DISPLACEMENT OF DNA-MOLECULAR-MIMIC σR1.1 FROM THE RNA POLYMERASE ACTIVE-CENTER CLEFT: SINGLE-MOLECULE FLUORESCENCE ANALYSIS

by

QUMIAO XU

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry 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, 2013

ABSTRACT OF THE DISSERTATION

Displacement of DNA-molecular-mimic σR1.1 from the RNA polymerase active-center cleft: single-molecule fluorescence analysis

By QUMIAO XU

Dissertation Director:

Richard H. Ebright

The N-terminal conserved region of σ (σR1.1) is a highly negatively charged segment with 90-100 amino acids. In free σ, it interacts with DNA-binding determinants on σ, preventing free σ from association with promoter DNA. In RNA polymerase (RNAP) holoenzyme, σR1.1 no longer masks DNA-binding determinants on σ, therefore enabling association with promoter DNA. Deletion of σR1.1 has significant effects on rates of formation of RNAP-promoter open complex (RPo).

The solution structure of σR1.1 has been solved by NMR. Ensemble fluorescence resonance energy transfer (FRET) analysis indicate that σR1.1 acts as a "molecular-mimic" of DNA by occupying the active-center cleft in RNAP holoenzyme, and must be displaced out of the active-center cleft upon formation of RPo. However, the precise positions and orientations of σR1.1 in holoenzyme and RPo have remained uncertain.

Recent crystal structures of σR1.1 in RNAP holoenzyme from two groups—Murakami and co-workers and Darst and co-workers—place σR1.1 in RNAP holoenzyme inside the RNAP active-center cleft, consistent with the ensemble FRET analysis, but the folds and

the rotational orientations of σR1.1 in the crystal structures from the two groups are different.

In this work, I have used systematic single-molecule FRET and distance-restrained docking to define the positions and rotational orientations of σR1.1 in RNAP holoenzyme in solution and in RPo in solution. The results for RNAP holoenzyme indicate that, in RNAP holoenzyme in solution, σR1.1 is located inside the active-center cleft, in a position and rotational orientation consistent with the crystal structure from Darst and co-workers. The results for RPo indicate that, in RPo in solution, σR1.1 is located outside the active-center cleft—at least 40 Å away from its position in holoenzyme—is positioned between the RNAP β' jaw and β dispensable region 1 (βDR1), and potentially makes direct protein-protein interactions with βDR1.

Deletion of βDR1 affects the rate and temperature-dependence of formation of RPo and alters the footprint of RPo. My results suggest that, upon formation of RPo, σR1.1 is displaced from the RNAP active-center cleft to a binding site on βDR1, and, as such, provide an explanation for previously detected effects of deleting βDR1.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor, Prof. Richard Ebright, for his support and guidance during the past six years. When I first joined the lab, I had very little experience on research and was unfamiliar with most experimental techniques. Many tumbles and mistakes were along the way. Prof. Ebright provided me with the opportunity to learn various cutting-edge techniques, and with his keen scientific insights and thorough resources, directed me to the right path to solve the problems in my projects.

I owe my gratitude to my committee members, Prof. Eddy Arnold, Prof. Bryce Nickels and Prof. Wilma Olson for their valuable advice on the work.

I am grateful to Dr. Jennifer Knight, for developing the structural models, which is an essential part of this work. Without her efforts, this work would not have been a complete story for dissertation.

I would like to thank Dr. Yon Ebright for her care and help during my most difficult time at graduate school. She has been a teacher and also a close friend to me. Her enthusiasm on work and curiosity about unknowns are traits I wish to acquire.

I would like to extend my gratitude to each of the past and present members of Ebright lab, especially Dr. Dongye Wang, Dr. Jayanta Mukhopadhyay, Dr. Sukhendu Mandal, Dr. Yu Zhang, Dr. Yu Feng, Dr. Juan Shen, Dr. Sergey Druzhinin, Dr. Anirban Chakraborty, Dr. Charles O'Brien, Mr. David Degen, Mr. Adam Hasemeyer, Ms Shuang (Carol) Liu, Ms Miaoxin Lin and Ms Soma Mandal for their help and advice.

I am also thankful to my current employer—Zoetis, Inc., and my manager Dr. Lisa Bergeron, for her understanding during my preparation for defense.

I want to share my joy and pride with my beloved boyfriend, Peng Du—without his patience, encouragement and love, it would have been very difficult for me to accomplish the task alone.

I will be forever indebted to my dearest parents—Mr. Xinhua Xu and Mrs. Qiongfang Li. Their enduring love has been my strongest support and will always be my source of strength no matter what difficulties I will face.

DEDICATION

Dedicated to my parents with love

Table of contents

List of tables

List of illustrations

1. Introduction

In cellular organisms, gene expression starts from transcription, the process by which genetic information encoded in DNA is copied into RNA. Most regulation of gene expression occurs at the level of transcription. RNA polymerase (RNAP) is the enzyme responsible for synthesizing RNA from DNA template, and is the target, directly or indirectly, of most regulation of transcription. In its simplest bacterial form, the enzyme comprises five subunits (β′, β, α^{I}, α^{II}, and ω), with a total molecular mass of around 400 kDa (Murakami and Darst, 2003). The catalytically competent core of RNAP is conserved in sequence, structure and function from bacteria to human (Cramer, 2002, 2004; Darst, 2001; Ebright, 2000; Lane and Darst, 2010a, b; Murakami and Darst, 2003; Werner, 2007; Werner and Grohmann, 2011). Archeal RNAP and eukaryotic RNAP contain five conserved subunits (A, B, D, L, and K in archaeal RNAP; RPB1, RPB2, RPB3, RPB11, and RPB6 in eukaryotic RNAP II) and also contain additional subunits (Cramer, 2002; Werner, 2007). Bacterial RNAP is the smallest and best characterized form of multi-subunit RNAP, making it the model system of choice for detailed structural and mechanistic studies.

1.1. Bacterial RNA polymerase core

Multi-subunit RNAP is a complex molecular machinery. Its simplest form, bacterial RNAP core is composed of five subunits: β′, β, α^{I}, α^{II}, ω, with dimensions of ~150 Å x ~ 100 Å x ~100 Å. The first high-resolution structure of bacterial RNAP was solved for *Thermus aquaticus* RNAP core (Zhang et al., 1999). Over a decade, a wealth of structural

information on bacterial RNAP have provided unprecedented insights into the structure and function of this essential enzyme (Darst, 2001; Darst et al., 2002; Murakami and Darst, 2003; Opalka et al., 2010; Vassylyev et al., 2002; Zhang et al., 1999). The structures have revealed that bacterial RNAP core has a shape reminiscent of a crab claw with two prominent pincers (Fig.1). The two pincers define an internal channel of ~ 27 Å in diameter, a size that can accommodate double-stranded DNA. The channel serves as the active-center cleft of the enzyme. The active site is located on the back wall of the channel, where a Mg^{2+} ion required for catalytic activity is chelated by three conserved aspartate residues. The active site of the enzyme is where synthesis of RNA from ribonucleoside triphosphate (NTP) takes place.

The largest subunit of RNAP, β′, forms one of the pincers, termed the "clamp" (Fig.1). The clamp opens and closes at different stages of transcription, allowing DNA to be loaded into the active-center cleft and remain in position during transcription elongation (Chakraborty et al., 2012). The other pincer is formed by the second largest subunit of RNAP, β. β′and β make up the active-center cleft, contain the determinants for the catalysis, and make extensive interactions with other subunits and nucleic acids (Zhang et al., 1999).

In addition to the active-center cleft, RNAP core holds two other distinct channels (Fig.1 B): the secondary channel, which mediates access of NTP substrates to the active center cleft; and the RNA-exit channel, which mediates egress of nascent RNA from the active-center cleft. The RNA-exit channel is covered by the β flap domain (Vassylyev et al., 2002).

Bacterial RNAP core contains two copies of α subunits: α^{I} and α^{II}. They have identical sequences, but make specific interactions with the rest of RNAP: α^{I} interacts with the β subunit and α^{II} interacts with the β' subunit (Zhang et al., 1999). Each α subunit consists of two independently folded domains: the N-terminal domain (αNTD) and the C-terminal domain (αCTD), connected by an ~ 20 amino-acid flexible linker (Ebright and Busby, 1995). αNTD dimerizes and is responsible for the assembly of the β and β′ subunit (Darst et al., 1998). αCTD is a DNA-binding module that has an important role at certain promoters (Blatter et al., 1994).

The smallest subunit of RNAP, ω, is not required for function of RNAP or bacterial growth under normal conditions (Minakhin et al., 2001). However, it has been shown to promote the assembly of the RNA polymerase by "latching" the N- and C-terminal regions of β′.

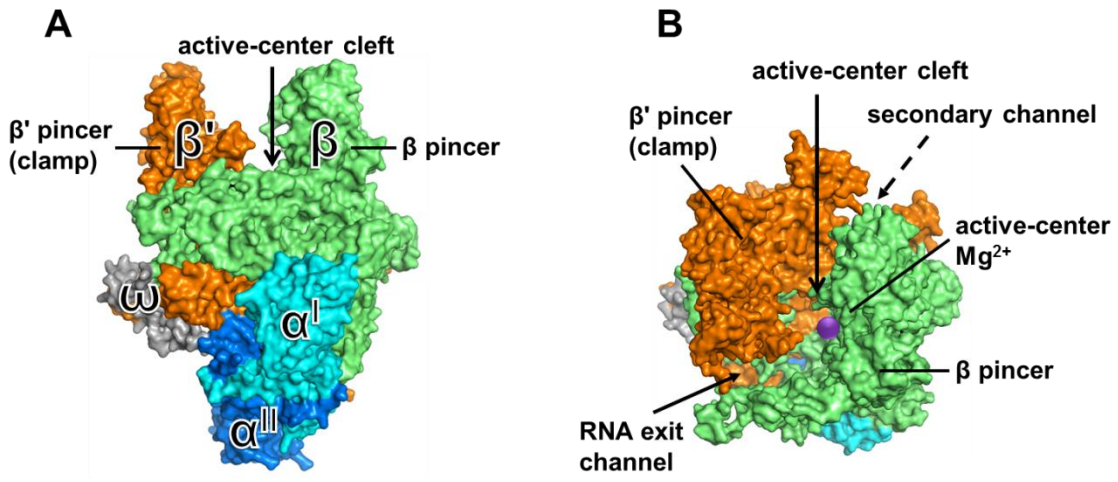


Fig. 1. Structure of RNAP core.
The structure of RNAP core is shown by surface representation (Zhang et al., 1999). β' is in orange; β is in green; α^{I} is in light blue; α^{II} is in dark blue; ω is in gray; the active-center Mg^{2+} is in violet.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A).

1.2. Bacterial RNA polymerase holoenzyme

Bacterial RNAP core is catalytically competent to carry out non-specific transcription initiation and transcription elongation, but is unable, by itself, to carry out promoter-specific transcription initiation (Wheeler et al., 1987). To carry out promoter-specific transcription initiation, RNAP core must associate with an initiation factor σ to form RNAP holoenzyme (subunit composition $\beta'\beta\alpha^{I}\alpha^{II}\omega\sigma$; reviewed in Borukhov and Nudler, 2003; Young et al., 2002). σ contains determinants for sequence-specific interaction with DNA and, through those determinants, targets RNAP holoenzyme to promoters (Busby and Ebright, 1994; Gross et al., 1998; Roberts and Roberts, 1996). In addition, σ plays critical roles in promoter unwinding, promoter escape, early elongation, transcriptional pausing, and response to transcriptional regulators (Campbell et al., 2008; Lee et al., 2012; Perdue and Roberts, 2011).

1.2.1 σ factors

Multiple σ factors are present in most bacterial species. However, typically a single σ is responsible for the majority of transcription in the cell, which is usually referred to the principal or housekeeping σ factor (Gross et al., 1998; Helmann and Chamberlin, 1988). Other σ factors that control specialized functions are considered alternative σ factors. Per bacterial species, there is only one principal σ factor, while the number of alternative σ factors varies. All the principal σ factors and most alternative σ factors belong to the σ^{70} family. Members in σ^{70} family share sequence and structure similarity with *Escherichia coli* principal σ factor, σ^{70} (Paget and Helmann, 2003). Many bacteria also have a second

distinct type of σ called the σ^{54} family, which share no sequence homology with the σ^{70} family and also utilize a distinct pathway of transcription initiation (Buck et al., 2000).

The most significant function of σ factors in transcription is promoter recognition. Promoters are specific DNA sequences that mark the transcriptional start site. Bacterial promoters contain two core sequence elements: the -10 element and the -35 element (Harley and Reynolds, 1987; Hawley and McClure, 1983; Keilty and Rosenberg, 1987; Siebenlist et al., 1980). The -10 element is located near position -10 relative to the transcription start site with the consensus sequence 5'-TATAAT-3'. The -35 element is located near position -35 relative to the transcription start site and has the consensus sequence 5'-TTGACA-3'. Strength of promoters is generally determined by the similarity of -10 and -35 regions to the consensus sequence. Some promoters have an "extended -10 element", located immediately upstream of the -10 element of the promoter with the consensus sequence 5'-TGN-3' (Keilty and Rosenberg, 1987; Mitchell et al., 2003). Such promoters can function well without a recognizable -35 region.

The principal σ factors of bacterial species--σ^{A} in *T. thermophilus* and *T. aquaticus*, σ^{70} in *E. coli*--contain five conserved regions (Fig.2): σ-region-1.1 (σR1.1), σ-region-2 (σR2), σ-region-3 (σR3), σ-region-3.2, and σ-region-4 (σR4) (Gross et al., 1998; Lonetto et al., 1992). σR1.1 and σR3.2 (also known as the σR3/ σR4 linker) are flexible, highly negatively charged segments. σR2, σR3, and σR4 are stably folded domains that contain determinants for sequence-specific interactions with, respectively, the promoter -10 element, the promoter extended -10 element, and the promoter -35 element (Campbell et al., 2002; Gross et al., 1998; Malhotra et al., 1996; Murakami and Darst, 2003; Young et

al., 2002).The σ non-conserved region (σNCR) is present in some principal σ factors, not related in size, sequence or structure (Campbell et al., 2002; Gruber and Gross, 2003).

Crystal structures of proteolytic fragments of σ suggested that σ factors are composed of domains connected by flexible linkers (Fig. 2, Campbell et al., 2002; Malhotra et al., 1996; Severinova et al., 1996). Each domain includes both an RNAP binding determinant and a DNA binding determinant. σR2 consists of three helices, with one of the helices forming the primary interface with RNAP core and another helix involved in DNA melting and recognition of -10 promoter element. σR2 and σR3 are connected by an exposed, flexible, but highly conserved loop. σR3 is a compact domain of three helices, one of which interacts with the extended -10 region of the promoter. σR4, composed of four helices, contains a major RNAP binding determinant and interacts with -35 promoter element. The flexible loop σR3.2 between σR3 and σR4, as well as the N-terminal region σR1.1, were degraded by proteolysis and not crystalized.

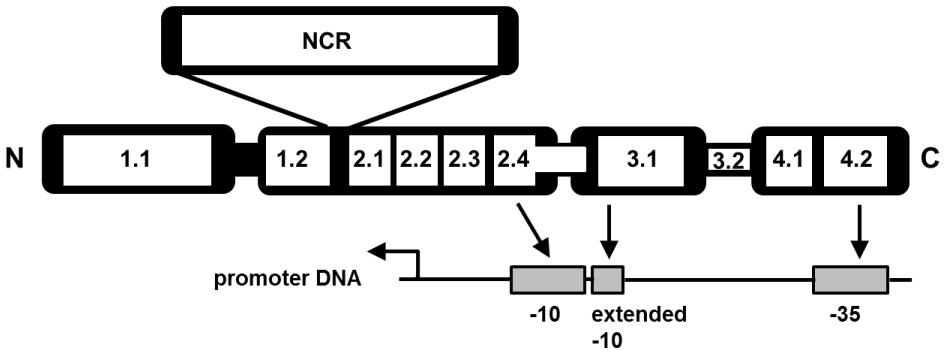


Fig. 2. **Sequence architecture of *E.coli* σ^{70}**.
Regions of σ^{70} are labeled and their specific interactions with promoter DNA are shown by arrows (adapted from Murakami and Darst, 2003).

