0.28	4.4E-13	52	0.29	4.5E-13	52
S4	**+3**	0.30	4.4E-13	52	0.31	4.5E-13	53
	+5	0.30	4.8E-13	53	0.31	4.8E-13	53
	+7	0.30	4.5E-13	52	0.31	4.5E-13	53
	+9	0.30	5.0E-13	53	0.31	4.8E-13	53
S6	**+3**	0.29	4.4E-13	52	0.31	4.5E-13	53
	+5	0.29	4.9E-13	53	0.31	4.9E-13	53
	+7	0.29	4.5E-13	52	0.31	4.6E-13	53
	+9	0.29	5.2E-13	53	0.31	4.9E-13	53
	+11	0.29	4.6E-13	52	0.31	4.6E-13	53
S8	**+3**	0.28	4.5E-13	52	0.30	4.7E-13	53
	+5	0.28	5.1E-13	53	0.30	5.0E-13	53
	+7	0.28	4.7E-13	52	0.30	4.7E-13	53
	+9	0.28	5.3E-13	53	0.30	5.1E-13	53
	+11	0.28	4.7E-13	52	0.30	4.8E-13	53
	+13	0.28	5.5E-13	53	0.30	5.3E-13	54

Table 5. Spectral properties of fluorescent probes in aDSC-scrunched complexes

DNA Probes		β106			β222		
		Q_D	J ($M^{-1}cm^{3)}$	R_o (Å)	Q_D	J ($M^{-1}cm^{3)}$	R_o (Å)
RP_o	**+3**	0.28	4.2E-13	51	0.33	5.4E-13	55
	+5	0.28	4.8E-13	52	0.33	5.5E-13	55
S2	**+3**	0.28	4.2E-13	51	0.30	5.4E-13	54
	+5	0.28	4.8E-13	52	0.30	5.5E-13	54
	+7	0.28	4.4E-13	52	0.30	5.3E-13	54
S4	**+3**	0.32	4.3E-13	52	0.31	5.4E-13	54
	+5	0.32	4.8E-13	53	0.31	5.6E-13	55
	+7	0.32	4.6E-13	53	0.31	5.3E-13	54
	+9	0.32	4.9E-13	54	0.31	5.5E-13	54
S6	**+3**	0.29	4.4E-13	52	0.33	5.4E-13	55
	+5	0.29	4.9E-13	53	0.33	5.6E-13	55
	+7	0.29	4.5E-13	52	0.33	5.3E-13	55
	+9	0.29	5.0E-13	53	0.33	5.7E-13	55
	+11	0.29	4.5E-13	52	0.33	5.2E-13	54
S8	**+3**	0.30	4.5E-13	52	0.29	5.5E-13	54
	+5	0.30	5.0E-13	53	0.29	5.8E-13	54
	+7	0.30	4.6E-13	53	0.29	5.4E-13	54
	+9	0.30	5.1E-13	53	0.29	6.0E-13	55
	+11	0.30	4.7E-13	53	0.29	5.4E-13	54
	+13	0.30	5.2E-13	54	0.29	6.3E-13	55

Table 5 (continued). Spectral properties of fluorescent probes in aDSC-scrunched complexes

DNA Probes		β937			β′284		
		Q_D	J ($M^{-1}cm^{3)}$)	R_o (Å)	Q_D	J ($M^{-1}cm^{3)}$)	R_o (Å)
RP_o	+3	0.28	4.5E-13	52	0.32	4.6E-13	53
	+5	0.28	4.8E-13	52	0.32	4.8E-13	53
S2	+3	0.29	4.5E-13	52	0.29	4.6E-13	52
	+5	0.29	4.8E-13	53	0.29	4.8E-13	53
	+7	0.29	4.4E-13	52	0.29	4.5E-13	52
S4	+3	0.30	4.6E-13	53	0.31	4.6E-13	53
	+5	0.30	4.8E-13	53	0.31	4.9E-13	53
	+7	0.30	4.5E-13	52	0.31	4.5E-13	53
	+9	0.30	5.0E-13	53	0.31	4.8E-13	53
S6	+3	0.28	4.5E-13	52	0.30	4.7E-13	53
	+5	0.28	5.0E-13	53	0.30	5.0E-13	53
	+7	0.28	4.5E-13	52	0.30	4.6E-13	53
	+9	0.28	5.2E-13	53	0.30	4.9E-13	53
	+11	0.28	4.5E-13	52	0.30	4.6E-13	53
S8	+3	0.28	4.6E-13	52	0.29	4.8E-13	53
	+5	0.28	5.1E-13	53	0.29	5.1E-13	53
	+7	0.28	4.7E-13	52	0.29	4.7E-13	52
	+9	0.28	5.2E-13	53	0.29	5.1E-13	53
	+11	0.28	4.7E-13	52	0.29	4.6E-13	52
	+13	0.28	5.5E-13	53	0.29	5.3E-13	53

Table 6. Steady-state anisotropies of fluorescent probes in DSC-scrunched complexes

Complex	Probe Position	anisotropy	Complex	Probe Position	anisotropy
RP_o	**β106**	0.25	S6	**β106**	0.26
	β222	0.22		**β222**	0.23
	β937	0.25		**β937**	0.24
	β′284	0.24		**β′284**	0.23
	+3	0.29		**+3**	0.31
	+5	0.27		**+5**	0.20
S2	**β106**	0.27		**+7**	0.26
	β222	0.21		**+9**	0.28
	β937	0.25		**+11**	0.29
	β′284	0.25	S8	**β106**	0.26
	+3	0.30		**β222**	0.22
	+5	0.30		**β937**	0.25
	+7	0.27		**β′284**	0.24
S4	**β106**	0.25		**+3**	0.30
	β222	0.22		**+5**	0.31
	β937	0.24		**+7**	0.27
	β′284	0.23		**+9**	0.27
	+3	0.30		**+11**	0.26
	+5	0.29		**+13**	0.31
	+7	0.28			
	+9	0.30			

Table 7. Steady-state anisotropies of fluorescent probes in aDSC-scrunched complexes

Complex	Probe Position	anisotropy	Complex	Probe Position	anisotropy
RP_o	**β106**	0.25	S6	**β106**	0.26
	β222	0.21		**β222**	0.24
	β937	0.25		**β937**	0.24
	β′284	0.25		**β′284**	0.24
	+3	0.28		**+3**	0.30
	+5	0.27		**+5**	0.21
S2	**β106**	0.26		**+7**	0.27
	β222	0.22		**+9**	0.27
	β937	0.25		**+11**	0.29
	β′284	0.26	S8	**β106**	0.26
	+3	0.29		**β222**	0.22
	+5	0.30		**β937**	0.24
	+7	0.28		**β′284**	0.25
S4	**β106**	0.24		**+3**	0.29
	β222	0.22		**+5**	0.31
	β937	0.25		**+7**	0.28
	β′284	0.22		**+9**	0.27
	+3	0.28		**+11**	0.27
	+5	0.28		**+13**	0.30
	+7	0.29			
	+9	0.30			

4.7. Systematic smFRET data: FRET efficiencies and distances

Probe-probe smFRET efficiency, *E*, was measured for each DNA-RNAP probe combination in each static scrunched state with either DSC or aDSC sequence using confocal optical microscopy with alternating laser excitation (ALEX), totaling 160 combinations (80 for DSC and 80 for aDSC). Data collection was performed in solution with freely diffusing sample at a single-molecule concentration. *E* histograms were plotted and fitted with Gaussian curves for each measurement. The *E* histograms specified equilibrium population distributions of *E*, distinguished the number of subpopulations with distinct *E*, and, for each subpopulation, defined mean *E*.
Nearly all of the complexes had a single population distribution or one subpopulation that made up more than 80% of the total population. In the cases where two subpopulations were seen, the scrunched complex was formed again as per the regular methods, but an all NTP mix was added to the newly formed scrunched complex at a concentration of 10 μM. The reaction mix was incubated for 10 min at room temp and then diluted in KT buffer for ALEX-smFRET analysis. The resulting histograms had one remaining population, which would be the one that was not functional in transcription.
Measurements were carried out in triplicates for each RNAP derivative. The probe-probe distance (*R*) was calculated from *E* and R_o using Equation (1). Values of R have precision of <5%.