1.2.2. Structure of RNA polymerase holoenzyme

The structures of *T. thermophilus* RNAP holoenzyme and *T. aquaticus* RNAP holoenzyme have been determined at 2.6 Å and 4 Å resolutions (Murakami et al., 2002b; Vassylyev et al., 2002). The structures reveal that the organization of core subunits is essentially identical in RNAP core and in RNAP holoenzyme, except that the β' pincer rotates, as a unit, ~16 °into the active-center cleft in RNAP holoenzyme.

In the holoenzyme, the globular domains of σ are spread out across the upstream face of the RNAP crab claw (Fig.3). The interactions between σ and core subunits are extensive--σR2 interacts with the β' pincer, in and above the RNAP active-center cleft; σR3 interacts with the base of the β flap; and σR4 interacts with the tip of the β flap. The linkers connecting each domain also make contacts with core subunits. σR3.2 is located in the RNA-exit channel, and must be displaced to permit access of nascent RNA to the RNA exit channel (when nascent RNA reaches a length of ~9-11 nt). σR1.1 is missing from the crystal structure of RNAP holoenzyme.

Although σ factors contain the determinants for interactions with promoter DNA, they are unable to bind with promoter DNA in the absence of RNAP core. Association with RNAP core induces the conformation change of σ and unmasks the DNA-binding elements in σ (Callaci et al., 1998, 1999; Callaci and Heyduk, 1998; Dombroski et al., 1993; Dombroski et al., 1992). In the holoenzyme, domains of σ are properly positioned to interact with promoter DNA (Fig.3). All the promoter-recognition determinants in σ are solvent exposed in the holoenzyme structure, with a spacing roughly consistent with

the predicted separation of their target promoter elements (Murakami, 2013; Murakami and Darst, 2003; Murakami et al., 2002a; Murakami et al., 2002b; Vassylyev et al., 2002).

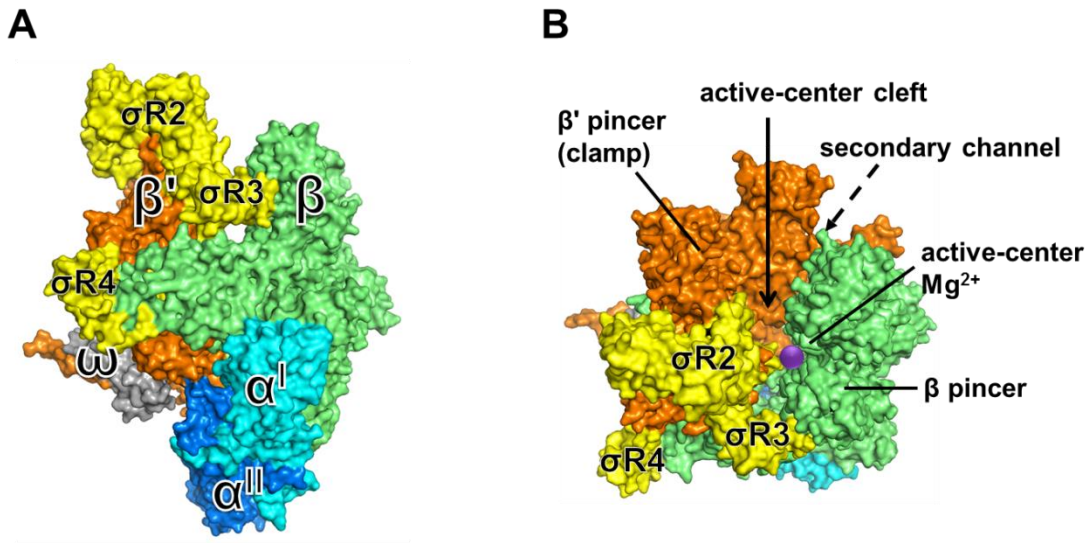


Fig. 3. Structure of RNAP holoenzyme.
The structure of RNAP holoenzyme is shown by surface representation (Vassylyev et al., 2002). β' is in orange; β is in green; α^{I} is in light blue; α^{II} is in dark blue; ω is in gray; σ^{70} is in yellow.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A). σR3.2 passes through the RNA-exit channel (not shown in this figure).

1.3. Transcription cycle

The transcription cycle can be divided into three main steps: initiation, elongation and termination. Upon association with initiation factor σ, RNAP carries out a series of reactions during transcription (Bai et al., 2006; Nudler, 2009; Peters et al., 2011; Saecker et al., 2011; Vassylyev, 2009):

(i) RNAP binds to promoter DNA, yielding an RNAP-promoter closed complex (RPc).

(ii) RNAP unwinds ~13 base pairs of promoter DNA surrounding the transcription start site, forming a single-stranded region ("transcription bubble"), and yielding an RNAP-promoter open complex (RPo).

(iii) RNAP begins synthesis of an RNA product as an RNAP-promoter initial transcribing complex (RPitc). During initial transcription, RNAP uses a "scrunching" mechanism, in which RNAP remains stationary on promoter DNA and unwinds and reels in downstream DNA in each nucleotide-addition cycle (Kapanidis et al., 2006; Revyakin et al., 2006).

(iv) After RNAP synthesizes an RNA product ~11 nt in length, RNAP breaks its interactions with the promoter, escapes from the promoter, and begins transcription elongation as an RNAP-DNA elongation complex (RDe). During transcription elongation, RNAP uses a "stepping" mechanism, in which RNAP translocates relative to DNA in each nucleotide-addition cycle (Abbondanzieri et al., 2005).

(v) When RNAP encounters a termination signal, RNAP releases the RNA product, and dissociates from DNA.

1.4. RNA polymerase-promoter open complex

Initiation of RNA synthesis from DNA template is a rate-limiting step during the transcription cycle, and is the step under most control (Browning and Busby, 2004). During transcription initiation, RNAP core associates with the initiation factor σ to form RNAP holoenzyme, and binds to promoter DNA to form an RNAP-promoter closed complex (RPc). RPc undergoes rapid isomerizations, melts ~ 13 nt surrounding the transcription start site, forms a transcription bubble, rendering accessible the genetic information in the template strand of DNA, to yield a stable RNAP-promoter open complex (RPo). RPo is the critical, catalytically competent, intermediate in transcription initiation, and modulation of the formation, stability, and activity of RPo is an important means of regulation of gene expression (Saecker et al., 2011).

The crystal structure of RPo has been determined at 2.9 Å resolution from the complex of *T. thermophiles* holoenzyme and a synthetic DNA scaffold (Zhang et al., 2012). The organization of RNAP subunits in RPo is similar to RNAP holoenzyme, except that the clamp closes by ~ 11 °relative to the crystal structure of holoenzyme (Vassylyev et al., 2002). The clamp conformation in RPo is the same as in the crystal structure of RNAP elongation complex, and is consistent with FRET results, indicating that the clamp closes upon formation of RPo and remains closed during elongation (Chakraborty et al., 2012; Vassylyev et al., 2007). The structure reveals that RNAP core and initiation factor σ make sequence-specific interactions with the nontemplate strand in the transcription bubble, and preorganize the downstream DNA to the same form in transcription elongation.

1.5. σR1.1

σR1.1 is present and conserved among primary σ factors. It is a highly acidic, flexible segment with 90 ~ 100 amino acids at the N-terminus of σ. Important biological functions have been ascribed to σR1.1. First, it interacts with the DNA-binding determinants on free σ, masking the DNA-determinants on σ, therefore preventing free σ from association with promoter DNA (Camarero et al., 2002; Dombroski et al., 1993; Dombroski et al., 1992). σ constructs lacking region 1.1 were able to bind promoter DNA whereas full-length σ exhibited only very weak DNA binding. In RNAP holoenzyme, σR1.1 no longer masks the DNA-binding determinants on σ, therefore enabling association with promoter DNA. Second, *E. coli* RNAP holoenzyme reconstituted with σ^{70} that lacks region 1.1 is easier to dissociate, compared with polymerase with full-length σ^{70}, suggesting that σR1.1 plays a role in stabilizing the interactions between σ^{70} and core (Hinton et al., 2006). Third, σR1.1 modulates the formation of RPo at certain promoters. Depending on the sequences of the promoters, the presence of σR1.1 can either increase or decrease the rates of formation of RPo (Vuthoori et al., 2001; Wilson and Dombroski, 1997). Last, σR1.1 is the target for bacterial phage T7 Gp2, a transcription regulator that inhibits *E. coli* RNAP. Full inhibition of formation of RPo by Gp2 requires the presence of σR1.1 (James et al., 2012; Sheppard et al., 2011).

The solution structure of the isolated domain of *Thermotoga maritima* σ^{A} σR1.1 (residues 29-95) has been determined by NMR (Schwartz et al., 2008). The structure shows that σR1.1has a core fold domain comprising of three α helices (Fig.4 B). Two short helices are roughly anti-parallel to one another and pack perpendicularly against the longest helix. The C-terminal tail is a long and flexible loop. σR1.1 has a mostly negative electrostatic

surface potential and contains a compact hydrophobic core formed by highly conserved residues from all three helices and part of the C-terminal tail. Due to the homology between σR1.1 of *T. maritima* σ^A and *E.coli* σ^{70}, the solution structure of *T. maritima* σ^A σR1.1 can serve as a structural model for σR1.1 of *E.coli* σ^{70} (Fig.4 A).

σR1.1 is missing from the crystal structures of RNAP holoenzyme and RPo described (1.2, 1.4). The conformational change of σR1.1 in free σ^{70}, holoenzyme and RPo is first revealed by footprinting experiments. Hydroxyl-radical protein footprinting shows that *E.coli* σ^{70} region 1.1 is exposed to hydroxyl radicals in free σ^{70}, substantially protected in the holoenzyme but exposed again in the binary complex of holoenzyme with promoter DNA (Nagai and Shimamoto, 1997). Systematic ensemble fluorescence resonance energy transfer (FRET) and distance-restrained docking analyzed the positions of *E.coli* σ^{70} domains in RNAP holoenzyme and in RPo (Mekler et al., 2002). The results suggested in RNAP holoenzyme, σR1.1 is located inside the RNAP active-center cleft, just above the floor of the downstream-duplex channel; in RPo, σR1.1 is located outside the RNAP active-center cleft—50 Å away from its position in RNAP holoenzyme—is positioned near the β pincer tip. It proposed that the highly negatively charged σR1.1 serve as a "molecular-mimic" of DNA by occupying the RNAP active-center cleft in RNAP holoenzyme, and must be displaced out of the RNAP active-center cleft upon formation of RPo.

The precise positions and rotational orientations of σR1.1 in RNAP holoenzyme and RPo have remained uncertain. Ensemble FRET analysis provided an estimation of positions of σR1.1 in RNAP holoenzyme and RPo, but no information on rotational orientations, mainly due to limited labeling sites (2 reporter probe sites on σR1.1, 4 reference probe

sites on RNAP core (Mekler et al., 2002)). Therefore, more precise and detailed mapping of σR1.1 in holoenzyme and RPo is required to understand the mechanisms of σR1.1 displacement and to define the interactions of σR1.1 with RNAP core subunits. Such structural models of σR1.1 in holoenzyme and RPo should also provide the necessary framework for analyzing the ever accumulating genetic, biochemical, and biophysical data.

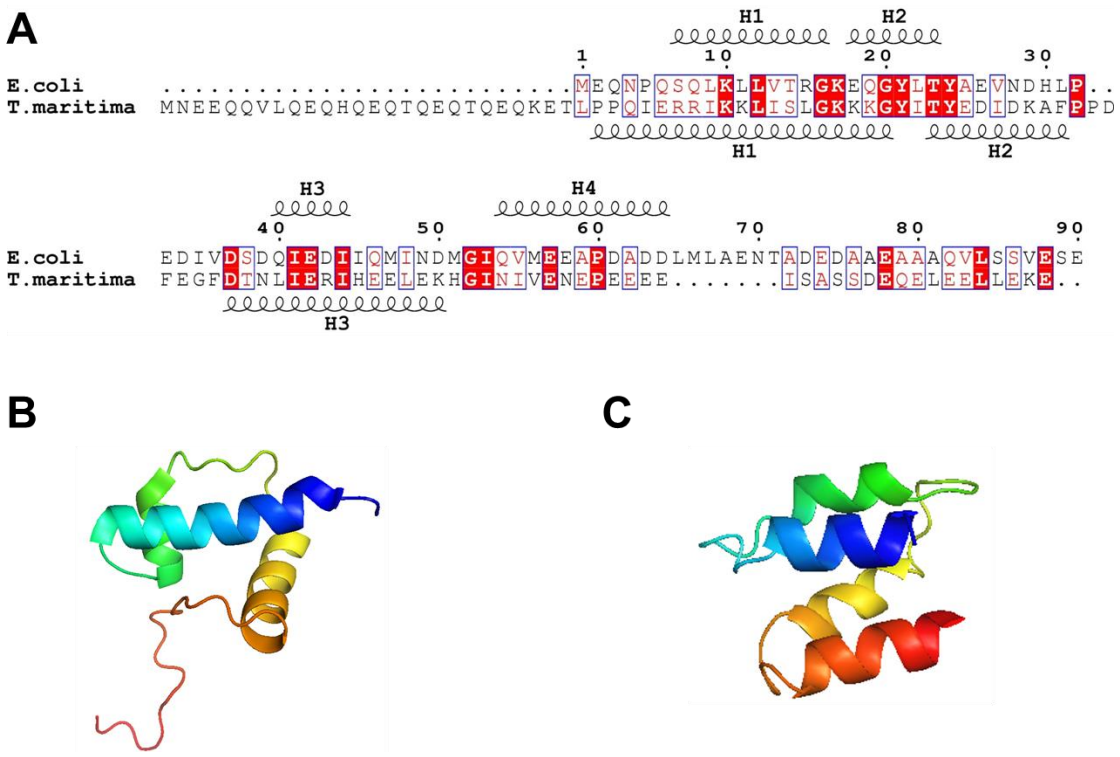


Fig. 4. Structures of σR1.1.
(A) Sequence alignment of *E. coli* σ^{70} σR1.1 and *T. maritima* σ^{A} σR1.1 using ClustalW2 program. Highlighted in red background are identical residues, in red font and blue square are conserved residues. Compositions of α-helices in σR1.1 are shown by springs above (*E. coli*) and below (*T. maritima*) the sequence alignment.
(B) Cartoon representation of the solution structure of *T. maritima* σ^{A} σR1.1 (Schwartz et al., 2008).
(C) Cartoon representation of the crystal structure of *E. coli* σ^{70} σR1.1 (Murakami 2013). The structures are rainbow spectrum colored from N-terminus (blue) to C terminus (red).

1.6. σR1.1: new structural information

Recently, crystal structures of σR1.1 in RNAP holoenzyme have become available from two different groups—Murakami and co-workers and Darst and co-workers (Murakami, 2013; Bae et al., 2013, in press). They both place σR1.1 inside the active-center cleft in RNAP holoenzyme, consistent with ensemble FRET analysis, but the folds and the rotational orientations of σR1.1 in the crystal structures from the two groups are different.

In the crystal structure of *E. coli* RNAP-σ^{70} holoenzyme from Murakami and co-workers (Murakami, 2013), σR1.1 is located inside the active-center cleft, surrounded by σR2, β lobe, and β′ clamp, a location consistent ensemble FRET analysis (Mekler et al., 2002). The fold of σR1.1 is different from that of the NMR structure of *T. maritima* σ^{A} σR1.1 (Fig.4 C). In Murakami's crystal structure, σR1. 1 (residues 6-64) possesses a "double-U" fold comprising four α-helices, distinct from the three-helix fold in the NMR structure (Schwartz et al., 2008).

In the crystal structure of *E. coli* RNAP-σ^{70} holoenzyme from Darst and co-workers (Bae et al., 2013, in press), σR1.1 comprises a core folded domain (residues 1-56) of three α-helices with similar topology to the NMR structure of σR1.1 (Schwartz et al., 2008). The globular domain of σR1.1 sits directly in the path of the downstream duplex DNA inside the RNAP active-center cleft, very close to the position proposed from ensemble FRET analysis (Mekler et al., 2002). The rotational orientation of σR1.1 relative to RNAP core is different from that in Murakami's crystal structure.

The crystal structures of σR1.1 in holoenzyme from the two groups provided detailed information on the positions and rotational orientations of σR1.1 in holoenzyme.

However, the results from the two groups conflicted with each other. Therefore, additional information from other sources of measurements is needed to differentiate the conflicting crystal structures of σR1.1.

1.7. Single-molecule fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET), a physical phenomenon that permits measurement of distances, has been applied to define the structure and mechanism in transcription (Mekler et al., 2002; Mukhopadhyay et al., 2001; Mukhopadhyay et al., 2003). Regarded as a "spectroscopic ruler", FRET allows accurate determination of distances in the range of ~20 Å to ~ 100 Å, about one-half the diameter of a transcription complex (Lilley and Wilson, 2000; Murakami and Darst, 2003; Selvin, 2000; Stryer and Haugland, 1967). FRET occurs in a system having a fluorescent probe serving as a donor and a second fluorescent probe serving as an acceptor, where the emission spectrum of the donor overlaps the excitation spectrum of the acceptor. In such a system, upon excitation of the donor at its excitation wavelength, energy can be transferred from the donor to the acceptor, resulting in excitation of the acceptor and emission at the acceptor's emission wavelength. 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 relation (Clegg, 1992):

$$E = 1/[1 + (R/R_0)^6] \quad (1)$$

Where R_0 is the characteristic Förster parameter for a given donor-acceptor pair, the value of which depends on the spectral properties of the fluorescent probes and the relative orientation of their dipole moments:

$$R_0 = 9780(n^{-4}\kappa^2 Q_D J)^{1/6} \text{ Å}$$

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. When $R = R_0$, the energy transfer is 50% efficient.

FRET efficiency E can be measured by a variety of ways, including a reduction in the fluorescent quantum yield of the donor, a corresponding shortening of the donor excited state lifetime, and an increased fluorescent emission from the acceptor (Clegg, 1992) . R_0 can be determined by experiments independent of energy transfer for each donor-acceptor pair. Thus, the donor-acceptor distance R can be derived from equation (1) with the values of E and R_o.

FRET analysis is conventionally carried out at ensemble level, which provides averaged, mean values for the system studied. In contrast to the ensemble FRET analysis, single-molecule FRET analysis allows measurements on a true molecular basis, thus reveals information on heterogeneous populations and dynamic processes (Ha et al., 1996). Single fluorescently labeled molecules can be studied when they are freely diffusing in solution or when they are surface immobilized (Deniz et al., 1999; Ha, 2001). Surface immobilized single-molecule analysis is especially useful in detecting time-dependent conformational change or dynamic events, by observing a particular molecule over a period of time. However, special care has to be taken to ensure minimal perturbation on the system (photophysical properties of the dye, structure and function of biological complexes) from immobilization.

In-solution single-molecule FRET (smFRET) measurements typically use set-ups where a femtoliter scale observation volume is defined by focused laser beams and confocal optics (Deniz et al., 1999; Schuler and Eaton, 2008). Molecules labeled with donor and acceptor fluorophores are present at sub-nanomolar concentration, freely diffusing in solution, by virtue of Brownian motion. As a single molecule transit through the observation volume, it is exposed to the light that excites the donor, generating fluorescence photon bursts corresponding to the excitation. When an acceptor is close to a donor, energy is partially transferred from donor to acceptor, which then emits at wavelength longer than donor. The photons from donor and acceptor are counted in their respective donor and acceptor detector channels. The FRET efficiency for each donor-acceptor pair is reported as a ratiometric parameter, determined by the ratio of photons detected in the acceptor channel and the number of photons detected in both donor and acceptor channels. Data from a large number of events (photon bursts) is summarized in one-dimensional FRET histograms with statistical analysis, which report on the presence of static and dynamic heterogeneity, on the presence of conformational changes and on molecular interactions.

The use of alternating-laser excitation (ALEX) has recently extended the applications of smFRET (Kapanidis et al., 2005; Kapanidis et al., 2004; Lee et al., 2005) . ALEX confocal microscopy employs a second laser that excites the acceptor at acceptor excitation wavelength, in addition to the laser that excites the donor at donor excitation wavelength. The two lasers are aligned, alternated at sub-millisecond time scale, and focused tightly into a femtoliter-scale observation volume (Fig.5 A). During the transit time of a freely diffusing fluorescent molecule through the observation volume (~ 1

millisecond), a finite number of excitation/emission cycles occur, resulting in a burst of fluorescence photons. When analyzing the stream of photons, two independent fluorescence ratios can be calculated: proximity ratio *E* (energy transfer efficiency) and stoichiometry *S*. The stoichiometry parameter *S* reports on the relative stoichiometry of donor and acceptor probes for each molecule observed. This allows the construction of a two-dimensional FRET histograms, an E/S plot, in which the *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). Thus, the desired doubly labeled species can be selected from donor-only and acceptor-only species, allowing accurate measurement at long donor-acceptor distances (6-10 nm). The corresponding *E* distributions can be plotted and fitted with Gaussians, in which the number of Gaussians correlates with the number of subpopulations and the mean of the Gaussian defines the mean *E* value of the subpopulation.

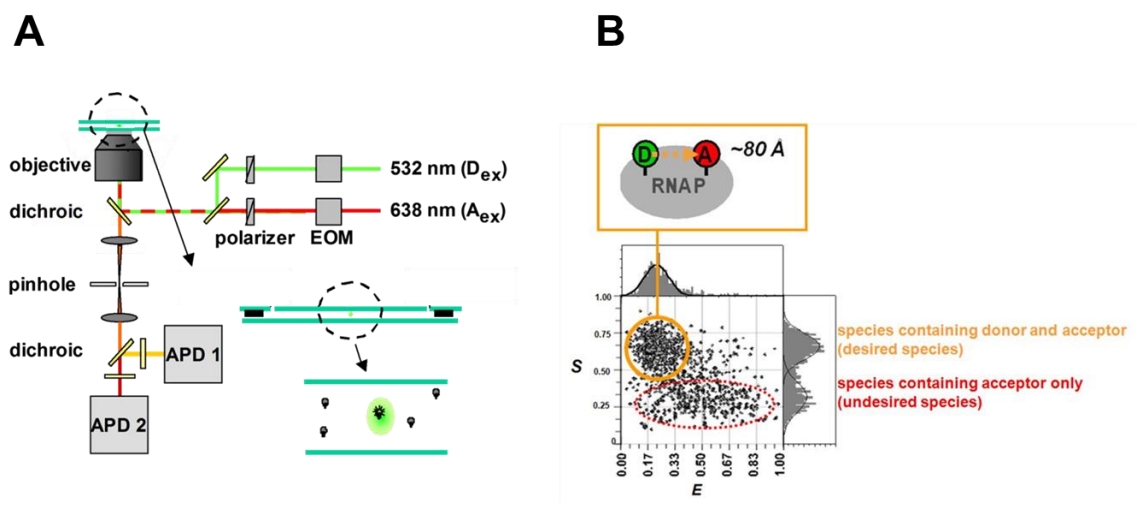


Fig. 5. Measurement of smFRET: two-color ALEX confocal microscopy.
(A) A diagram of the experimental set-up. In actually experiments, excitation of the two lasers at 532 nm and 638 nm wavelengths is modulated using AOM (acousto-optical modulator), resulting in alternation at microsecond time-scale. Excitation laser beams are coupled through optical fiber, directed to a dichroic mirror that reflects into the objective. Excitation light is tightly focused to a femtoliter-scale observation volume in a sample chamber (inset at right center). As a single fluorescent molecule diffuses through the observation volume (inset at lower right), it is excited by the laser beams and the generated fluorescence emission travels down though the objective. The pinhole is used to reject the out-of-focus light, so as to isolate femtoliter observation volume. The fluorescence emission signals from donor and acceptor are further directed into respective donor and acceptor detection channels, counted by APDs (avalanche photodiodes).
(B) Sorting of single molecules using two-dimensional E/S plot. Each dot in the plot represents a single observation for a single molecule during transit through the observation volume. For each single observation, two ratiometric parameters are calculated-a donor-acceptor stoichiometry parameter (*S*) and a donor-acceptor smFRET efficiency (*E*), resulting in distributions of *S* and *E* over a number of observations. The distribution of *S* values 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). Considering only observations from the doubly labeled desired species, the distribution of *E* values on the E-axis (histogram at top) defines mean *E* and permits calculation of mean donor-acceptor distance, *R* (image at top). In the case of two or more subpopulations are present, the number of peaks in the distribution defines the number of distinguishable subpopulations. For each distinguishable subpopulation, mean *E* and mean *R* can be defined. (Adapted from Chakraborty et al., 2012)

2. Experimental strategy

The objectives of this work are to define the positions and the rotational orientations of σR1.1 in RNAP holoenzyme in solution and in RPo in solution, therefore to differentiate the crystal structures of σR1.1 from two different groups, and to provide first information identifying regions of σR1.1 and regions of RNAP core that interacts in RPo.

The method of choice is systematic single-molecule fluorescence resonance energy transfer (smFRET). The experimental strategy involves three main parts:

(1) generation of a series of RNAP derivatives, with a fluorescent probe incorporated in σR1.1, and a complementary fluorescent probe incorporated in RNAP core (Strategy I) or σ^{70} (Strategy II);

(2) measurement of probe-probe smFRET distances in the context of holoenzyme and RPo using fluorescently labeled RNAP derivatives;

(3) automated, computer-aided distance-restrained docking of σR1.1 to the structures of holoenzyme and RPo using FRET distance restraints.

2.1. Generation of fluorescently labeled RNAP derivatives

Depending on the choice of reference probe sites, two sets of experimental strategies were performed.

Experimental strategy I involves:

(1) Incorporation of the fluorescent probe tetramethylrhodamine (TMR) at each of 4 reporter probe sites within σR1.1;

(2) Incorporation of the fluorescent probe Alexa Fluor 647 (A647) at each of 6 reference probe sites within RNAP core;

(3) Preparation of RNAP holoenzyme derivatives from TMR-labeled σ^{70} and A647-labeled core.

In step 1 of the procedure, I incorporated TMR at each of four sites within σR1.1: namely, residue 14, 36, 46 and 59 of σ (Fig.6) Two of the probe sites, residue 14 and 59 were previously used in ensemble FRET analysis (Mekler et al., 2002; Mukhopadhyay et al., 2003). Probe sites 36 and 46 are located on the helix both in NMR structure and crystal structure (Murakami, 2013; Schwartz et al., 2008). Four probe sites are well separated, allowing accurate determination of the position and orientation of σR1.1. I prepared four σ-TMR derivatives with TMR labeled at each site. To prepare each σ derivative, I first prepared a σ derivative containing a single Cys residue at the site of interest, and then performed Cys-specific chemical modification to introduce TMR at the Cys residue.

In step 2 of the procedure, I incorporated A647 at each of six sites within core: namely, residue 106, 222, 357, 379, 643 and 937 of β (Fig.7). Two of the probe sites, residue 643 and 937 were previously used in ensemble FRET analysis (Mekler et al., 2002). One of the probe sites, residue 106 was previously used in single-molecule FRET analysis (Chakraborty et al., 2012). Six probe sites are well separated in the structure of core, are located at the periphery of core, and are within the distance range where FRET is very sensitive to any changes in distance (~20-100 Å). The probe sites are surface-exposed, non-conserved residues, and are not involved in structural stability and function of RNAP (confirmed by transcription assays, see Results). I prepared six core-A647 derivatives with A647 labeled at each site. To prepare each core derivative, I used a procedure

comprising unnatural amino acid mutagenesis (Chin et al., 2002; Wang et al., 2001), Staudinger ligation (Chakraborty et al., 2010; Saxon and Bertozzi, 2000) and RNAP reconstitution (Tang et al., 1995). The procedure (Fig.8) involved (i) preparation of β subunits containing 4-azidophenylalanine at the sites of interest (accomplished by expressing engineered genes containing nonsense codons at sites of interest using cells that contained an engineered suppressor-transfer RNA (tRNA)/aminoacyltRNA-synthase pair in media supplemented with 4-azidophenylalanine); (ii) incorporation of the fluorescent probe A647 into β subunits at the sites of interest by azide-specific chemical modification (accomplished by Staudinger ligation using A647-phosphine derivatives); and (iii) *in vitro* reconstitution of RNAP core (β-A647/β′/α^{I}/α^{II}/ω).

In the step 3 of the procedure, I prepared RNAP holoenzyme derivatives from each of the σ-TMR derivatives and each of the core-A647 derivatives, yielding 24 distinct RNAP holoenzyme derivatives. Transcriptional activities of RNAP derivatives are comparable to wild-type RNAP, suitable for subsequent smFRET analysis.

Experimental strategy II involves:

(1) Incorporation of the fluorescent probe A647 at each of 4 reporter probe sites within σR1.1;

(2) Incorporation of the fluorescent probe Cy3B at each of 6 reference probe sites within σR2, σR3, and σR4;

(3) Preparation of RNAP holoenzyme derivatives from A647/Cy3B-labeled σ^{70} and unlabeled wild-type RNAP core.

In step 1 and step 2 of the analysis, I incorporated the fluorescent probes A647 and Cy3B at each of a series of sites on σ, yielding doubly labeled σ derivatives. Each σ derivative

contained A647 at one of the four labeling sites on σR1.1, and Cy3B at one of the six reference probe sites on other regions of σ. The labeling sites on σR1.1 are residue 14, 36, 46 and 59 as in experimental strategy I. Incorporation of the fluorescent probe A647 into σR1.1 involved the procedure comprising unnatural amino acid mutagenesis and Staudinger ligation. The six reference probe sites on σR2, σR3 and σR4 are residue 108, 304, 366, 396, 459 and 569. Four of the probe sites, residue 366, 396, 459, and 569 were previously used in ensemble FRET analysis and single molecule FRET analysis (Kapanidis et al., 2006; Mukhopadhyay et al., 2003). Six probe sites are well separated in the structure of σ, are located at upper half of RNAP, and are located opposite to the positions of probe sites on core (Fig.9). The probe sites are surface-exposed, non-conserved residues, and are not involved in interactions with promoter DNA or RNAP core (confirmed by transcription assays, see Results). Incorporation of the fluorescent probe Cy3B into σ-A647 derivatives containing a single Cys residue at the site of interest was accomplished by Cys-specific chemical modification.

In step 3 of the analysis, I prepared RNAP holoenzyme derivatives from each of the σ-Cy3B/A647 derivatives and wild-type RNAP core, yielding 24 distinct RNAP holoenzyme derivatives. Transcriptional activities of RNAP derivatives are comparable to wild-type RNAP, suitable for subsequent smFRET analysis.

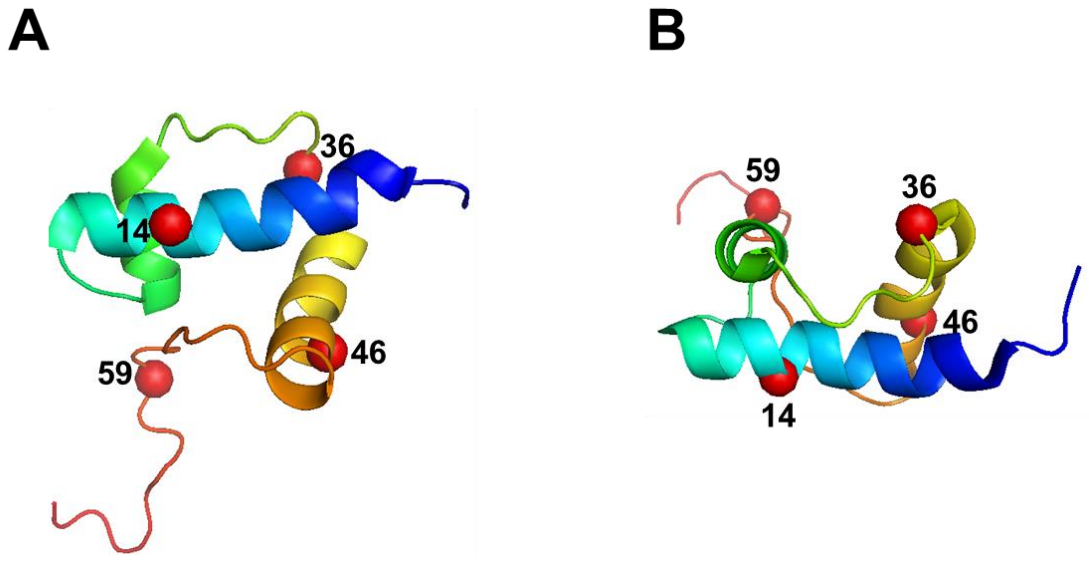


Fig. 6. Reporter probe sites on σR1.1.
The solution structure of *T. maritima* σ^A σR1.1 is represented by cartoon. Red spheres represent the labeling sites on σR1.1. Residues are numbered as in *E. coli* σ^{70}. The structure is rainbow spectrum colored from N-terminus (blue) to C-terminus (red).
(A) A view corresponding to Schwartz et al., 2008.
(B) A view with -90 °rotation about x-axis relative to (A).