Table 8. Systematic smFRET data for DSC-scrunched complexes

DNA Probes		β106			β222		
		E	R_o (Å)	R (Å)	E	R_o (Å)	R (Å)
RP_o	+3	0.61	51	47	0.71	55	47
	+5	0.65	52	47	0.70	55	48
S2	+3	0.64	51	46	0.66	54	48
	+5	0.65	52	47	0.67	54	48
	+7	0.60	52	48	0.66	54	48
S4	+3	0.68	52	46	0.67	54	48
	+5	0.65	53	48	0.65	54	49
	+7	0.67	53	47	0.64	54	49
	+9	0.60	54	50	0.65	54	49
S6	+3	0.66	51	46	0.52	55	54
	+5	0.66	53	47	0.58	55	52
	+7	0.65	52	47	0.57	55	52
	+9	0.67	53	47	0.65	55	50
	+11	0.59	52	49	0.68	55	48
S8	+3	0.65	52	47	0.49	54	54
	+5	0.65	53	48	0.51	54	54
	+7	0.61	53	49	0.47	54	55
	+9	0.69	54	47	0.55	55	53
	+11	0.73	53	45	0.48	54	55
	+13	0.61	54	50	0.43	55	58

Table 8 (continued). Systematic smFRET data for DSC-scrunched complexes

DNA Probes		β937			β′284		
		E	R_o (Å)	R (Å)	E	R_o (Å)	R (Å)
RP_o	+3	0.74	51	43	0.69	53	46
	+5	0.76	52	43	0.68	53	47
S2	+3	0.72	51	44	0.70	52	45
	+5	0.76	52	43	0.63	52	48
	+7	0.75	52	43	0.56	52	50
S4	+3	0.71	52	45	0.69	53	46
	+5	0.75	53	44	0.65	53	48
	+7	0.74	52	44	0.63	53	48
	+9	0.76	53	44	0.53	53	52
S6	+3	0.70	52	45	0.74	53	44
	+5	0.77	53	43	0.71	53	46
	+7	0.78	52	42	0.67	53	47
	+9	0.73	53	45	0.60	53	50
	+11	0.71	52	45	0.52	53	52
S8	+3	0.73	52	44	0.72	53	45
	+5	0.77	53	43	0.76	53	44
	+7	0.78	52	42	0.67	53	47
	+9	0.73	53	45	0.48	53	54
	+11	0.68	52	46	0.72	53	45
	+13	0.74	53	45	0.58	54	51

Table 9. Systematic smFRET data for aDSC-scrunched complexes

DNA Probes		β106			β222		
		E	R_o (Å)	R (Å)	E	R_o (Å)	R (Å)
RP_o	+3	0.68	51	45	0.69	55	48
	+5	0.63	52	48	0.67	55	49
S2	+3	0.65	51	46	0.61	54	50
	+5	0.63	52	48	0.59	54	51
	+7	0.60	52	48	0.61	54	50
S4	+3	0.72	52	45	0.59	54	51
	+5	0.68	53	47	0.66	55	49
	+7	0.67	53	47	0.61	54	50
	+9	0.60	54	50	0.57	54	52
S6	+3	0.70	52	45	0.58	55	52
	+5	0.64	53	48	0.64	55	50
	+7	0.62	52	48	0.66	55	49
	+9	0.61	53	49	0.48	55	56
	+11	0.59	52	49	0.54	54	53
S8	+3	0.63	52	48	0.61	54	50
	+5	0.65	53	48	0.65	54	49
	+7	0.66	53	47	0.66	54	48
	+9	0.27	53	63	0.41	55	58
	+11	0.29	53	61	0.52	54	53
	+13	0.36	54	59	0.61	55	51

Table 9 (continued). Systematic smFRET data for aDSC-scrunched complexes

DNA Probes		β937			β′284		
		E	R_o (Å)	R (Å)	E	R_o (Å)	R (Å)
RP_o	+3	0.67	52	46	0.70	53	46
	+5	0.71	52	45	0.66	53	48
S2	+3	0.70	52	45	0.56	52	50
	+5	0.72	53	45	0.66	53	47
	+7	0.67	52	46	0.59	52	49
S4	+3	0.72	53	45	0.72	53	45
	+5	0.67	53	47	0.65	53	48
	+7	0.76	52	43	0.60	53	49
	+9	0.65	53	48	0.53	53	52
S6	+3	0.73	52	44	0.72	53	45
	+5	0.63	53	48	0.56	53	51
	+7	0.75	52	43	0.60	53	49
	+9	0.73	53	45	0.56	53	51
	+11	0.70	52	45	0.51	53	52
S8	+3	0.64	52	47	0.63	53	48
	+5	0.72	53	45	0.62	53	49
	+7	0.73	52	44	0.57	52	50
	+9	0.50	53	53	0.45	53	55
	+11	0.45	52	54	0.42	52	55
	+13	0.43	53	56	0.57	53	51

4.8. Determination of NT ssDNA locations in scrunched complexes

Positions of nontemplate ssDNA nucleotides in static scrunched complexes were determined by distance-restrained rigid body docking using the FPS program. Dye distribution was modeled using an accessible-volume (AV) approach (Muschielok et al., 2008), and the AV was used to calculate an approximated fixed mean position of the dye. Measured distances to each of the four probe locations on RNAP derivatives were used as restraints to map the positions of each labeled nucleotide in each scrunched complex as represented by the appropriate nucleotide docking body (dG, dT, or dC). The other docking body was the 4YLN-derived ITC with NT-truncated nucleic acid scaffold (missing NT -5 to+2) incorporated. There were 4 donor probe restraints for each nucleotide position.

For each labeled nucleotide position (2 in RP_o, 3 in S4, etc.), 200 initial structures were obtained by a coarse search step executing a 2-Å clash tolerance. From those 200 structures, the top 20 were refined using clash tolerance of 0.5 Å. Each of those 20 FRET-derived positions was mapped to the RNAP crystal structure for each scrunched state: RP_o, S2, S4, S6, and S8. The nucleotides were mapped as colored spheres representing the 4′C of the nucleotide with all other atoms of the nucleotide removed for clarity. Spheres deemed unviable (see Materials and Methods 3.13) were shown as gray spheres. Five scrunched states were mapped for DSC scrunched complexes and five scrunched states were mapped for aDSC scrunched complexes.

4.8.1. NT ssDNA locations in RP_o for DSC and aDSC complexes

For the RP_o complexes, NT positions +3 and +5 were modeled (Figure 13). In DSC RP_o, all twenty solutions for both +3 and +5 were clustered in a small spherical space within the active center cleft. All of the solutions were within ~10 Å of the +3 and +5 positions solved for RP_o in crystal structure (Zhang et al., 2012), thus providing a validation of the methodology used in this work. Similar results occurred with the aDSC RP_o, as all twenty solutions for both +3 and +5 fit in a small spherical space within the active center cleft and were within ~12 Å of the +3 and +5 positions solved for RP_o in crystal structure. The only difference was the some of the +5 positions in aDSC RP_o were slightly downstream of the main tight cluster, along the path of the NT ssDNA. This may be due to fewer interactions with RNAP and σ in the discriminator region, allowing for more flexibility in the NT ssDNA.