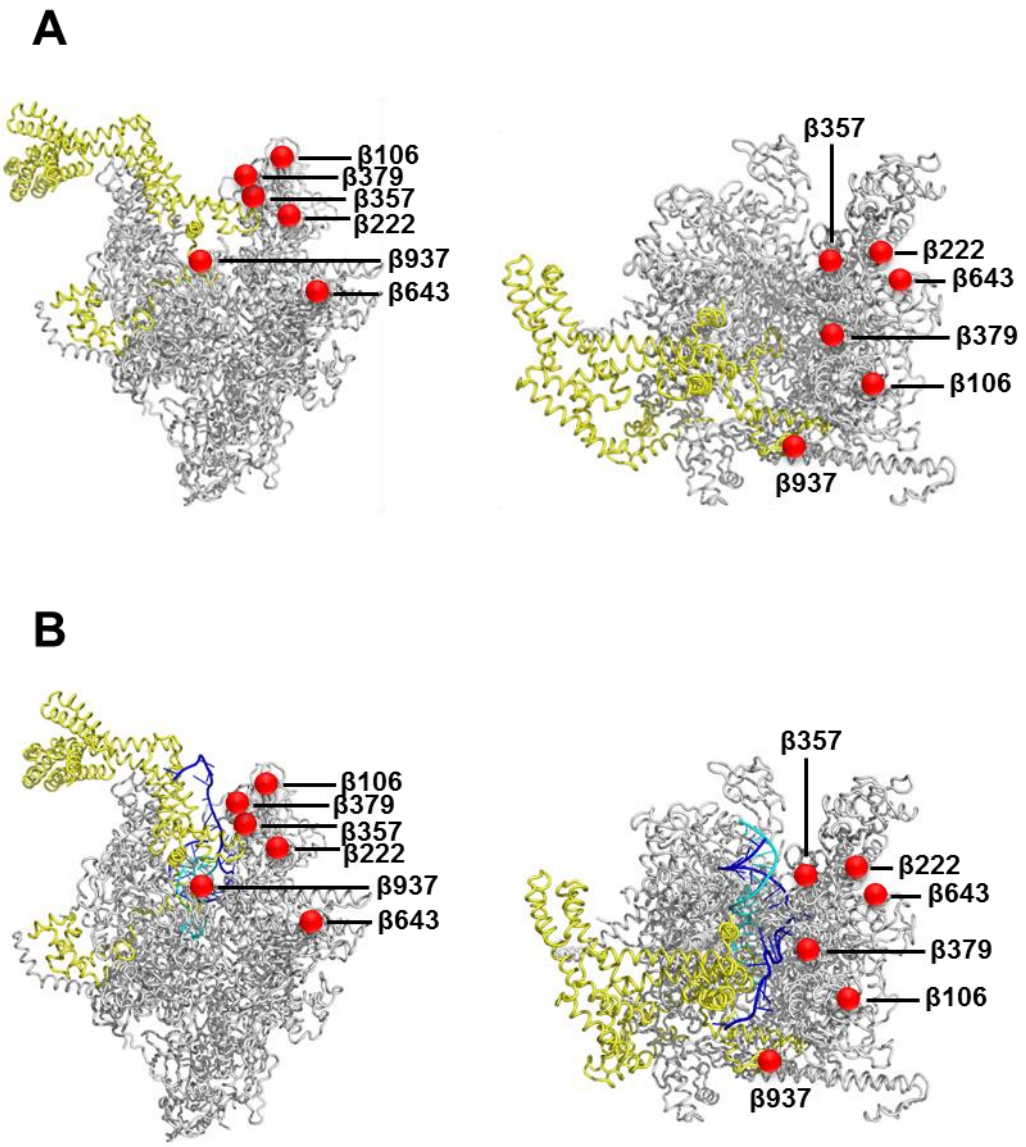


Fig. 7. Reference probe sites on RNAP core (Experimental strategy I).
Red spheres represent the labeling sites on RNAP core. RNAP core is in gray; σ^{70} is in yellow; DNA template strand is in cyan; DNA non-template strand is in blue. Left, upstream face; Right, top face (view into active-center cleft, -90 °rotation about x-axis relative to left).
(A) Structure of *E. coli* RNAP holoenzyme as in Murakami 2013.
(B) Structural model of RPo with DNA incorporated from the crystal structure of *T. thermophilus* RPo (Zhang et al., 2012)

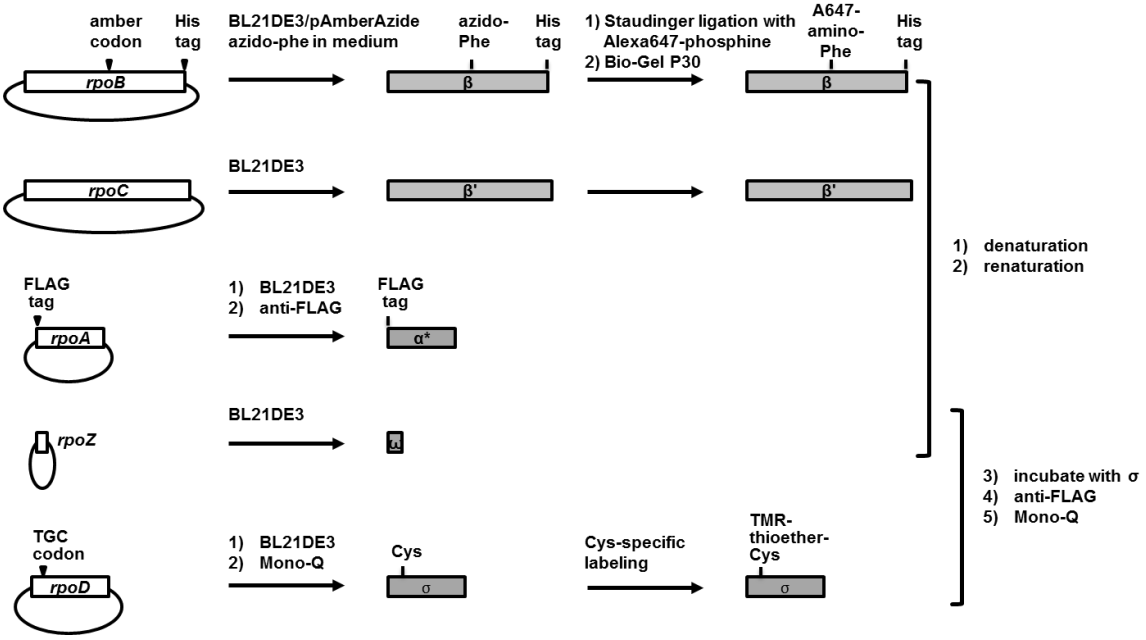


Fig. 8. **Preparation of RNAP holoenzyme derivative with a donor fluorophore (TMR) on σR1.1 and an acceptor fluorophore (Alexa647) on β subunit.**
Preparation of β derivative containing 4-azidophenylalanine was accomplished by expressing engineered genes containing nonsense codon at the site of interest along with an engineered suppressor-transfer RNA (tRNA)/aminoacyl tRNA-synthase pair in media supplemented with 4-azidophenylalanine. Alexa647 was incorporated into β derivative by Staudinger ligation using Alexa647 phosphine derivatives. RNAP core derivative was reconstituted *in vitro* from Alexa647-labeled β derivative, wild-type β′ and ω, and α*(FLAG-αNTD^{I}-GSGGSG-αNTD^{II}). σ bearing single Cys residue at σR1.1 was produced and labeled with TMR by Cys-specific modification. 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.

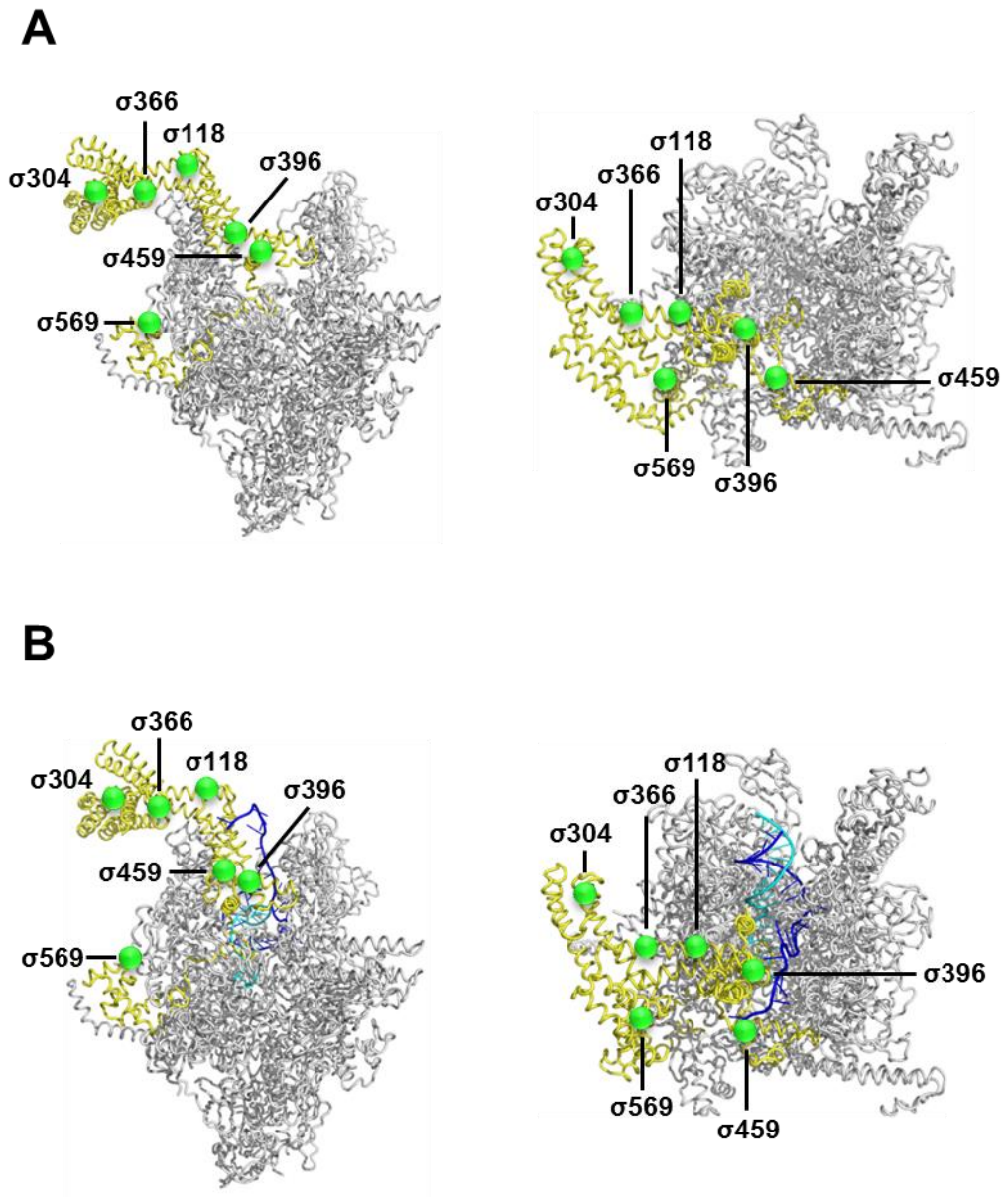


Fig. 9. Reference probe sites on σR2, σR3, and σR4 (Experimental strategy II). Green spheres represent the labeling sites on σ. RNAP core is in gray; σ^{70} is in yellow; DNA template strand is in cyan; DNA non-template strand is in blue. Left, upstream face; Right, top face (view into active-center cleft, -90 °rotation about x-axis relative to left).
(A) Structure of *E. coli* RNAP holoenzyme as in Murakami 2013.
(B) Structural model of RPo with DNA incorporated from the crystal structure of *T. thermophilus* RPo (Zhang et al., 2012).

2.2. Data collection: smFRET efficiencies and distances

In-solution smFRET data acquisition was performed using ALEX confocal microscopy (Fig.5 A), for each of RNAP holoenzyme derivatives prepared in experimental strategy I and experimental strategy II, and, both in the context of holoenzyme and in the context of RPo.

For each measurement, data analysis of the photon bursts generated a 2-dimensional *E*/*S* plot, with which the doubly labeled species were selected for plotting *E* distributions. The equilibrium distribution of *E* values was fitted with Gaussians, allowing distinction of subpopulations, and for each subpopulation, the mean *E* values were defined. For samples that exhibit two Gaussian distributions, the major subpopulation was identified (if the minor subpopulation is less than one-third of the major one) and the *E* value of the major subpopulation was used for model construction.

Förster parameter R_0 was determined for each RNAP derivatives, both in the context of holoenzyme and in the context of RPo. Therefore, the mean probe-probe distances for each RNAP derivatives can be derived from Equation (1) using values of *E* and R_0. In total, 47 probe-probe distances for holoenzyme, and 48 probe-probe distances for RPo were employed in the subsequent computational modeling.

2.3. Distance-restrained docking

In collaboration with Dr. Jennifer Knight (*Schrödinger*), an automated, objective distance-restrained docking algorithm was performed to dock σR1.1 onto the structures of RNAP holoenzyme and RPo using experimental FRET distance restraints (Materials and Methods, distance-restrained docking). The starting models of RNAP holoenzyme and RPo were constructed using current crystal structural information on *E.coli* RNAP holoenzyme and *T.thermophilus* initial transcribing complex (Murakami, 2013; Zhang et al., 2012). σR1.1 was modeled using residues 35-87 of the solution structure of *T.maritima* σ^A σR1.1. In both holoenzyme and RPo, 100,000 trial configurations of σR1.1 were sampled by Markov-Chain Monte-Carlo Simulations, to define the translational and rotational orientations of σR1.1 relative to the structures of holoenzyme and RPo that best fit the FRET distance restraints. The results were post-processed to eliminate configurations that place σR1.1 in steric clash with RNAP or DNA, and eliminate configurations that have no contact between σR1.1 and RNAP or DNA. Of the retained configurations, three to four solutions with the lowest FRET penalties were identified as the optimal solutions.

3. Materials and Methods

3.1. Plasmids

Plasmids used in this work were summarized in Table 1. pET21d-rpoB-H6 encodes C-terminal hexahistidine-tagged *E.coli* RNAP β subunit (Wang, 2008). Amber stop codon (TAG) substitution at one of the residues 106, 222, 357, 379, 643, and 937 was introduced by site-directed mutagenesis.

pEVOL-*p*AzF encodes an evolved *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (MjTyrRS) and tyrosine amber suppressor $tRNA^{Tyr}{}_{CUA}$ ($mutRNA_{CUA}$) pair(Chin et al., 2002; Young et al., 2010), which specifically charges for unnatural amino acid 4-azidophenylalanine. Two copies of the *M. jannaschii* MjTyrRS gene were constructed in the plasmid, under the control of an inducible promoter (P*aaBAD*) and a constitutive promoter (P*glns*). The plasmid was kindly provided by Peter Schultz, The Scripps Research Institute, La Jolla CA.

pET28a-NF-α^{I}NTD- α^{II}NTD encodes a N-terminal flag tagged fusion protein of two αNTDs (Wang, 2008). Plasmids pMKSe2, pT7β′, and pT7ω encoding *E. coli* RNAP β, β′, and ω subunits respectively were previously described (Naryshkin et al., 2001; Severinov et al., 1993).

Plasmids pGEMD encoding the *E.coli* σ^{70} subunit, pGEMD (-Cys) encoding a σ^{70} derivative with no Cys residues, and pGEMD (-Cys) derivatives encoding σ^{70} derivatives with single Cys residues at positions 14, 59, 366, 396, 459, 569 were described elsewhere (Bown et al., 1999; Callaci et al., 1998; Igarashi and Ishihama, 1991; Mekler et al., 2002;

Mukhopadhyay et al., 2001; Owens et al., 1998). pGEMD (-Cys) derivatives encoding σ^{70} derivatives with single Cys residues at positions 36, 46, 118, 304 were prepared by site-directed mutagenesis using pGEMD (-Cys) as a template. pGEMD (-Cys) derivatives encoding σ^{70} derivatives with amber stop codon (TAG) substitutions at residues 14, 36, 46, 59 were produced by site-directed mutagenesis using pGEMD (-Cys) as a template. pGEMD (-Cys) derivatives encoding σ^{70} derivatives with amber stop codon (TAG) substitutions at residues 14, 36, 46, 59 and single Cys residues at positions 118, 304, 366, 396, 459, 569 were constructed by site-directed mutagenesis using pGEMD (-Cys)-14TAG, pGEMD (-Cys)-36TAG, pGEMD (-Cys)-46TAG, pGEMD (-Cys)-59TAG as templates.

3.2. Site-directed mutagenesis

Site-directed mutagenesis was carried out according to the instruction manual of QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, catalog #200524). PCR primers were purchased from IDT (Integrated DNA Technologies), purified by standard desalting.

3.3. Fluorescent probes

Fluorescent donor probe tetramethylrhodamine-5-maleimide (I) was purchased from Invitrogen (catalog #T6027) (Fig.10).

Another fluorescent donor probe Cy3B maleimide (II) was purchased from GE healthcare (catalog # PA63130) (Fig.10).

Fluorescent acceptor probe A647-phosphine was synthesized from Alexa Fluor 647 Carboxylic Acid, Succinimidyl Ester (Invitrogen, catalog # A-20106), following three steps adapted from the procedures in Chakraborty et al., 2010 (Chakraborty et al., 2010) (Fig.10):

1) A647-pentalnoyl-ethylenediaminyl-trityl (IV).

   Dissolve compound III (Alexa Fluor 647 Carboxylic Acid, Succinimidyl Ester; 5.0 mg; 4.0 μmole) and N-Trityl-1,2-ethanediamine hydrobromide (Fluka, catalog #00589; 23 mg; 60 μmole) in 1 ml anhydrous dimethylformamide (DMF). Triethylamine (TEA; 10 μl; 71 μmol) was added and the reaction was carried out at room temperature for 1 hour with stirring. The reaction mixture was dried under vacuum, dissolved in 40% Acetonitrile/60% H_2O. Product IV was purified by reversed-phase HPLC (solvent A: H_2O; solvent B: 100% acetonitrile; gradient: 10 to 75% B in 30 min at 2 ml/min) and dried under vacuum.

2) A647-pentanoyl-ethylenediamine (V)

   Add appropriate amount of chloroform to compound IV. Then add one half volume of trifluoroacetic acid (TFA) to the mixture, so that ratio by volume for chloroform and TFA is 2 to 1. The reaction mixture was incubated for 30 min at room temperature, and was dried under vacuum. The product V was re-dissolved with 10% Acetonitrile/0.1% TFA/89.9% H_2O, purified using reversed-phase HPLC (solvent A: 0.1% TFA in H_2O; solvent B: 100% acetonitrile; gradient: 10 to 75% B in 30 min at 2 ml/min) and dried under vacuum. MS (MALDI): calculated, *m/z* 901 (MH^+); found, 901.

3) A647-pentanoly-ethylenediaminyl-phosphine (A647-phosphine; VI)

Dissolve compound V in 300 μl of 50%DMF/50% degassed H_2O. Quantify the amount of compound V by UV/Vis spectrophotometry. The typical yield was 2.2 mg/2.5 μmol. N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDAC; Sigma-Aldrich; 9.6 mg; 50 μmol) in 50 μl of degassed H_2O, N-Hydroxysulfosuccinimide sodium salt (NHSS; Sigma-Aldrich; 9.6 mg; 35 μmol) in 50 μl of degassed H_2O, and 1-Methyl-2-Diphenylphosphinoterephthalate (MDPT; Wang 2008; 14.5 mg; 35 μmol) in 200 μl of DMF were combined with compound V. N,N-Diisopropylethylamine (DIPEA; Sigma-Aldrich; 13 μl; 72.5 μmol) was added to the reaction mixture, followed by incubation for 3 hours at 37 ℃ with stirring.

An alternative route for synthesizing VI: Dissolve compound V in 500 μl of anhydrous DMF. If not completely soluble, add 10 μl of DIPEA. Dissolve NHS-phosphine (Thermo Scientific; #06162; 10 mg; 21.6 μmol) in in 500 μl of anhydrous DMF, which was then combined with V. The reaction was incubated for 3 hours at 37 ℃ with stirring.

The reaction mixture was divided into 6 aliquots and dried under vacuum. It was re-dissolved in 30%Acetonitrile/0.1% TFA/69.9% H_2O, before purification by reversed-phase HPLC (solvent A: 0.1% TFA in H_2O; solvent B: 100% acetonitrile; gradient: 30 to 100% B in 30 min at 2 ml/min) and dried under vacuum. MS (MALDI): calculated, *m/z* 1248.4 (MH^+); found, 1247.4. The minor peak was

mainly the oxidized product, with MS (MALDI): calculated, *m/z* 1264 (MH^+); found, 1263.3.

3.4. Preparation of holoenzyme derivatives for experimental strategy I

The procedure was summarized in Fig.9.

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

Inclusion bodies containing *E.coli* RNAP wild-type β subunit, inclusion bodies containing *E.coli* RNAP wild-type β′ subunit, and inclusion bodies containing *E.coli* RNAP wild-type ω subunit were prepared as in Naryshkin et al., 2001.

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

The α fusion protein (Flag-αNTD^{I}-GSGGSG-αNTD^{II}) comprising an N-terminally Flag-tagged N-terminal domain of the first *E. coli* RNAP α subunit (α residues 1-235; αNTD^{I}), followed by a linker of GlySerGlyGlySerGly, followed by the N-terminal domain of the second *E. coli* RNAP α subunit (α residues 1-235; αNTD^{II}) was prepared as follows: *E. coli* strain BL21(DE3) (Invitrogen, Inc.) was transformed with plasmid pET28a-NF-αNTD^{I}-αNTD^{II} (Table 1). 50 ml LB medium containing 40 μg/ml kanamycin were inoculated with a single colony and incubated for 16 h at 37 ℃ with shaking. 10 ml of the overnight culture were transferred into 1 L LB medium containing 40 μg/ml kanamycin. The culture were incubated at 37 ℃ with shaking, induced by addition of IPTG to 1 mM when OD_{600} reaches 0.6, and incubated for an additional 3 h at 37 ℃. Cells were harvested by centrifugation (4,500xg; 20 min at 4 ℃), 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)] and lysed using an Avestin EmulsiFlex-C5 cell disrupter (Avestin, Inc.). Lysates were centrifuged (20,000xg; 20 min at 4 ℃), supernatants were collected, and protein was precipitated by addition of ammonium sulfate (35 g per 100 ml supernatant) followed by centrifugation (20,000xg; 10 min at 4 ℃). Pellets were 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. Columns were washed with 50 ml TBS and were eluted with 25 ml (collect 1-ml per fraction) of TBS containing 0.1 mg/ml FLAG peptide (Sigma Aldrich, #F3290). Fractions containing αNTD^{I}-αNTD^{II} were pooled, and precipitated by addition of ammonium sulfate and stored in aliquots as ammonium sulfate pellets at -80 ℃. Typical yields of Flag-αNTD^{I}-GSGGSG-αNTD^{II} were 50 mg/L.

3.4.3. Preparation of β derivatives

4-azidophenylalanine was incorporated into β subunit as follows (Chin et al., 2002). *E.coli* competent cells BL21 (DE3) were co-transformed with pEVOL-*p*AzF, which encodes the aaRS/tRNA pair specifically for 4-azidophenylalanine, and pETd-rpoB-H6 bearing amber mutation at selected positions (Table 1). Single colonies from the transformants were inoculated into 50 ml of LB medium supplied with 100 μg/ml ampicillin and 35 μg/ml chloramphenicol, and the culture were grown for 16 h at 37 ℃ with shaking. The cells were harvested by centrifugation (4000xg; 10 min at 4 ℃), resuspended in 10 ml 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, and 1 μg/ml biotin], and transferred to 500 ml of M9+ medium containing 100 μg/ml ampicillin, 35 μg/ml chloramphenicol and 1 mM 4-azido-L-phenylalanine (Chem-Impex International, Inc., Catalog #06162). The cultures were grown in dark at 37 ℃ with shaking. The growth was monitored by measuring absorbance of the cell culture at 600 nm. Arabinose was added to the cultures at a final concentration of 0.02% until OD_{600} = 0.5, while isopropyl β-D-1-thiogalactopyranoside (IPTG) was supplied at a final concentration of 1 mM until OD_{600} = 0.6. Cells were grown for another 3h after induction, harvested by centrifugation (4500xg; 30 min at 4 ℃) and stored at -80 ℃.

Inclusion bodies containing β subunits were purified as in Naryshkin et al., 2001 with modifications. Cell pellet from 500 ml culture was suspended in 25 ml lysis buffer [40 mM Tris-HCl, pH 7.9, 300 mM KCl, 10 mM EDTA, and 0.2% (w/v) sodium deoxycholate (Sigma Aldrich)]. Cells were lysed using an Avestin EmulsiFlex-C5 (Avestin Inc.). Cell lysis was centrifuged at 24,500g for 30 min and the pellet containing the inclusion bodies was collected. The pellet was suspended in 20 ml of lysis buffer plus 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 24,500 g at 4 ℃. The washing procedure was repeated once using the lysis buffer plus 0.5% TritonX-100. The pellet was then suspended in 8 ml of storage buffer (40mM Tris-HCl, pH 7.9, 300mM KCl, 10mM EDTA, and 10% glycerol), divided into 1-ml aliquots, and stored at -80 ℃. The purified inclusion bodies contain ~50 mg proteins for 500 ml culture.

4-azidophenylalanine-containing β subunit was labeled with Alexa647-phosphine through Staudinger-Bertozzi ligation (Chakraborty et al., 2010). Inclusion bodies of about 2 mg proteins were dissolved in 0.5 ml of denaturing labeling buffer (50 mM Tris-HCl, pH 7.9, 6 M guanidine hydrochloride), and the exact amount of proteins were determined by Bradford assay. Appropriate amount of fluorescent dye A647 phosphine was dissolved in 50 μl of denaturing labeling buffer, and was added to the protein solution at a dye-to-protein molar ratio of 8 to 1. This reaction, namely, Staudinger-Bertozzi ligation, was carried out in darkness, at 37 ℃ with rotation for 16 h. The reaction mixture was then loaded onto a 15 ml column packed with Bio-Gel P-30 Gel (Bio-Rad, #150-4154), which was pre-equilibrated in the denaturation buffer (50 mM Tris–HCl, pH 7.9, 6 M guanidine hydrochloride, 10 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 10% glycerol, and 10mM DTT), followed by elution in the same buffer. The fractions containing labeled proteins were identified by US-Vis spectroscopy, combined and stored for the following experiments.

Labeling efficiencies and specificities were quantified as in Chakraborty et al., 2010.

3.4.4. Preparation of TMR-labeled σ^{70}

Unlabeled *E.coli* σ^{70} and TMR-labeled σ^{70} were prepared as in Mukhopadhyay et al., 2003 with modifications.

pGEMD enconding *E.coli* σ^{70}, pGEMD (-Cys) encoding a σ^{70} derivative with no Cys residue or a pGEMD (-Cys) derivative encoding a σ^{70} derivative with a single Cys residue at position 14, 36, 46, or 59, were transformed into *E.coli* BL21 (DE3) competent cells (Table). Single colonies were inoculated into 50 ml LB medium with 100 μg/ml

ampicillin, and incubated for 16 h at 30 ℃ with shaking. 10 ml of the overnight culture was inoculated into 500 ml of LB medium with 100 μg/ml ampicillin. The culture were subsequently grown at 37 ℃ with shaking until OD_{600} reached 0.6, supplemented with IPTG to 1 mM, and grown for another 3 hours before harvest by centrifugation (4500xg; 30 min at 4 ℃).

Inclusion bodies containing σ^{70} or σ^{70} derivatives were purified as in 3.4.3. The purified inclusion bodies contain ~100 mg protein per 500 ml culture.

Inclusion bodies were solubilized in the denaturing reconstitution buffer (50 mM Tris–HCl, pH 7.9, 6 M guanidine hydrochloride, 10 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 10% glycerol, and 10mM DTT), centrifuged to remove particles and adjusted to 2.5 mg/ml by the same buffer. The protein solution (50 ml) was transferred into a dialysis membrane (Spectra/Por Dialysis Membrane, MWCO 3500 Da, #132724) and dialyzed against TGEβ buffer [50 times the volume of the protein solution; 20 mM Tris–HCl, pH 7.9, 0.1 mM EDTA, 5% glycerol, and 5 mM 2-mecaptomethanol (2-ME)] containing 0.2 M NaCl (36 hours at 4 ℃; two changes of buffer).

The sample was centrifuged (24500xg; 20 min at 4 ℃) to remove particulates and applied to 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 dithiothreitol (DTT)] containing 0.3 M NaCl. The column was washed with 16 ml TGED buffer containing 0.3M NaCl and eluted in 2-ml fractions by a 160-ml linear gradient of 0.3-0.5 M NaCl in TGED (σ^{70} typically eluted at 0.36 M NaCl in TGED). The peak fractions were analyzed on SDS-PAGE and fractions containing σ^{70} or σ^{70} derivatives were pooled. The target

proteins were precipitated by addition of 0.25 g/ml ammonium sulfate, incubated for 30 min at 4 ℃, followed by centrifugation(16000xg; 20 min at 4 ℃), and were stored as ammonium-sulfate pellet in aliquots at -80 ℃. Yields typically are 40-50 mg. Purities typically are >95%.

Labeled σ^{70} derivatives were prepared according to Cys-specific chemical modification (Kim et al., 2008). An aliquot for ammonium-sulfate pellet containing 5-6 mg of σ^{70} derivative was dissolved in 1 ml maleimide-labeling buffer (50 mM Hepes, pH 7.1, 0.3 M NaCl, 0.1 mM EDTA, and 5% glycerol). The protein was subjected to a solid-phase reduction on immobilized reductant columns (Thermo Scientific, #77701), to cleave any pre-existing disulfide bonds. Appropriate amount of fluorescent dye tetramethylrhodamine-5-maleimide (Invitrogen, # T6027) was freshly dissolved in 10 μl dimethyl sulfoxide (DMSO), and was immediately added to the protein solution at a dye-to-protein molar ratio of 10 to 1. For a typical labeling reaction, the reaction mixtures (1 ml) contain: 40 μM σ^{70} derivative and 400 μM tetramethylrhodamine-5-maleimide in maleimide-labeling buffer. After 2 hours incubation at 4 ℃, products were applied to 15 ml column packed with Bio-Gel P-30 Gel (Bio-Rad, #150-4154), pre-equilibrated in TGED buffer containing 0.2 M NaCl, followed by elution in the same buffer. The fractions containing labeled proteins were identified by US-Vis spectroscopy, pooled, concentrated, and stored in 20 mM Tris–HCl, pH 7.9, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol in aliquots at -80 ℃. Labeling efficiency was calculated as in Mukhopadhyay et al., 2003.

3.4.5. Reconstitution of holoenzyme derivatives

Labeled RNAP derivatives were prepared as in Naryshkin et al., 2001 with modifications.

Reconstitution of RNAP core derivatives was carried out under denaturing conditions, from labeled β derivatives, inclusion bodies containing wild-type β′ subunit, inclusion bodies containing ω subunit, and purified flag-αNTD^{I}-GSGGSG-αNTD^{II}. The reconstitution mixtures (15 ml) typically contained: 1 mg (6.7 nmol) A647-labeled β derivative (point to) , 4.2 mg (26.8 nmol) β′, 0.7 mg (67 nmol) ω, and 0.8mg (13.3 nmol) flag-αNTD^{I}-GSGGSG-αNTD^{II} in denaturation buffer (50 mM Tris–HCl, pH 7.9, 6 M guanidine hydrochloride, 10 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 10% glycerol, and 10mM DTT). Reconstitution mixtures were transferred into the dialysis membrane (Spectra/Por Dialysis Membrane, MWCO 3500 Da, #132724) and dialyzed against 7500 ml renaturation buffer (50 mM Tris–HCl, pH 7.9, 200 mM KCl, 10 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 20% glycerol, and 5 mM 2-ME; 36 hours at 4 ℃; two changes of buffer; last dialysis omitted 2-ME in the buffer).