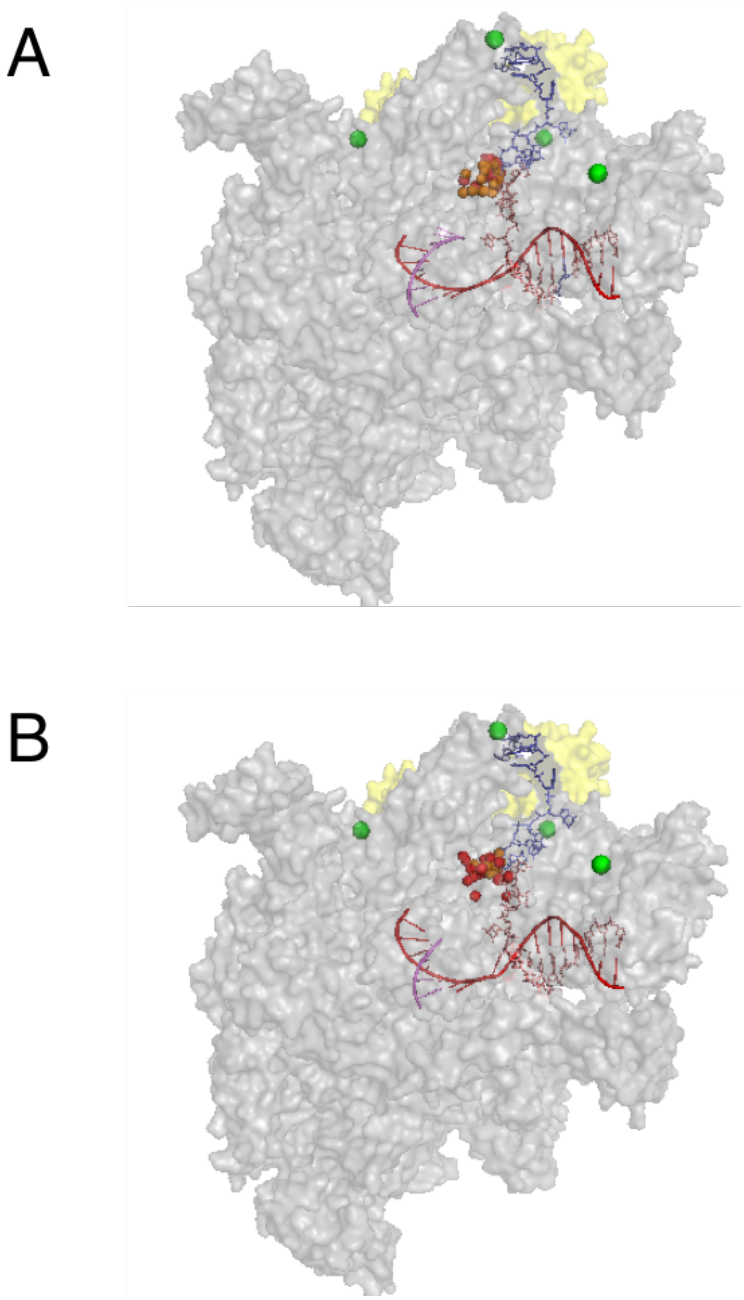

Figure 13. Comparison of NT strand positions in RP_o for DSC and aDSC complexes.
Sideview of DSC (A) and aDSC (B) complexes with top 20 FRET-defined locations of RP_o NT positions +3 and +5, indicated as orange and red spheres, respectively. RNAP reference probes sights are indicated as green spheres. RNAP core is in gray; σ is in yellow. The crystallographic structure of the RP_o scaffold used in this work has been inserted after modeling as a reference; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

4.8.2. NT ssDNA locations in S2 for DSC and aDSC complexes

For the S2 complexes, containing two nucleotides of scrunched DNA, NT positions +3, +5, and +7 were modeled (Figure 14). In DSC S2, all twenty solutions for +3, nineteen solutions for +5 and eighteen solutions for +7 were clustered in a small spherical space within the active center cleft. One position for +5 and two positions for +7 were located outside of the active center cleft on the outer surface of β. All three positions were determined to be too far away from NT ssDNA to be viable positions for study and were subsequently ignored. The remaining solutions for +3, +5, and +7 were all tightly clustered in a similar position to +3 and +5 positions in RP_o within the active center cleft, indicating that two additional nucleotides can be maintained with the active center cleft during initial transcription.

All of the modeled positions in S2 aDSC for +3, +5, and +7 were located within the active center cleft, with most of them found in the same spherical space as the S2 DSC solutions. As in aDSC RP_o, a few of the +5 solutions, and now a few of the +7 solutions, were slightly downstream of the main grouping along the path of NT ssDNA. This further suggests more flexibility of the NT strand within the ssDNA region in scrunched ITCs. The tight bunching of most of the aDSC NT solutions, as well as the placement of all model solutions within the active center cleft, does indicate that two nucleotides of scrunched DNA can be accommodated within the active center cleft in aDSC complexes.

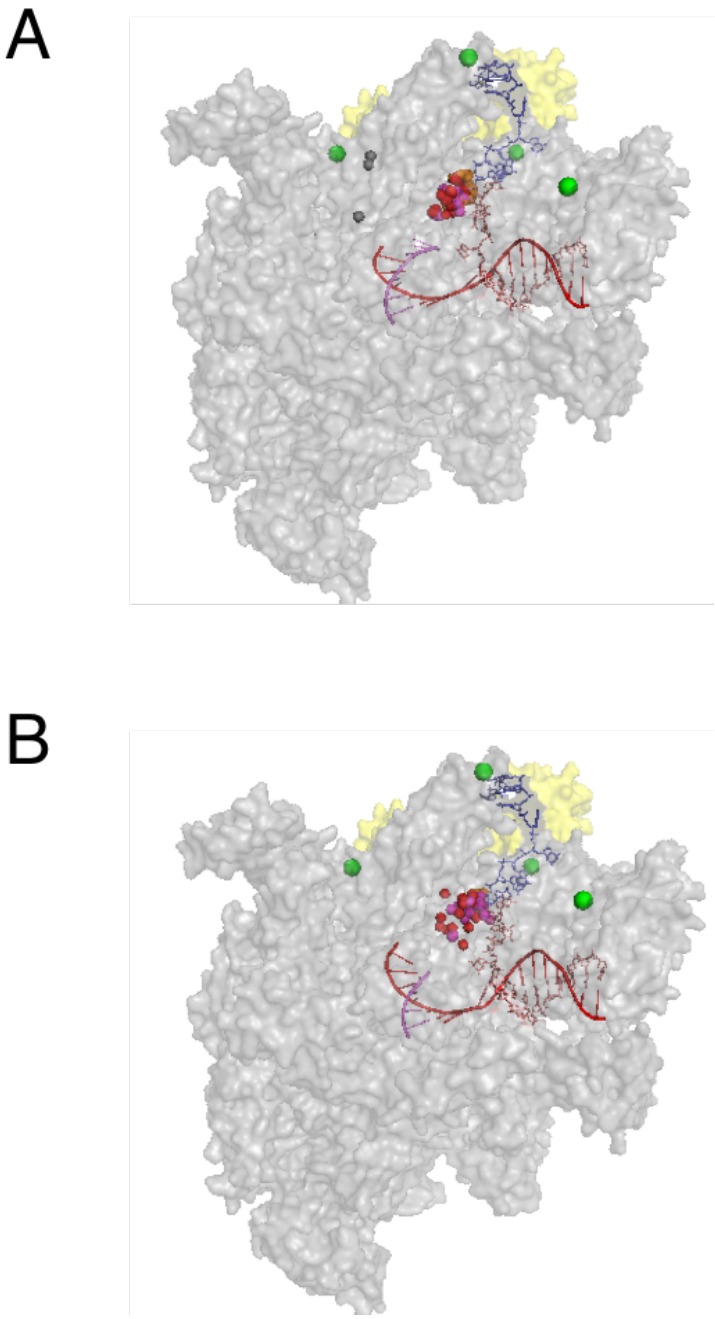

Figure 14. Comparison of NT strand positions in S2 for DSC and aDSC complexes. Sideview of DSC (A) and aDSC (B) complexes with top 20 FRET-defined locations of S2 NT positions +3, +5, and +7, indicated as orange, red, and magenta spheres, respectively. Reference spheres that are positionally impossible are in dark gray. RNAP reference probes sights are indicated as green spheres. RNAP core is in gray; σ is in yellow. The crystallographic structure of the RP_o scaffold used in this work has been inserted after modeling as a reference; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