Renatured RNAP core derivatives were incubated with corresponding TMR-labeled σ^{70} derivatives to form target RNAP holoenzyme derivatives. Renatured RNAP core derivatives were centrifuged (24500xg; 20 min at 4 ℃) to remove particulates, supplemented with TMR-labeled σ^{70} derivative at a σ^{70} to core molar ratio of 2:1, and incubated at 30 ℃ for 45 min.

Reconstituted RNAP holo derivatives were purified by anti-flag affinity chromatography and mono-Q ion exchange chromatography. The protein sample was loaded onto a 3-ml column packed with anti-FLAG M2 affinity gel pre-equilibrated with buffer A (50 mM

Tris-HCl, pH 7.4, 150 mM NaCl, 0.1 mM EDTA, and 5% glycerol). The column was washed with 50 ml of buffer A. The bound protein was eluted by 15 ml buffer A containing 100 μg/ml FLAG peptide. Eluted sample was applied to 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, and 1 mM DTT) containing 0.3 M NaCl. The column was washed with 16 ml of the equilibration buffer and eluted in 2-ml fractions of a 160-ml linear gradient of 0.3-0.5 M NaCl in TGED. The peak fractions were analyzed on SDS-PAGE and fractions containing RNAP holo derivatives were identified, pooled, concentrated and stored in 20 mM Tris–HCl, pH 7.9, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol in aliquots at -80 ℃.

3.5. Preparation of holoenzyme derivatives for experimental strategy II

3.5.1. Preparation of σ^{70} derivatives labeled with Alexa 647 and Cy3B

BL21 (DE3) competent cells were co-transformed with a pGEMD (-Cys) derivative (which encodes a σ^{70} derivative bearing amber mutation at 14, 36, 46, or 59 and single Cys residue at 118, 304, 366, 396, 459, or 569; Table), and pEVOl-pAzF (Young et al., 2010). Cell production and purification of inclusion bodies were done as in 3.4.3. The purified inclusion bodies contain ~35 mg proteins per 500 ml culture, with purity over 80%.

The labeling reaction mixture (1 ml) contained: 43 μM σ^{70} derivatives, 2.2 mM A647-phosphine in denaturing labeling buffer (50 mM Tris-HCl, pH 7.9, 6 M guanidine hydrochloride). The reaction was carried out and purified as described in 3.4.3.

The concentration of σ^{70} derivatives was adjusted to about 0.5 mg/ml (about 8 ml) by the denaturing reconstitution buffer before dialysis. Renaturation and purification of A647-labeled σ^{70} derivatives were done as described in 3.4.4. Fractions containing σ^{70} derivatives were pooled and stored at 4 ℃ for the next treatment. The recovery ratio of σ^{70} derivative was typically 30% and the purity was over 95%.

The A647-labeled σ^{70} derivatives were then conjugated with Cy3B maleimide. First, the σ^{70} derivative in TGED buffer was undergone buffer exchange with maleimide-labeling buffer (50 mM Hepes, pH 7.1, 0.3 M NaCl, 0.1 mM EDTA, and 5% glycerol) using centrifugal filters (Millipore, Amicon Ultra MWCO 30,000 Da). The reaction mixture (500 μl) contained 15 μM σ^{70} derivative and 150 μM Cy3B maleimide in maleimide-labeling buffer. Doubly labeled σ^{70} derivatives were purified and stored as described in 3.4.4.

The labeling efficiency for a σ^{70} derivative labeled by A647 and Cy3B was determined by UV/Vis absorption spectra as follows:

$$eff_{A647} = (OD_{655} / \varepsilon_{A647,655}) / [\sigma^{70}]$$

$$eff_{Cy3B} = (OD_{555} / \varepsilon_{Cy3B,555}) / [\sigma^{70}]$$

where $\varepsilon_{A647,655}$ is the molar extinction coefficient for Alexa 647 at 555 nm (239,000 M^{-1} cm^{-1}), $\varepsilon_{Cy3B,555}$ is the molar extinction coefficient for Cy3B at 555 nm (130,000 M^{-1} cm^{-1}) and $[\sigma^{70}]$ is the concentration of σ^{70} determined by Bradford assay. Efficiencies for labeling typically are ~ 95%.

3.5.2. Reconstitution of holoenzyme derivatives

RNAP derivatives for experimental strategy II were prepared freshly before smFRET measurements. σ^{70} derivatives labeled with A647 and Cy3B were incubated with 2 times molar excess of *E. coli* RNAP core (Epicentre, #C90100) for 25 min at 37 ℃.

3.6. Ribogreen transcription assay

Transcriptional activities of RNAP derivatives were measured using ribogreen transcription assays, which included the following steps:

1) RNAP holoenzyme derivatives and wild-type RNAP holoenzyme were incubated with DNA fragment N25 (Revyakin et al., 2006) for 20 min at 37 ℃. 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, and 10 μg/ml bovine serum albumin(BSA)]. The negative control omitted RNAP in the reaction.

2) NTPs were added to the mixtures at a final concentration of 100 μM each. Transcription reactions were allowed to proceed for 1 h at 37 ℃.

3) DNA templates were digested by addition of 1 μl of 5 mM $CaCl_2$ and 2 U of DNase I (Ambion, #AM2222) in each reaction, and were incubated for 90 min at 37 ℃.

4) Reaction mixtures were supplemented with 1/500 diluted ribogreen (Invitrogen; #R11491) in 100 μl TE (10 mM Tris-Cl, pH8.0, 1 mM EDTA) and incubated for 10 min at room temperature.

5) Reaction mixtures were transferred to Corning 3686 assay plate. Fluorescence intensity were measured by Tecan GENios Pro plate reader (λex = 485 nm; λem = 535 nm).

Fluorescence intensity (FI) is the readout for the ribogreen assay. RNAP transcription activity is quantified by the ratio of FI for the sample to FI for negative control (which does not contain RNAP). Transcription activity for a RNAP derivative is expressed as a percentage of FI for the RNAP derivative over FI for the wild-type RNAP.

3.7. smFRET sample preparation

For RNAP holoenzyme, 20 μl transcription buffer (TB; 50 mM Tris-HCl, pH8.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 5% glycerol, and 100 μg/ml BSA; filtered) containing 20 nM RNAP holoenzyme derivatives was warmed at 37 ℃. Aliquots (0.2 μl) were transferred to the pre-warmed 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] and incubated for 5 min before smFRET data collection.

For RPo, 20 μl TB containing 20 nM RNAP holo derivatives and 66 nM DNA fragment *lacCONS*-14(-107/+56) (prepared according to procedures in Mukhopadhyay et al., 2003;

sequence as in Chakraborty et al., 2012) was incubated for 25 min at 37 ℃. Aliquots (0.2 μl) were transferred to the pre-warmed tubes containing 40 μl of KT buffer and incubated for 5 min before smFRET data collection.

10 μl of sample in KT was added to the 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.

3.8. smFRET set-up and data acquisition

smFRET data was collected by confocal microscopy with alternating laser excitation (ALEX). The set-up (Fig.5 A) was essentially as described in (Kapanidis et al., 2005; Kapanidis et al., 2004). Direct excitation of the donor was provided by a green laser (532 nm; Compass 215M-20; Coherent, Inc.), and direct excitation of the acceptor was provided by a red laser (638 nm; Radius 635-25; Coherent, Inc.). Lasers were operated at continuous-wave excitation intensities of 160-200 μW at 532 nm and 70-80 μW at 638 nm and were alternated at 25 μs intervals using an acousto-optical modulator (Neos Technologies, Inc.). The excitation beams were coupled through a single-mode fiber, collimated and directed to an Olympus IX71 inverted microscope (Olympus America, Inc.). Reflected on a dichroic beamsplitter, the beams were focused 20 μm from the bottom coverslip through a 60x oil-immersion objective. Desired temperature (37 ℃ for this study) during data acquisition was maintained by heating the objective using the objective heater system from Bioptechs. Fluorescence emission from the sample was collected through the objective, focused through a 100 μm pinhole, spectrally split by a

dichroic mirror into donor emission channel and acceptor emission channel. The photons from both channels were passed through filters (for TMR and Cy3B, 585BP70; for Alexa 647, 650LP; Chroma), and their photo arrival times were recorded by two avalanche photodiode detectors (APD; SPCM-AQR-15; Perkin-Elmer, Inc.). Data was collected in a 30-min time period for each sample.

3.9. smFRET data analysis

For data analysis, 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 emissions $F_{D_{exc}}^{A_{em}}$, $F_{D_{exc}}^{D_{em}}$, $F_{A_{exc}}^{A_{em}}$, *and* $F_{A_{exc}}^{D_{em}}$. For each above-threshold photon burst (> 15-30 photons), the stoichiometry parameter (S) was calculated, as follows (Kapanidis et al., 2005; Kapanidis et al., 2004):

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

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

The donor-acceptor energy-transfer-efficiency parameter, E, was calculated as:

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

Two dimensional *E*-*S* histograms allow identification and sorting of diffusing species. The parameter *S* permits identification of molecules containing both donor and acceptor (D-A, *S* = 0.3-0.8; desired species), molecules containing only a donor (D only, *S*~1; undesired species), and molecules containing only an acceptor (A only, *S* <0.3; undesired species). For species containing both donor and acceptor (D-A), one-dimensional *E* histograms were plotted and fitted with Gaussian functions. The width of distribution and the mean E value can be extracted from each subpopulation.

Donor-acceptor distances (*R*) were calculated as follows (Clegg, 1992):

$$R = R_0[(1/E) - 1]^{1/6} \qquad \text{Equation (2)}$$

where R_0 is the distance at which 50% of the energy is transferred, given by the function (Clegg, 1992):

$$R_0 = 9780(n^{-4}\kappa^2 Q_D J)^{1/6} \text{ Å} \qquad \text{Equation (3)}$$

where *n* is the refractive index of the medium (1.4, Clegg, 1992), κ^2 is a geometrical factor depending on the relative orientation of the donor and acceptor transition dipoles [approximated as 2/3-justified by fluorescence anisotropy measurements (Table 5) indicating donor and acceptor reorient on the time scale of the donor excited-state life time, and, in most cases, also by the fact that *E* < 0.5 ((Wu and Brand, 1992), Table 6,7)], Q_D is quantum yield of the donor in the absence of the acceptor (Table 3,4), and *J* is the spectral overlap integral of the donor emission spectrum and acceptor absorption spectrum (Table 3 and 4, determined using corrected spectra for donor-only and acceptor-only controls).

3.10. Measurements of fluorescence quantum yields

The relative quantum yield of a fluorophore was determined by comparison to a reference fluorophore with a well-known quantum yield. Rhodamine 101(Q = 1.0 in ethanol, T. Karstens and K. Kobs, 1980) was used as the reference fluorophore for Cy3B and TMR. Cy5 (Q = 0.27 in PBS, (Mujumdar et al., 1993)) was used as reference fluorophore for Alexa 647. The quantum yield (Q) of a given sample was calculated by (Lakowicz, 1999):

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

where I is the integrated fluorescence intensity, n is the refractive index of solvent, and OD is the optical density (absorption). The subscript R refers to the reference fluorophore of known quantum yield. Measurements were carried out in solution in a sub-micro fluorometer cuvette with 10-mm path length. The values were recorded in the context of RNAP holoenzyme and RPo, using donor-only and acceptor-only proteins (Table 3,4).

3.11. Measurements of fluorescence anisotropy

Steady-state fluorescence anisotropies were measured in solution containing donor-only or acceptor-only RNAP derivatives, in the context of holoenzyme and RPo (Table 5). Anisotropies were measured using a T-format spectrofluorometer equipped with excitation and emission polarizers (PTI photon technology, Inc), at 37 ℃. Excitation and emission wavelengths were 550 nm and 580 nm for the donor, and 635 nm and 670 nm for the acceptor. Anisotropy (A) was calculated as (Chen and Bowman, 1965):

$$A = (I_{VV} - GI_{VH})/(I_{VV} + 2GI_{VH})$$

where I_{VV} and I_{VH} are fluorescence intensities with the excitation polarizer at the vertical position and the emission polarizer at, respectively, the vertical position and the horizontal position. G is the grating correction factor, which is measured by:

$$G = I_{HV}/I_{HH}$$

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

3.12. Distance-restrained docking

This part of analysis was performed by Dr. Jennifer L. Knight (Schrödinger).

3.12.1. Distance-restrained docking: generation of starting models

The structure of RP_o was modeled by Ec RNAP holoenzyme (PDB accession code 4IGC Murakami, 2013) with the downstream DNA duplex incorporated from *T. thermophilus* RP_o (PDB accession code 4G7H; Zhang et al., 2012) and σR1.1 removed. The structure of RNAP holoenzyme was modeled with an open-clamp conformation as in Chakroborty et al. 2012, by rotating the β' pincer (*E. coli* β' residues 8-345 and 1326-1340; *E.coli* β residues 1318-1342; and *E. coli* σ residues 95-448) of *E. coli* RNAP holoenzyme structure 16° about an axis defined by Cα atoms of β' pincer residues 345 and 1262. σR1.1 was modeled using residues 35-87 of the NMR structure of *T. maritima* σR1.1 (PDB accession code 2K6X; Schwartz et al., 2008). Probes and linkers were modeled onto the structures using Maestro (Maestro, version 9.3, Schrödinger, LLC: New York, 2012) and sterically-allowed conformations of probes and linkers were identified using

MacroModel's Conformational Search utility within Maestro (MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2013). All linker torsion angles were sampled in 30° increments for TMR-Mal and Cy3B-Mal or 120° increments for Alexa647 and conformations which had van der Waals energy < 5000 kcal/mol in the OPLS2005 force field (Jorgensen and Tirado-Rives, 1988; Kaminski et al., 2001) were accepted. For each accepted conformation, a probe pseudoatom, corresponding to the center of the probe chromophore, was defined.

3.12.2. Distance-restrained docking: FRET-only penalty function

Modeled configurations (Y) of σR1.1-RNAP holoenzyme and σR1.1-RP$_o$ were assigned a penalty based on their deviation from experimental FRET distances. For each configuration Y, the apparent donor-acceptor distance corresponding to the ith FRET restraint was defined to be:

$$R_{Y,i}^{calc} = R_{o,i}\left(\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[1+\left(\frac{R_{d_m a_n}}{R_{o,i}}\right)^6\right]^{-1}\right)^{-1} - 1\right)^{\frac{1}{6}}$$

where M is the total number of donor pseudoatoms d_m, and N is the total number of acceptor pseudoatoms a_n. $\pi_{FRET}(Y, i)$, the configuration-specific, donor-acceptor-pair-specific penalty term for the deviation between the calculated donor-acceptor distance ($R_{Y,i}^{calc}$) and the corresponding experimental distance (R_i^{obs}), and $\pi_{FRET}(Y)$, the corresponding configuration-specific, global penalty function, were calculated as:

$$\pi_{FRET}(Y,i) = \left(2\pi\sigma_i^2\right)^{-1/2} exp\left[\frac{-\left(R_{Y,i}^{calc} - R_i^{obs}\right)^2}{2\sigma_i^2}\right]$$

$$\pi_{FRET}(Y) = \prod_{i=1}^{I} \pi_{FRET}(Y,i)$$

where σ_i is the error associated with the target distance restraint, R_i^{obs}, and I is the total number of donor-acceptor pairs (48 for RP_o; 47 for RNAP holoenzyme). Values of R_i^{obs} > 0.5 R_o and < 1.75 R_o were assigned uncertainties (σ_i) corresponding to 15% of the experimental target distance. Values of R_i^{obs} < 0.5 R_0 were assigned an upperbound of 0.5 R_0 with σ_i = 15% of 0.5 R_0. Values of R_i^{obs} > 1.75 R_0 were assigned a lowerbound of 1.75 R_0 with σ_i = 15% of 1.75 R_0. Experimental data for which two peaks were observed, the distance associated with the predominant peak was used with σ_i = 15% of R_i^{obs}. Where both peaks were comparable or could not be resolved, the average efficiency was used to compute the apparent distance with σ_i = 30% of R_i^{obs}.