4.8.3. NT ssDNA locations in S4 for DSC and aDSC complexes

For the S4 complexes, containing four nucleotides of scrunched DNA, NT positions +3, +5, +7, and +9 were modeled (Figure 15). In DSC S4, all twenty solutions for +3, +5, and +7 positions, and eighteen solutions for +9 were located within the active center cleft, most of which were clustered in a small spherical space within the active center cleft. Two positions for +9 were located outside of the active center cleft on the outer surface of β in locations determined to be too far away from NT ssDNA to be viable positions for study, and were subsequently ignored. A few other +9 solutions were found in the active center cleft, but more towards the base of the cleft. These could represent a situation where the bases stack and contort more inward in the transcription bubble during scrunching. However, these may not be viable positions, as they are located in a region that may cause steric clash with the nitrogenous bases of the template strand ssDNA normally present in a full transcription bubble. The remaining solutions for +3, +5, +7, and +9 are all tightly clustered in a similar location as in RP_o and S2, indicating that four additional nucleotides can be maintained with the active center cleft during initial transcription.

All of the modeled positions in S4 aDSC for +3, +5, and +9 were located within the active center cleft, with most of them found in the same spherical space as the S4 DSC solutions. Two positions for +7 were located well outside of the active center cleft on the outer surface of β, and were determined unviable. As in aDSC RP_o and S2, the positions downstream of +3 were not quite with the main dense spheroid. Solutions for +5 were minimally downstream along the NT ssDNA path, and solutions for +7 and +9 could be

found further downstream along this path. This further suggests flexibility of the NT strand in the ssDNA region in scrunched ITCs. The tight bunching of most of the aDSC NT solutions, as well as the placement of all viable model solutions within the active center cleft does indicate that four nucleotides of scrunched DNA can be accommodated within the active center cleft in aDSC complexes.

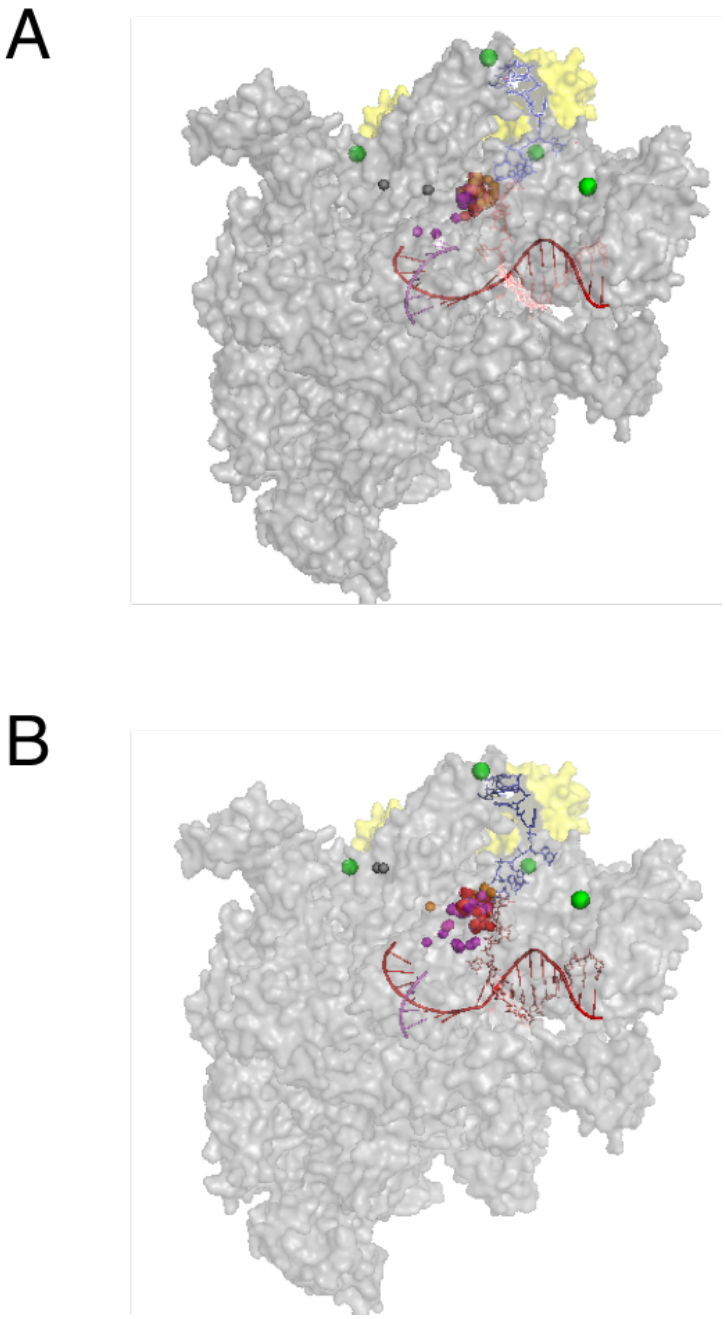

Figure 15. Comparison of NT strand positions in S4 for DSC and aDSC complexes. Sideview of DSC (A) and aDSC (B) complexes with top 20 FRET-defined locations of S4 NT positions +3, +5, +7, and +9, indicated as orange, red, magenta, and purple spheres, respectively. Reference spheres that are positionally impossible are in dark gray. RNAP reference probes sights are indicated as green spheres. RNAP core is in gray; σ is in yellow. The crystallographic structure of the RP_o scaffold used in this work has been inserted after modeling as a reference; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

4.8.4. NT ssDNA locations in S6 for DSC and aDSC complexes

For the S6 complexes, containing six nucleotides of scrunched DNA, NT positions +3, +5, +7, +9 and +11 were modeled (Figure 16). In DSC S6, all twenty solutions for +3, +5, and +9 positions, and eighteen solutions for +11 were located within the active center cleft, most of which were clustered in a small spherical space within the active center cleft. Two positions for +11 were modeled outside of the active center cleft on the outer surface of β in functionally unviable locations, and thus ignored. Seven of the twenty positions for +7 were not modeled within the active center cleft, with six of those being in unviable positions spread around the outer surface of RNAP. One of the +7 docked locations was located slightly outside of the cleft near the interface between the downstream portion of the -10 element and σ, and could be a possible position where the NT scrunched DNA could exit out of the cleft. However, a majority of the +7 positions were still found within the active center cleft, and nearly every solution for +3, +5, +7, and +9 were all clustered in a similar location as in RP_o, S2, and S4, indicating that six additional nucleotides can be maintained with the active center cleft during initial transcription.

All of the modeled positions in S6 aDSC for +5, and most of the positions for +3, +7, +9, and +11 were located within the active center cleft in in the same spherical locations as RP_o, S2, and S4, but they were much less compact in S6 aDSC complexes. Eight positions for +7 and four positions for +11 were located well outside of the active center cleft on the outer surface of β, and were determined unviable. Three positions each for +3 and +9 were located just outside the active center cleft near the interface between the -

10 element and σ, suggesting possible locations where the scrunched DNA could exit the active center cleft. The much less dense grouping within the active center cleft suggests much more sampling of locations by the NT ssDNA in the S6 complex than was seen in the DSC complex, due to less stabilizing interactions with the protein. While not as tightly bunched as in previous complexes, the preponderance of solutions for each NT position does indicate that six nucleotides of scrunched DNA can be accommodated within the active center cleft in aDSC complexes. However, sampling outside of the active center cleft does begin to occur for aDSC complexes at six nucleotides of scrunched DNA.

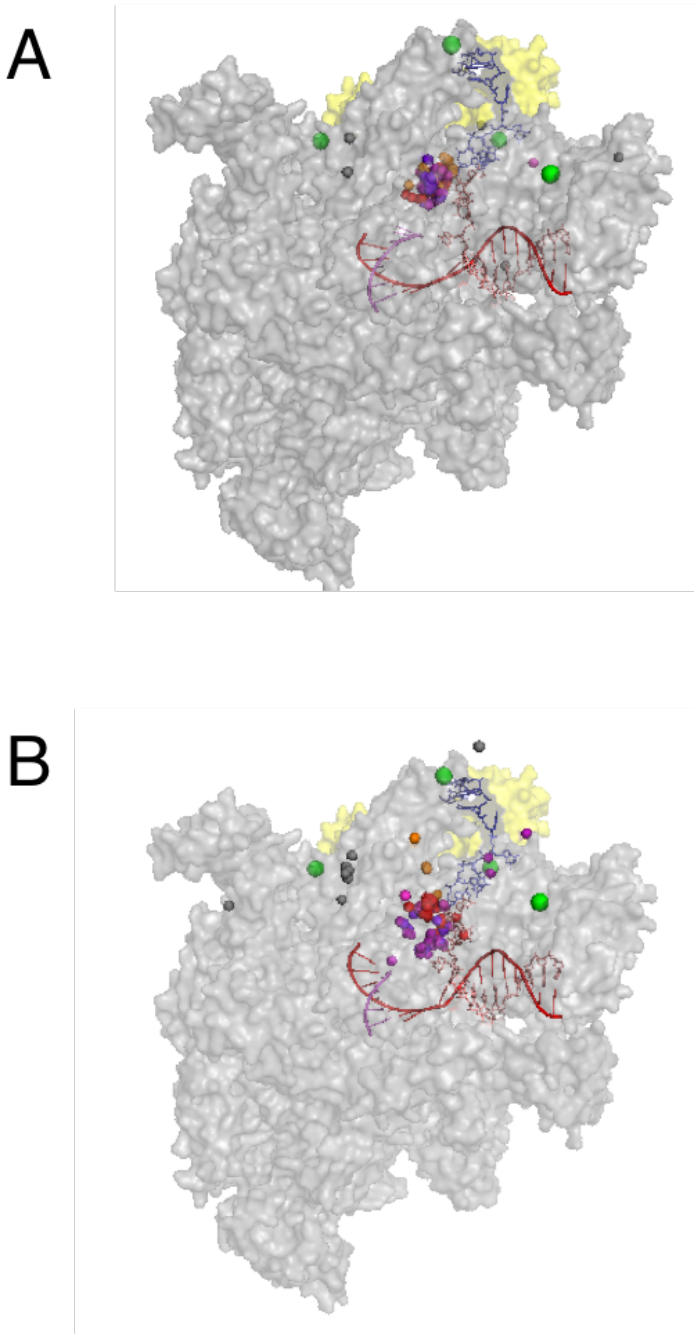

Figure 16. Comparison of NT strand positions in S6 for DSC and aDSC complexes. Sideview of DSC (A) and aDSC (B) complexes with top 20 FRET-defined locations of S6 NT positions +3, +5, +7, +9, and +11, indicated as orange, red, magenta, purple, and dark purple spheres, respectively. Reference spheres that are positionally impossible are in dark gray. RNAP reference probes sights are indicated as green spheres. RNAP core is in gray; σ is in yellow. The crystallographic structure of the RP_o scaffold used in this work has been inserted after modeling as a reference; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

4.8.5. NT ssDNA locations in S8 for DSC and aDSC complexes

For the S8 complexes, containing eight nucleotides of scrunched DNA, NT position +3, +5, +7, +9, +11 and +13 were modeled (Figure 17). In DSC S8, +9, +11, and +13 had all twenty solutions within the active center cleft, while +3 had nineteen, +5 had seventeen, and +7 had ten. All of the positions for +3 and +5, and five of the positions for +7 were sampled to unviable locations around the outside surface of RNAP. The other five solutions for +7 docked locations was located radially above the active center cleft and could be possible positions where the NT scrunched DNA could exit out of the cleft. Most, if not all, of the positions for +3, +5, +9, +11, and +13 were still found within the active center cleft. The dispersion of the solutions within the cleft was much more wide than previous scrunched states, suggesting more possible sampling of locations in S8 complexes. While there was indication that some of the NT ssDNA is exiting the active center cleft in S8 complexes, especially the most central portions of the ssDNA (+7) these results indicate that most of the eight additional nucleotides can be accommodated within the active center cleft during initial transcription.

In aDSC S8, only +13 had all twenty solutions within the active center cleft. +3 had seventeen, +5 had seventeen, +7 had ten, +9 had eleven, and +11 had ten docking solutions within the active center cleft. These solutions formed two separate clusters within the active center cleft, with the upstream NT locations (+3, +5, +7) closer to upstream transcription bubble DNA and the downstream positions (+9, +11, +13) located closer to the downstream junction of the transcription bubble. There were many unviable sampling locations for aDSC S8: +3 had two, +5 had three, +7 had six, +9 had eight, and

+11 had ten. Four positions each for +7 and one position each for +3 and +9 were located radially above the active center cleft, suggesting possible locations where the scrunched DNA could exit the active center cleft. As in DSC S8, the was some indication in aDSC that some of the NT ssDNA is exiting the active center cleft in S8 complexes, especially the most central portions of the NT ssDNA (+7 and +9). Additionally, there was some sampling outside the active center cleft by the most upstream NT ssDNA (+3) in both S6 and S8 aDSC complexes, which was not seen in the DSC complexes. While there was indication that some of the NT ssDNA is exiting the active center cleft in S8 aDSC complexes, especially the most central portions of the NT ssDNA, each position in aDSC S8 had at least half of its solutions within the active center cleft, indicating that most of the eight additional nucleotides can be accommodated within the active center cleft during initial transcription.