3.12.3. Distance-restrained docking: Markov-Chain Monte-Carlo simulations

In each of RNAP RPo and RNAP holoenzyme models, Markov-chain Monte Carlo simulations (Metroplis et al., 1953) employing FRET restraints for were performed. In each simulation, 50 random probe pseudoatoms were selected to model each chromophore. Distance-restrained docking was performed essentially as in Knight et al., 2005. 100, 000 trial configurations (Y) were generated from previously-accepted configurations (X) by translations and rotations of σR1.1 in which translations and angles of rotation were selected from Gaussian distributions about values in X and the axis of rotation was selected from a Fisher distribution about values in X (Fisher, 1953). For each configuration Y, its penalty was computed and configurations were accepted with probability $\alpha(X,Y)$, as follows:

$$\alpha(X,Y) = min\left(1, \frac{\pi_{FRET}(Y)}{\pi_{FRET}(X)}\right)$$

Four independent simulations were performed with different starting configurations. Sampling parameters were selected such that 20-30% acceptance rates were obtained with each simulation. Specifically, σR1.1 translations and angles of rotation were selected from Gaussian distributions about values in X (σ_{trans} =2.0 Å; $\sigma_{rotation}$ =0.075 radians) and the axis of rotation was selected from a Fisher distribution about values in X (30.0 radians; Fisher, 1953). The simulation trajectories were post-processed such that only accepted configurations were retained that demonstrated some contact between σR1.1 and RNAP, but no severe clashes; ie minimum σR1.1(Cα atom)-[RNAP(Cα atom) or DNA(P atom)] distances were between 2 and 8 Å. Of the retained configurations, three to four solutions with the lowest FRET penalties were identified as the optimal solutions.

Table 1: Plasmids

Plasmids	Characteristics	Source
pET21d-rpoB-H6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB(CH6)	Wang, 2008
pET21d-rpoB106TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB106amber(CH6)	Wang, 2008
pET21d-rpoB222TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB222amber(CH6)	Wang, 2008
pET21d-rpoB357TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB357amber(CH6)	This work
pET21d-rpoB379TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB379amber(CH6)	This work
pET21d-rpoB643TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB643amber(CH6)	This work
pET21d-rpoB937TAG-CH6	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoB937amber(CH6)	This work
pEVOL-*p*AzF	Cm^R; ori-p15A; P_{araBAD}-aaRS/$tRNA^{Tyr}_{CUA}$	Young et al., 2010
pET28a-NF-α^{I}NTD-α^{II}NTD	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoA(1-235)-rpoA(1-235)(NFLAG)	Wang, 2008
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
pGEMD (-Cys)	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD Cys free	Owens et al., 1998

Table 1: Plasmids (continued)

Plasmids	Characteristics	Source
pGEMD (-Cys)-14C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14Cys	Mekler et al., 2002
pGEMD (-Cys)-36C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36Cys	This work
pGEMD (-Cys)-46C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46Cys	This work
pGEMD (-Cys)-59C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59Cys	Callaci et al., 1998
pGEMD (-Cys)-14TAG-118C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 118Cys	This work
pGEMD (-Cys)-14TAG-304C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 304Cys	This work
pGEMD (-Cys)-14TAG-366C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 366Cys	This work
pGEMD (-Cys)-14TAG-396C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 396Cys	This work
pGEMD (-Cys)-14TAG-459C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 459Cys	This work
pGEMD (-Cys)-14TAG-569C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 14amber 569Cys	This work
pGEMD (-Cys)-36TAG-118C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 118Cys	This work
pGEMD (-Cys)-36TAG-304C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 304Cys	This work
pGEMD (-Cys)-36TAG-366C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 366Cys	This work
pGEMD (-Cys)-36TAG-396C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 396Cys	This work

Table 1: Plasmids (continued)

Plasmids	Characteristics	Source
pGEMD (-Cys)-36TAG-459C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 459Cys	This work
pGEMD (-Cys)-36TAG-569C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 36amber 569Cys	This work
pGEMD (-Cys)-46TAG-118C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 118Cys	This work
pGEMD (-Cys)-46TAG-304C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 304Cys	This work
pGEMD (-Cys)-46TAG-366C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 366Cys	This work
pGEMD (-Cys)-46TAG-396C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 396Cys	This work
pGEMD (-Cys)-46TAG-459C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 459Cys	This work
pGEMD (-Cys)-46TAG-569C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 46amber 569Cys	This work
pGEMD (-Cys)-59TAG-118C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59amber 118Cys	This work
pGEMD (-Cys)-59TAG-304C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59amber 304Cys	This work
pGEMD (-Cys)-59TAG-366C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59amber 366Cys	This work
pGEMD (-Cys)-59TAG-396C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59amber 396Cys	This work
pGEMD (-Cys)-59TAG-459C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59amber 459Cys	This work
pGEMD (-Cys)-59TAG-569C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 59 amber 569Cys	This work

Table 1: Plasmids (continued)

Plasmids	Characteristics	Source
pGEMD (-Cys)-14TAG	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD Cys free 14amber	This work
pGEMD (-Cys)-36TAG	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD Cys free 36amber	This work
pGEMD (-Cys)-46TAG	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD Cys free 46amber	This work
pGEMD (-Cys)-59TAG	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD Cys free 59amber	This work
pGEMD (-Cys)-118C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 118Cys	This work
pGEMD (-Cys)-304C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 304Cys	This work
pGEMD (-Cys)-366C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 366Cys	Callaci et al., 1998
pGEMD (-Cys)-396C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 396Cys	Owens et al., 1998
pGEMD (-Cys)-459C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 459Cys	Bown et al., 1999
pGEMD (-Cys)-569C	Ap^R; ori-pBR322; $P_{\phi 10}$-rpoD 569Cys	Mukhopadhyay et al., 2001

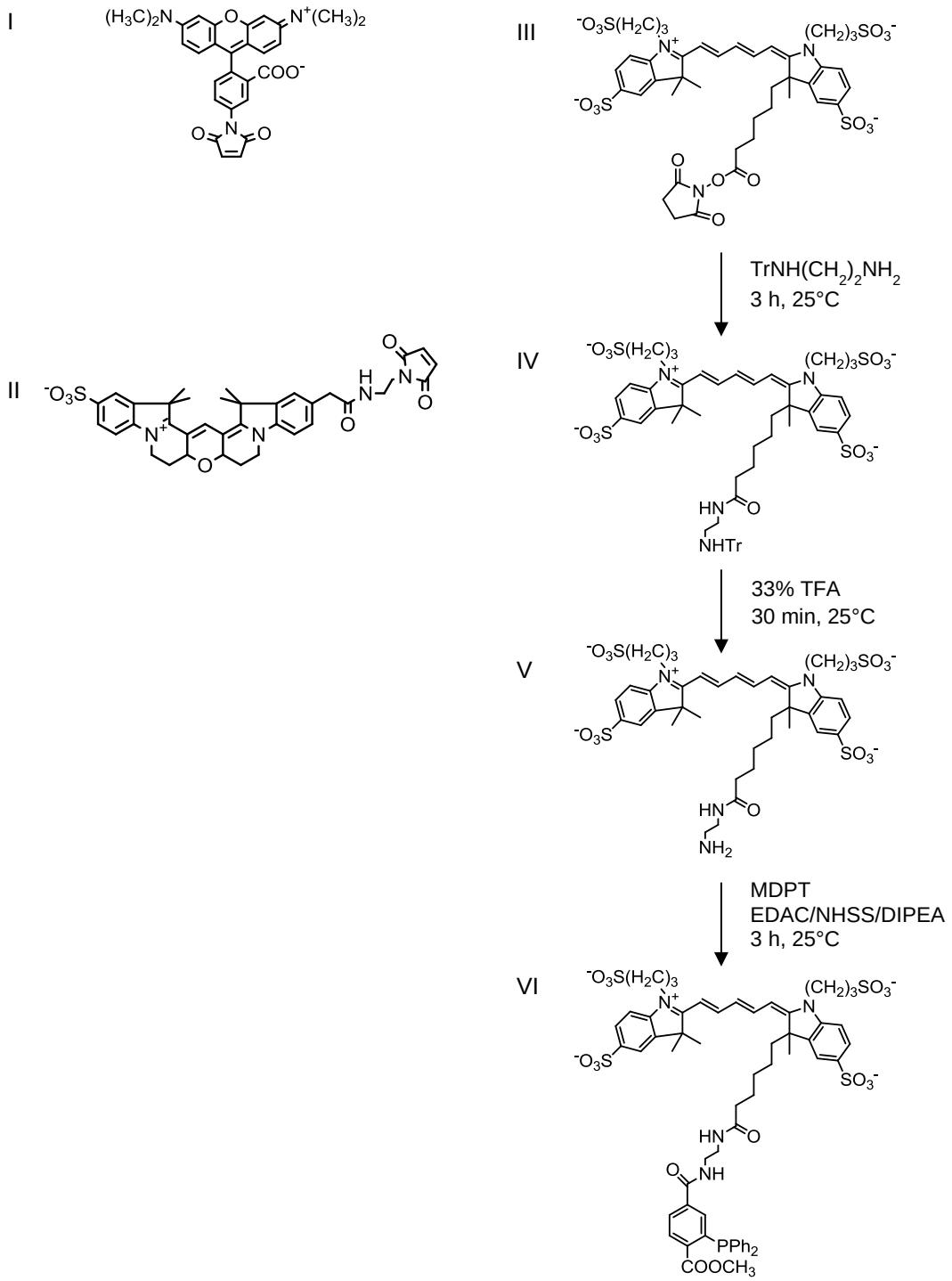


Fig. 10. Fluorescent probes.
I: tetramethylrhodamine-5-maleimide. II: Cy3B-maleimide. III-VI: Synthesis of Alexa647-phosphine[20Å].

4. Results

4.1. Site-specific incorporation of TMR into σ subunit

Specific labeling of σ subunit with TMR was accomplished by site-directed mutagenesis and Cys-specific chemical modification. The procedure involved: (1) generation of expression vector for σ subunit bearing a single Cys residue at the position of interest on σR1.1 (with other Cys residues mutated to Ala); (2) refolding and purification of σ subunits; (3) incorporation of fluorescent probe TMR into σ, by Cys-specific chemical modification using TMR-maleimide.

TMR-labeled σ derivatives showed high labeling efficiencies (> 95%) and high labeling specificities (>90%). (Fig. 11)

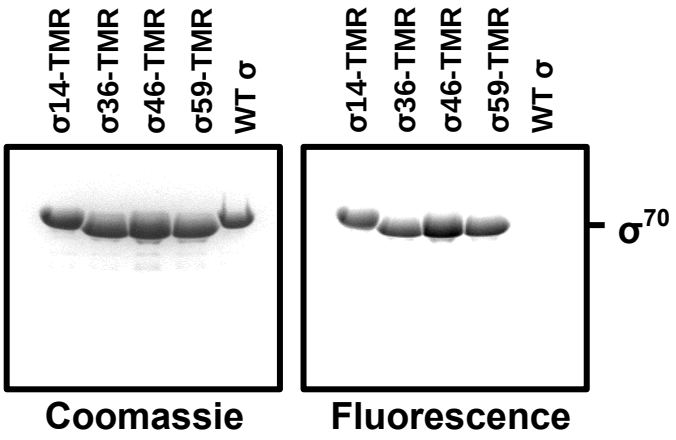


Fig. 11. Site-specific incorporation of fluorescent probe TMR into σ subunit.
σ subunits with single Cys residues at 14, 36, 46, or 59 and Cys-free σ were reacted with TMR-maleimide. Products from the labeling reactions were analyzed by 4-20% SDS-PAGE. *Left*: σ derivatives detected by Coomassie staining. *Right*: Fluorescence scanning of the same gel with 532nm excitation and 580 nm bandpass emission filters. Labeling specificities are over 90%. Labeling efficiencies are over 95%.

4.2. Site-specific incorporation of A647 into β subunit

Site-specific labeling of β subunit with A647 was achieved by unnatural amino acid mutagenesis and Staudinger ligation. The procedure involved: (1) incorporation of nonsense amber stop codon (TAG) into the gene expressing β subunit at the position of interest by site-directed mutagenesis; (2) incorporation of 4-azidophenylalanine into β subunit at the site of interest, by expressing the protein in cells which contain an orthogonal suppressor-transfer RNA (tRNA)/aminoacyl-tRNA-synthetase pair, in medium supplemented with 4-azidopheylalanine; (3) incorporation of fluorescent probe Alexa 647 into β subunits by azide-specific chemical modification, through Staudinger ligation using phosphine derivatives of Alexa 647.

Inclusion bodies containing β subunit derivatives were analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE), detected by Coomassie staining (Fig.12 A). Full-length β subunit containing 4-azidophenylalanine accounted for 40%-60% of the total protein expressed in inclusion bodies.

A647-labeled β derivatives were analyzed by SDS-PAGE, detected by Coomassie staining and fluorescence scanning. The gel picture showed high labeling efficiencies (> 90%) and high labeling specificities (>90%, Fig.12 B).

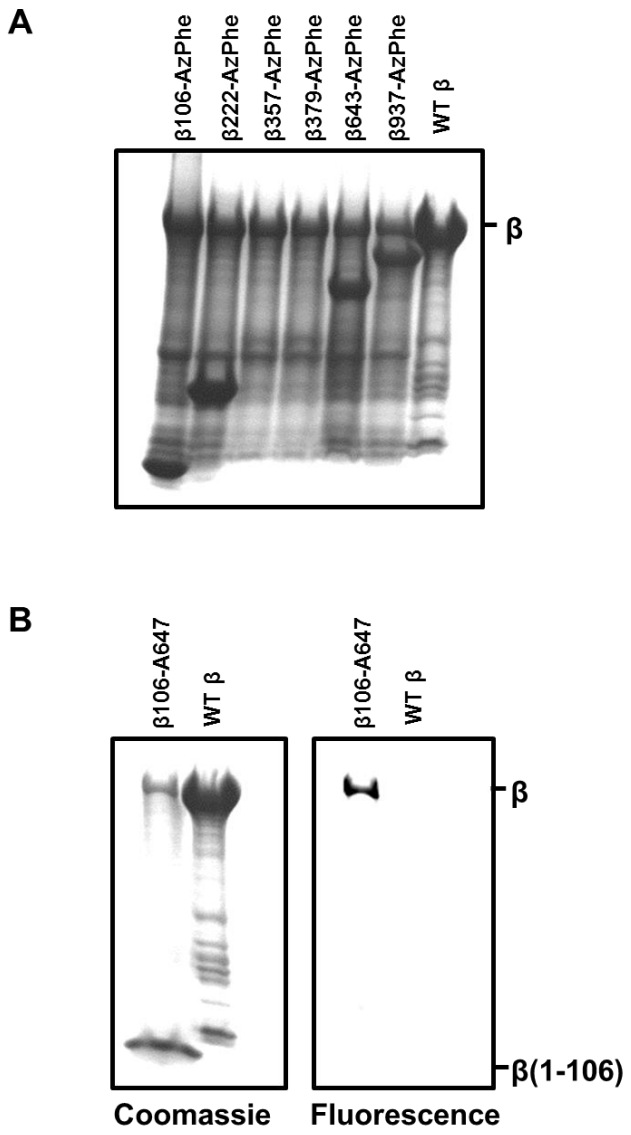


Fig. 12. Site-specific incorporation of Alexa647 into RNAP β subunit.
(A) Incorporation of 4-azidophenylalanine(AzPhe) at specified positions in RNAP β subunit. Inclusion bodies containing β subunit with AzPhe incorporated at 106, 222, 357, 379, 643 or 937 were analyzed along with inclusion bodies containing wild-type β subunit by 4-20% SDS-PAGE. Full-length β was indicated, and truncated β were expressed at various levels for each positions.
(B) Incorporation of Alexa647 into RNAP β subunit. β106 labeled with Alexa647 phosphine was shown as an example here, compared with wild-type β through the same reaction. *Left*: Products of labeling reactions of β106-AzPhe (A) and β with Alexa647-phosphine, as detected by Coomassie staining. *Right*: Fluorescence scanning of the same gel with 633nm excitation and 670 nm bandpass emission filters. Fluorescence labeling is specific to β106-AzPhe and only for full-length products.