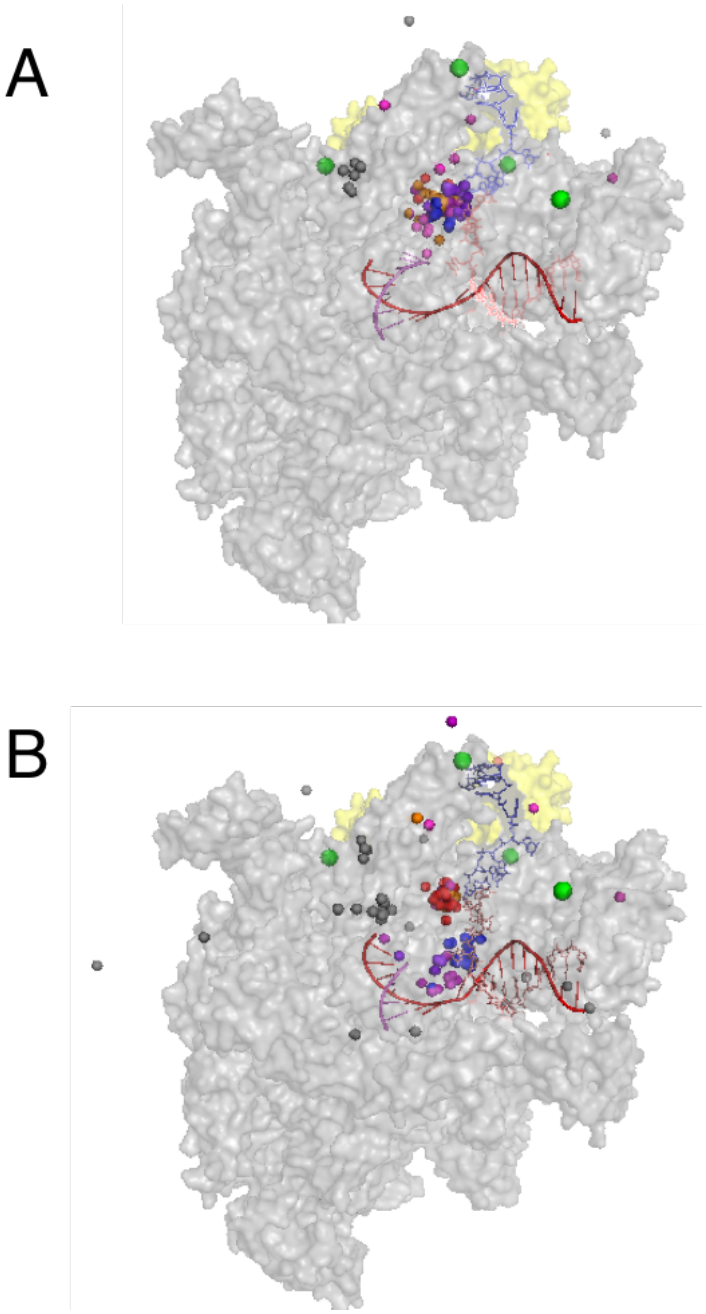

Figure 17. Comparison of NT strand positions in S8 for DSC and aDSC complexes. Sideview of DSC (A) and aDSC (B) complexes with top 20 FRET-defined locations of S8 NT strand positions +3, +5, +7, +9, +11 and +13, indicated as orange, red, magenta, purple, dark purple, and dark blue spheres, respectively. Reference spheres that are positionally impossible are in dark gray. RNAP reference probes sights are indicated as green spheres. RNAP core is in gray; σ is in yellow. The crystallographic structure of the RP_o scaffold used in this work has been inserted after modeling as a reference; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

4.9. Comparison of DSC NT ssDNA locations to scrunched crystal structures

Concurrent work in the lab was done forming these static scrunched complexes with the same scaffold DNA (Zhang, unpublished work). Structures were solved for DSC S2, S4, and S6 at 2.9 Å, 3.1 Å, and 3.5 Å resolution, respectively. The top 10 FRET-derived solutions for each labeled NT position were aligned to the crystal structures for comparison (Figure 18). In S2, the entire DNA structure was resolved, which allowed us to compare our data with the crystal results. For all three positions, +3, +5, and +7, the modeling results were all within 10 Å of their corresponding nucleotide in the structure. In S4, all but four nucleotides of the NT strand were resolved, including +3 and +5. The FRET-derived solutions for +3 and +5 are all within 12 Å of their corresponding nucleotide in the structure. Both the S2 and S4 results validate our methodology based on the close proximity of FRET-modelled positions to crystal structure positions. In S6, nearly all of the nontemplate strand between the -10 element and the downstream dsDNA was unresolved. The docked structures for +3, +5, +7, and +9 do form a tight grouping with the active center cleft, suggesting they can be accommodated within the cleft. Only three positions, all in +7, are not in the cleft, and two of those solutions sit on the outside of β (removed in the crystal structure) and are functionally unviable based on distance.

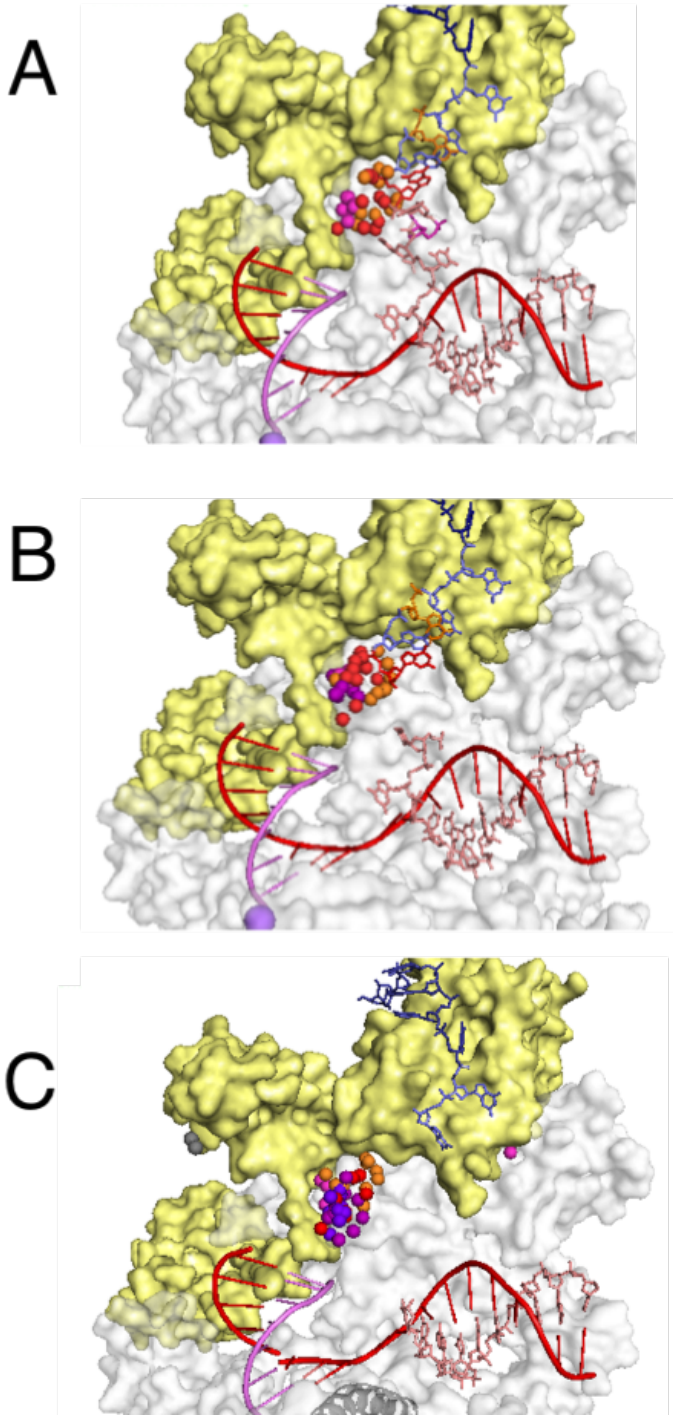

Figure 18. FRET-derived NT strand positions fit to crystal structures of the static scrunched complexes used in this work.
Crystal structures of S2, S4, and S6 using the static scrunched scaffolds were solved in *Thermus thermophilus* (Zhang, unpublished work). RNAP core is in light gray; σ is in yellow; –10 element of DNA nontemplate strand in blue; discriminator element of DNA nontemplate strand in light blue; rest of DNA nontemplate strand in pink; DNA template strand in red; RNA in violet.

(A) Sideview of S2 complex, zoomed in to view active center. Top 10 FRET-defined positions of S2 NT strand positions +3, +5, and +7 are indicated as orange, red, and magenta spheres, respectively. The +3, +5, and +7 nucleotides solved in the crystal structure have been colored to match the probes.
(B) Sideview of S4 complex, zoomed in to view active center. Top 10 FRET-defined positions of S4 NT strand positions +3, +5, +7, and +9 are indicated as orange, red, magenta, and purple spheres respectively. The +3 and +5 nucleotides solved in the crystal structure have been colored to match the probes.
(C) Sideview of S6 complex, zoomed in to view active center. Top 10 FRET-defined positions of S6 NT strand positions +3, +5, +7, +9, and +11 are indicated as orange, red, magenta, purple, and dark purple spheres, respectively.

5. Discussion

In order to further understand the mechanism of DNA scrunching during initial transcription, this work sought to define the path of nontemplate strand single-stranded DNA at multiple scrunched lengths. In order to achieve this, static initial transcription complexes were generated containing nontemplate strand DNA with 0, 2, 4, 6, or 8 nucleotides scrunched de novo. These complexes were well characterized as stable and functional for use in FRET-derived model development.

Using the static scrunched complexes, this work created several detailed models of initial transcribing complexes at many levels: 0 nucleotides scrunched (RP_o), 2 nucleotides scrunched (S2), 4 nucleotides scrunched (S4), 6 nucleotides scrunched (S6), and eight nucleotides scrunched (S8). The models were developed using smFRET determined distances between nucleotide positions on the nontemplate strand and surrounding positions in RNAP. Mapping of the DNA positions in the nontemplate strand was performed using FRET-derived distance restraining docking methodology.

The RP_o model was used as a method validation and showed to be accurate to within 10 Å of previously solved crystallographic structures of the nontemplate strand in RP_o. Since each model increased scrunching in the nontemplate strand by two nucleotides, they provided a stepwise method to investigate the trajectory of the scrunched segments.

The models demonstrated that scrunched nontemplate strand DNA can be accommodated in the active center cleft up to at least six nucleotides in length and possibly up to eight nucleotides in length.

The models indicated that, as the level of scrunching increases, the more central portion of the scrunched DNA, approximately 7 nucleotides downstream of the downstream edge of the -10 element (+7 in this work), became less and less constrained in the active center. At six nucleotides scrunched, there was a small portion of positional sampling by the +7 position. At eight nucleotides scrunched, that outward sampling by the central DNA portion became more pronounced. This is logical, since the nontemplate strand DNA in transcription bubble is constrained upstream by the -10 element and downstream by the dsDNA duplex, while the more central portion of the DNA would have less restrictions of movement.

This work also created models to investigate the role that the discriminator element in nontemplate strand DNA may play during scrunching. A consensus discriminator element, comprised of a GGG sequence immediately downstream of the -10 element, has been shown to make strong interactions with σR1.2 during initial transcription, further stabilizing the NT ssDNA in the transcription bubble. The same stepwise set of scrunched models, from RP_o to S8, were developed with a weak anticonsensus discriminator element (aDSC, with CCC sequence), and the pathway of nontemplate strand DNA was investigated and compared with the previous models containing a consensus discriminator sequence (DSC). The aDSC RP_o model further validated the methodology used in this work as its mapped positions were within 12 Å of their crystallographically-determined counterparts. As the scrunching increased up to two and four nucleotides scrunched, the aDSC-modeled positions demonstrated more flexibility in their locations within the active center cleft, as compared with the denser sphere of DSC-modeled positions. At six and eight nucleotides scrunched, the middle portion (+7, +9)

of the aDSC scrunched DNA began to model to portions outside of the active center cleft to a much larger degree than DSC models showed. This suggests two things: (1) the interactions of the consensus discriminator sequence with the RNAP seems to have a constraining effect on the movement of nontemplate strand DNA during scrunching, and (2) without that effect, as was the case with the antidiscriminator element, the scrunched nontemplate strand DNA is able to sample more positions within the active center cleft, and may be able to expand out of the active center cleft more quickly. Additionally, the +3 position, which is the most downstream position of the discriminator element, began to model some positions outside of the active center cleft at six and eight nucleotides scrunched in aDSC complexes. This further highlights the greater flexibility seen in the aDSC scrunched complexes, both in the more central portion of the NT ssDNA and the upstream portion of the NT ssDNA that includes the discriminator element.

These FRET-derived rigid body docking models provided a stepwise look into the path of scrunched nontemplate strand DNA and sequence elements that may influence that pathway. Further refinement of the models would further elucidate details of the scrunching pathway. Incorporation of more reference probe positions in RNAP would increase the restraints and thus improve the accuracy of determining an unknown position (Knight et al., 2005). Beyond sharpening the accuracy of the model positions, it would be useful investigate the scrunching pathway beyond eight nucleotides scrunched in order to determine the point of consistent emergence from the active center cleft. Attempts at forming larger complexes (i.e., S10) using the scrunched scaffold method in this study were unsuccessful in forming competent, functional scrunched complexes.

Elucidating the path of NT ssDNA provided some important insights, but it clarifies only one half of the mystery, as the pathway of the template strand DNA during scrunching would provide a full picture of the scrunching trajectories. Recent work has suggested that scrunching of the template strand DNA during initial transcription plays a role in displacing σR3.2 from the RNA exit channel and helping promoter escape, but this was only able to be investigated up to three nucleotides scrunched (Winkelman et al., 2015). Full transcription bubble scaffolds intended to form static scrunched complexes do novo were investigated during this study, but were unable to be formed.

References

Bai, L., Santangelo, T.J., and Wang, M.D. (2006). Single-molecule analysis of RNA polymerase transcription. Annu Rev Biophys Biomol Struct *35*, 343-360.

Brodolin, K., Zenkin, N., and Severinov, K. (2005). Remodeling of the sigma70 subunit non-template DNA strand contacts during the final step of transcription initiation. J Mol Biol *350*, 930-937.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. Nature *221*, 43-46.

Busby, S., and Ebright, R.H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell *79*, 743-746.

Cai, Q., Kusnetzow, A.K., Hideg, K., Price, E.A., Haworth, I.S., and Qin, P.Z. (2007). Nanometer distance measurements in RNA using site-directed spin labeling. Biophys J *93*, 2110-2117.

Callaci, S., Heyduk, E., and Heyduk, T. (1998). Conformational changes of Escherichia coli RNA polymerase sigma70 factor induced by binding to the core enzyme. J Biol Chem *273*, 32995-33001.

Callaci, S., and Heyduk, T. (1998). Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from Escherichia coli RNA polymerase are modulated by an interaction with the core enzyme. Biochemistry *37*, 3312-3320.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell *9*, 527-539.

Cantor, C.R. (1980). The conformation of biological macromolecules / Charles R. Cantor, Paul R. Schimmel (San Francisco: W. H. Freeman).

Carpousis, A.J., and Gralla, J.D. (1985). Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation. J Mol Biol *183*, 165-177.

Chakraborty, A., Wang, D., Ebright, Y.W., and Ebright, R.H. (2010). Azide-specific labeling of biomolecules by Staudinger-Bertozzi ligation phosphine derivatives of fluorescent probes suitable for single-molecule fluorescence spectroscopy. Methods Enzymol *472*, 19-30.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R.*, et al.* (2012). Opening and closing of the bacterial RNA polymerase clamp. Science *337*, 591-595.

Chen, R.F., and Bowman, R.L. (1965). Fluorescence Polarization: Measurement with Ultraviolet-Polarizing Filters in a Spectrophotofluorometer. Science *147*, 729-732.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. Journal of the American Chemical Society *124*, 9026-9027.

Clegg, R.M. (1992). Fluorescence resonance energy transfer and nucleic acids. Methods in enzymology *211*, 353-388.

Cramer, P. (2002). Multisubunit RNA polymerases. Curr Opin Struct Biol *12*, 89-97.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J.*, et al.* (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337-352.

Darst, S.A. (2001). Bacterial RNA polymerase. Curr Opin Struct Biol *11*, 155-162.

Darst, S.A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. Proc Natl Acad Sci U S A *99*, 4296-4301.

Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998). Insights into Escherichia coli RNA polymerase structure from a combination of x-ray and electron crystallography. J Struct Biol *124*, 115-122.

deHaseth, P.L., Zupancic, M.L., and Record, M.T., Jr. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. J Bacteriol *180*, 3019-3025.

Deniz, A.A., Dahan, M., Grunwell, J.R., Ha, T., Faulhaber, A.E., Chemla, D.S., Weiss, S., and Schultz, P.G. (1999). Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Forster distance dependence and subpopulations. Proc Natl Acad Sci U S A *96*, 3670-3675.

Dombroski, A.J., Walter, W.A., and Gross, C.A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. Genes Dev *7*, 2446-2455.

Dombroski, A.J., Walter, W.A., Record, M.T., Jr., Siegele, D.A., and Gross, C.A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell *70*, 501-512.

Ebright, R.H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol *304*, 687-698.

Ebright, R.H., and Busby, S. (1995). The Escherichia coli RNA polymerase alpha subunit: structure and function. Curr Opin Genet Dev *5*, 197-203.

Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimasauskas, S.*, et al.* (2006). A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. Mol Cell *23*, 97-107.

Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol *63*, 141-155.

Ha, T. (2001). Single-molecule fluorescence resonance energy transfer. Methods *25*, 78-86.

Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R., and Weiss, S. (1996). Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proc Natl Acad Sci U S A *93*, 6264-6268.

Harley, C.B., and Reynolds, R.P. (1987). Analysis of E. coli promoter sequences. Nucleic Acids Res *15*, 2343-2361.

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell *125*, 1069-1082.

Haugen, S.P., Ross, W., Manrique, M., and Gourse, R.L. (2008). Fine structure of the promoter-sigma region 1.2 interaction. Proc Natl Acad Sci U S A *105*, 3292-3297.

Helmann, J.D., and Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. Annu Rev Biochem *57*, 839-872.

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

Igarashi, K., and Ishihama, A. (1991). Bipartite functional map of the E. coli RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. Cell *65*, 1015-1022.

Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P.J., Berger, S., Restle, T., Goody, R.S., Gohlke, H., and Seidel, C.A. (2012). A toolkit and benchmark study for FRET-restrained high-precision structural modeling. Nat Methods *9*, 1218-1225.

Kapanidis, A.N., Laurence, T.A., Lee, N.K., Margeat, E., Kong, X., and Weiss, S. (2005). Alternating-laser excitation of single molecules. Acc Chem Res *38*, 523-533.

Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., and Weiss, S. (2004). Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. Proc Natl Acad Sci U S A *101*, 8936-8941.

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

Karstens, T., and Kobs, K. (1980). Rhodamine-B and Rhodamine-101 as Reference Substances for Fluorescence Quantum Yield Measurements. J Phys Chem-Us *84*, 1871-1872.

Knight, J.L., Mekler, V., Mukhopadhyay, J., Ebright, R.H., and Levy, R.M. (2005). Distance-restrained docking of rifampicin and rifamycin SV to RNA polymerase using

systematic FRET measurements: developing benchmarks of model quality and reliability. Biophys J *88*, 925-938.

Komissarova, N., and Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. Proc Natl Acad Sci U S A *95*, 14699-14704.

Korzheva, N., and Mustaev, A. (2001). Transcription elongation complex: structure and function. Curr Opin Microbiol *4*, 119-125.

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

Lakowicz, J.R. (2006). Principles of fluorescence spectroscopy, 3rd edn (New York: Springer).

Lamond, A.I., and Travers, A.A. (1985). Stringent control of bacterial transcription. Cell *41*, 6-8.

Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M., and Ebright, R.H. (2004). Catabolite activator protein: DNA binding and transcription activation. Curr Opin Struct Biol *14*, 10-20.

Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2005). Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys J *88*, 2939-2953.

Lilley, D.M., and Wilson, T.J. (2000). Fluorescence resonance energy transfer as a structural tool for nucleic acids. Curr Opin Chem Biol *4*, 507-517.

Malhotra, A., Severinova, E., and Darst, S.A. (1996). Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase. Cell *87*, 127-136.

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

Mekler, V., Minakhin, L., and Severinov, K. (2011). A critical role of downstream RNA polymerase-promoter interactions in the formation of initiation complex. J Biol Chem *286*, 22600-22608.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.

Mitchell, J.E., Zheng, D., Busby, S.J., and Minchin, S.D. (2003). Identification and analysis of 'extended -10' promoters in Escherichia coli. Nucleic Acids Res *31*, 4689-4695.

Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? Mol Cell *20*, 335-345.

Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R.*, et al.* (2008). The RNA polymerase "switch region" is a target for inhibitors. Cell *135*, 295-307.

Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W., and Ebright, R.H. (2001). Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell *106*, 453-463.

Mukhopadhyay, J., Mekler, V., Kortkhonjia, E., Kapanidis, A.N., Ebright, Y.W., and Ebright, R.H. (2003). Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. Methods Enzymol *371*, 144-159.

Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R.M., and Ebright, R.H. (2004). Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. Mol Cell *14*, 739-751.

Murakami, K.S. (2013). X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. J Biol Chem *288*, 9126-9134.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol *13*, 31-39.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science *296*, 1285-1290.

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

Muschielok, A., Andrecka, J., Jawhari, A., Bruckner, F., Cramer, P., and Michaelis, J. (2008). A nano-positioning system for macromolecular structural analysis. Nat Methods *5*, 965-971.

Naryshkin, N., Kim, Y., Dong, Q., and Ebright, R.H. (2001). Site-specific protein-DNA photocrosslinking. Analysis of bacterial transcription initiation complexes. Methods Mol Biol *148*, 337-361.

Nickels, B.E., and Hochschild, A. (2004). Regulation of RNA polymerase through the secondary channel. Cell *118*, 281-284.

Pemberton, I.K., Muskhelishvili, G., Travers, A.A., and Buckle, M. (2000). The G+C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. J Mol Biol *299*, 859-864.

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

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science *262*, 1407-1413.

Saxon, E., and Bertozzi, C.R. (2000). Cell surface engineering by a modified Staudinger reaction. Science *287*, 2007-2010.

Schuler, B., and Eaton, W.A. (2008). Protein folding studied by single-molecule FRET. Curr Opin Struct Biol *18*, 16-26.

Selvin, P.R. (2000). The renaissance of fluorescence resonance energy transfer. Nat Struct Biol *7*, 730-734.

Severinov, K., Soushko, M., Goldfarb, A., and Nikiforov, V. (1993). Rifampicin region revisited. New rifampicin-resistant and streptolydigin-resistant mutants in the beta subunit of Escherichia coli RNA polymerase. J Biol Chem *268*, 14820-14825.

Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E.N., Roberts, J.W., Chait, B.T., and Darst, S.A. (1996). Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit. J Mol Biol *263*, 637-647.

Sharp, M.M., Chan, C.L., Lu, C.Z., Marr, M.T., Nechaev, S., Merritt, E.W., Severinov, K., Roberts, J.W., and Gross, C.A. (1999). The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. Genes Dev *13*, 3015-3026.

Srivastava, A., Talaue, M., Liu, S., Degen, D., Ebright, R.Y., Sineva, E., Chakraborty, A., Druzhinin, S.Y., Chatterjee, S., Mukhopadhyay, J.*, et al.* (2011). New target for inhibition of bacterial RNA polymerase: 'switch region'. Curr Opin Microbiol *14*, 532-543.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231-232.

Straney, D.C., and Crothers, D.M. (1987). A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. J Mol Biol *193*, 267-278.

Stryer, L., and Haugland, R.P. (1967). Energy transfer: a spectroscopic ruler. Proc Natl Acad Sci U S A *58*, 719-726.

Tang, H., Severinov, K., Goldfarb, A., and Ebright, R.H. (1995). Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant Escherichia coli RNA polymerase. Proc Natl Acad Sci U S A *92*, 4902-4906.

Travers, A.A. (1980). Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. J Bacteriol *141*, 973-976.

Travers, A.A., and Burgessrr (1969). Cyclic re-use of the RNA polymerase sigma factor. Nature *222*, 537-540.

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

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163-168.

Werner, F. (2007). Structure and function of archaeal RNA polymerases. Mol Microbiol *65*, 1395-1404.

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol *9*, 85-98.

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

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

Wu, P., and Brand, L. (1992). Orientation factor in steady-state and time-resolved resonance energy transfer measurements. Biochemistry *31*, 7939-7947.

Xu, Q., and Rutgers University. Graduate School--New Brunswick. (2013). Displacement of DNA-molecular-mimic σR1.1 from the RNA polymerase active-center cleft single-molecule fluorescence analysis, pp. xi, 102 p.

Young, B.A., Gruber, T.M., and Gross, C.A. (2002). Views of transcription initiation. Cell *109*, 417-420.

Young, T.S., Ahmad, I., Yin, J.A., and Schultz, P.G. (2010). An enhanced system for unnatural amino acid mutagenesis in E. coli. J Mol Biol *395*, 361-374.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811-824.

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

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