4.3. Site-specific incorporation of fluorescent probes into RNAP holoenzyme (Experimental strategy I)

RNAP holoenzyme derivatives containing fluorescent probes were reconstituted *in vitro* from individual subunits. Fluorescently labeled and unlabeled subunits of RNAP core (β-Alexa 647/β′/α^{I}/α^{II}/ω) were combined, denatured, renatured, mixed with labeled σ subunit, and purified to prepare intact RNAP holoenzyme derivatives (β-Alexa 647/β′/α^{I}/α^{II}/ω/σ-TMR). RNAP containing only donor fluorophore and RNAP containing only acceptor fluorophore were also made in parallel using unlabeled β subunit and unlabeled σ subunit respectively.

Donor-acceptor (D+A) labeled RNAP, donor (D) labeled RNAP, acceptor (A) labeled RNAP and unlabeled RNAP were analyzed by SDS-PAGE, detected by Coomassie staining and fluorescence scanning. Correct subunit stoichiometries and fluorescence emission were observed for these prepared RNAP derivatives (Fig.13). The labeling efficiencies and specificities for labeled RNAP derivatives were generally > 90%.

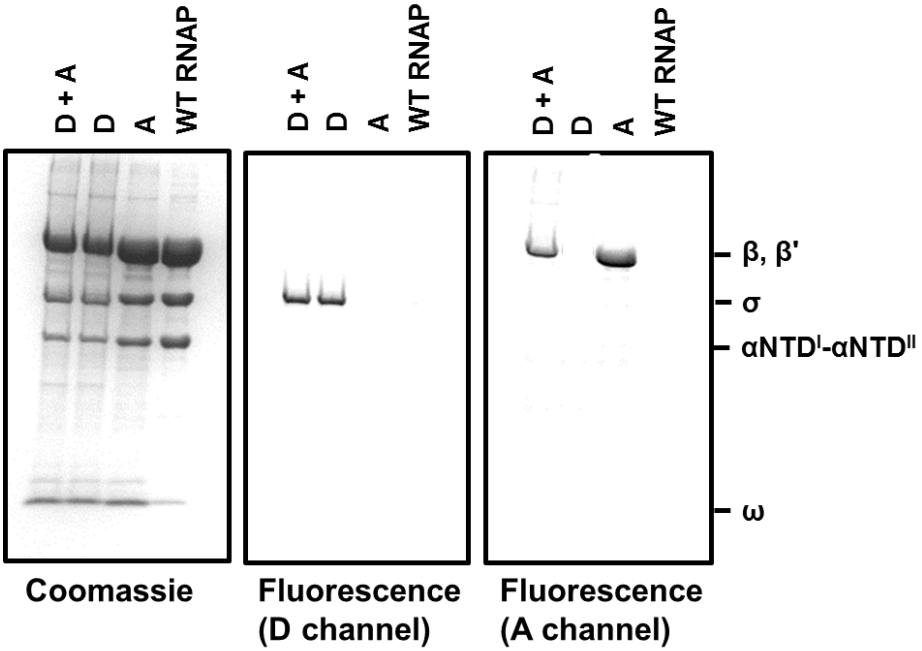


Fig. 13. Site-specific incorporation of fluorescent probes into RNAP holoenzyme. RNAP holoenzyme derivatives were reconstituted from Alexa647-labeled β derivative (or WT β subunit for donor-only RNAP), β′ subunit, Flag-αNTD^{I}-GSGGSG-αNTD^{II}, ω and TMR-labeled σ derivative (or WT σ subunit for acceptor-only RNAP). Reconstituted RNAP holoenzyme derivatives were analyzed by 4-20% SDS-PAGE.
Left: RNAP holoenzyme derivatives detected by Coomassie staining. *Middle*: Fluorescence scanning of the same gel with 532nm excitation and 580 nm bandpass emission filters. *Right*: Fluorescence scanning of the same gel with 633nm excitation and 670 nm bandpass emission filters.

4.4. Production of RNAP holoenzyme derivatives with Cy3B/Alexa 647-labeled σ (Experimental strategy II)

Site-specific incorporation of Cy3B and Alexa 647 into σ subunit was accomplished by unnatural amino acid mutagenesis, Staudinger ligation, and Cys-specific modification. The procedure involved: (1) site-directed mutagenesis of the gene expressing σ subunit, to incorporate a nonsense amber stop codon (TAG) at the position of interest on σR1.1, and a single Cys codon at one of the reference probe sites on σR2-4 (with other Cys residues mutated to Ala); (2) preparation of σ subunit containing 4-azidophenylalanine at the position of TAG codon, accomplished by growing cells which contain an engineered suppressor-transfer RNA (tRNA)/aminoacyl-tRNA-synthetase pair, in a medium supplemented with 4-azidopheylalanine; (3) incorporation of fluorescent probe Alexa 647 into σ through Staudinger ligation using phosphine derivatives of Alexa 647; (4) refolding and purification of Alexa 647-labeled σ derivatives; (5) incorporation of Cy3B into Alexa647-labeled σ derivative through Cys-specific modification using Cy3B-maleimide. Singly labeled σ derivatives, with only donor fluorophore or acceptor fluorophore were prepared in parallel.

Analysis of singly labeled and doubly labeled σ derivatives showed correct labeling by the intended fluorophore with high labeling efficiency and specificity (Fig.14). RNAP holoenzyme derivatives in experimental strategy II were reconstituted from unlabeled RNAP core and labeled σ derivatives.

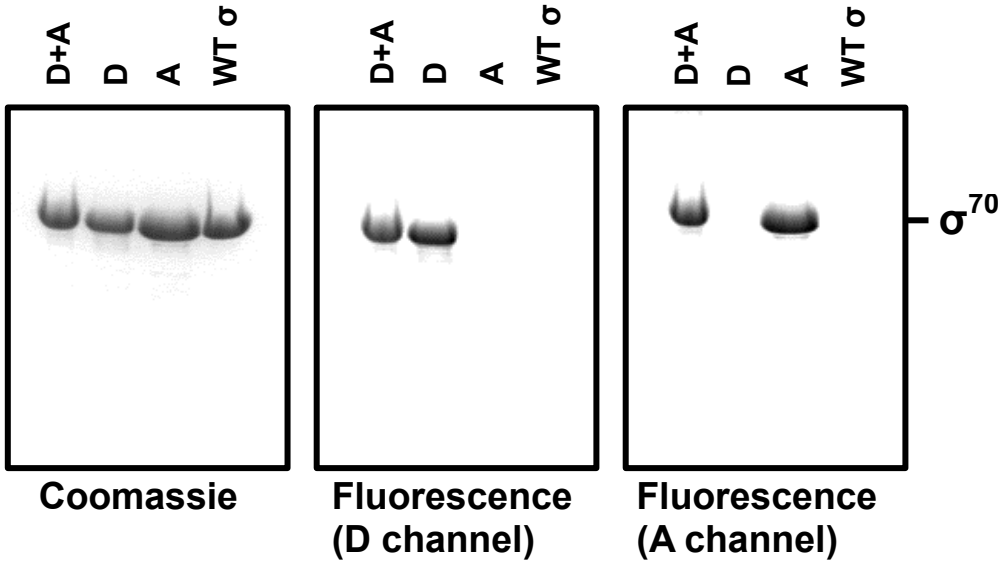


Fig. 14. Site-specific incorporation of Alexa647 and Cy3B into σ^{70}.
Labeled σ derivatives and unlabeled σ were analyzed by 4-20% SDS-PAGE.
Left: Doubly labeled σ, Cy3B-labeled σ, Alexa647-labeled σ and unlabeled σ detected by Coomassie staining. *Middle*: Fluorescence scanning of the same gel with 532 nm excitation and 580 nm bandpass emission filters. *Right*: Fluorescence scanning of the same gel with 633 nm excitation and 670 nm bandpass emission filters. Labeling specificities are over 90%. Labeling efficiencies are over 95%.

4.5. Transcriptional activities of RNAP derivatives

To confirm that incorporation of fluorescent probes into RNAP has no effect on the function of RNAP, I performed ribogreen transcription assays on all RNAP derivatives. Relative transcriptional activities of RNAP derivative were expressed as percentages of activities of wild-type RNAP polymerases (see Materials and Methods 3.6). The results (Table 2) showed that over half of the RNAP derivatives maintained >90% activities, and all of the RNAP derivatives exhibited >60% activities. The results suggested that incorporation of fluorescent probes into RNAP has minimal effect on the function of RNAP, and justified the use of labeled RNAP derivatives in the following FRET measurements.

Table 2. Relative transcription activities of RNAP derivatives

		σR1.1			
		14	36	46	59
β	106	82%	78%	80%	68%
	222	60%	100%	100%	61%
	357	99%	100%	100%	100%
	379	68%	100%	100%	100%
	643	64%	63%	78%	60%
	937	100%	100%	100%	84%
σR2	108	80%	100%	100%	78%
	304	78%	100%	100%	78%
	366	78%	100%	100%	84%
	396	69%	100%	100%	78%
σR3	459	100%	100%	100%	73%
σR4	569	100%	100%	100%	73%

4.6. Spectral properties of fluorescent probes in RNAP holoenzyme and RPo

Spectral properties of fluorescent probes can be affected by the environment, mainly due to interactions between the fluorophore with other species in the surrounding environment. Therefore, measurements of quantum yield (Q), spectral overlap integral (J) and Förster parameter R_0 need to be determined for donor-acceptor fluorescent probes at each labeling sites, both in the context of RNAP holoenzyme and in the context of RPo.

Donor quantum yield (Q_D) and acceptor quantum yield (Q_A, data not shown) were determined for fluorescent probes at each labeling sites using donor-only and acceptor-only proteins, in samples of RNAP holoenzyme and RPo. The results (Table 3 for RNAP derivatives in experimental strategy I and Table 4 for RNAP derivatives in experimental strategy II) show that the values of Q_D vary at different labeling positions. Förster parameter R_0 was calculated for each donor-acceptor labeled RNAP derivative from Q_D and donor emission-acceptor excitation spectral overlap integral (J) using equation (2) (Materials and Methods, 3.9). For the same donor-acceptor pair of fluorescent probes (TMR-A647 in Table 3 and Cy3B-A647 in Table 4), the coefficient of variation for R_0 is around 3%, mainly resulting from the differences on Q_D. Comparing values for each individual RNAP derivative from holoenzyme to RPo, values of Q_D, J and R_0 remain essentially unchanged.

Steady-state anisotropies of labeled RNAP were also measured in samples of RNAP holoenzyme and RPo, using donor-only and acceptor-only RNAP derivatives. The results (Table 5) showed that all anisotropies values were below 0.36, low relative to calculated anisotropies of probes linked to molecules of ~0.5 MDa and having restricted local

motion (Cantor and Schimmel, 1980), indicating that probes reorient on the time scale of the probe life times, and validating use of $\kappa^2 = 2/3$ in calculation of Ro. The results also suggested that anisotropies remain relatively similar in RNAP holoenzyme and RPo.

Overall, R_0 and anisotropy values were similar in holoenzyme and RPo, suggesting that binding of DNA did not affect the spectral properties of fluorescently labeled RNAP, and further indicating that any changes in smFRET efficiencies were due to the changes in donor-acceptor distance.

Table 3. **Spectral properties of fluorescent probes in RNAP holoenzyme and RPo**

Probe sites		Holo			RPo		
Core	σR1.1	Q_D	J ($M^{-1}cm^3$)	R_0 (Å)	Q_D	J ($M^{-1}cm^3$)	R_0 (Å)
106	14	0.24	5.8E-13	53	0.26	5.5E-13	53
222	14	0.24	6.5E-13	54	0.26	6.1E-13	54
357	14	0.24	6.5E-13	54	0.26	6.6E-13	54
379	14	0.24	6.2E-13	53	0.26	6.2E-13	54
643	14	0.24	6.9E-13	54	0.26	6.1E-13	54
937	14	0.24	7.9E-13	55	0.26	5.9E-13	53
106	36	0.34	6.1E-13	56	0.34	5.8E-13	56
222	36	0.34	6.8E-13	57	0.34	6.4E-13	57
357	36	0.34	6.7E-13	57	0.34	6.7E-13	57
379	36	0.34	6.4E-13	57	0.34	6.4E-13	57
643	36	0.34	7.2E-13	58	0.34	7.2E-13	58
937	36	0.34	8.2E-13	59	0.34	8.2E-13	59
106	46	0.33	5.9E-13	56	0.29	5.4E-13	54
222	46	0.33	6.5E-13	56	0.29	6.1E-13	55
357	46	0.33	6.5E-13	57	0.29	6.5E-13	55
379	46	0.33	6.2E-13	56	0.29	6.1E-13	55
643	46	0.33	6.9E-13	57	0.29	6.0E-13	55
937	46	0.33	7.9E-13	58	0.29	5.8E-13	54
106	59	0.35	5.8E-13	56	0.35	5.5E-13	56
222	59	0.35	6.4E-13	57	0.35	6.2E-13	57
357	59	0.35	6.5E-13	57	0.35	6.6E-13	57
379	59	0.35	6.1E-13	56	0.35	6.2E-13	57
643	59	0.35	6.9E-13	58	0.35	6.1E-13	56
937	59	0.35	7.8E-13	59	0.35	5.9E-13	56

Spectral properties of fluorescent probes in RNAP holoenzyme and RPo of experimental strategy I. Q_D, quantum yield of donor; J, spectral overlap integral; R_0, Förster radius.

Table 4. Spectral properties of fluorescent probes in RNAP holoenzyme and RPo

Probe sites		Holo			RPo		
σ	σR1.1	Q_D	J ($M^{-1}cm^3$)	R_0 (Å)	Q_D	J ($M^{-1}cm^3$)	R_0 (Å)
118	14	0.46	7.6E-13	61	0.50	8.0E-13	63
304	14	0.44	7.7E-13	61	0.45	8.1E-13	62
366	14	0.43	7.6E-13	61	0.45	8.1E-13	62
396	14	0.59	7.8E-13	64	0.60	8.3E-13	65
459	14	0.37	8.1E-13	60	0.38	8.5E-13	60
569	14	0.44	7.6E-13	61	0.43	7.9E-13	61
118	36	0.46	9.2E-13	63	0.50	8.9E-13	64
304	36	0.44	9.3E-13	63	0.45	9.1E-13	63
366	36	0.43	9.2E-13	63	0.45	9.1E-13	63
396	36	0.59	9.4E-13	66	0.60	9.2E-13	66
459	36	0.37	9.6E-13	62	0.38	9.4E-13	62
569	36	0.44	9.2E-13	63	0.43	8.9E-13	62
118	46	0.46	1.0E-12	65	0.50	1.0E-12	65
304	46	0.44	1.1E-12	64	0.45	1.1E-12	64
366	46	0.43	1.0E-12	64	0.45	1.1E-12	64
396	46	0.59	1.1E-12	68	0.60	1.1E-12	68
459	46	0.37	1.1E-12	63	0.38	1.1E-12	63
569	46	0.44	1.0E-12	64	0.43	1.0E-12	64
118	59	0.46	7.8E-13	62	0.50	7.8E-13	62
304	59	0.44	7.9E-13	62	0.45	7.9E-13	62
366	59	0.43	7.8E-13	61	0.45	7.9E-13	62
396	59	0.59	8.0E-13	64	0.60	8.1E-13	65
459	59	0.37	8.2E-13	60	0.38	8.3E-13	60
569	59	0.44	7.8E-13	61	0.43	7.7E-13	61

Spectral properties of fluorescent probes in RNAP holoenzyme and RPo of experimental strategy I. Q_D, quantum yield of donor; J, spectral overlap integral; R_0, Förster radius.

Table 5. Steady-state anisotropies of fluorescent probes in RNAP derivatives

RNAP	anisotropy		RNAP	anisotropy	
	Holo	RPo		Holo	RPo
σ14TMR-RNAP	0.29	0.25	σ14A647-RNAP	0.34	0.27
σ36TMR-RNAP	0.31	0.28	σ36A647-RNAP	0.36	0.30
σ46TMR-RNAP	0.26	0.25	σ46A647-RNAP	0.36	0.27
σ59TMR-RNAP	0.29	0.22	σ59A647-RNAP	0.35	0.27
β106A647-RNAP	0.28	0.28	σ118Cy3B-RNAP	0.25	0.27
β222A647-RNAP	0.28	0.27	σ304Cy3B-RNAP	0.31	0.29
β357A647-RNAP	0.32	0.30	σ366Cy3B-RNAP	0.26	0.27
β379A647-RNAP	0.30	0.30	σ396Cy3B-RNAP	0.25	0.25
β643A647-RNAP	0.36	0.36	σ459Cy3B-RNAP	0.29	0.29
β937A647-RNAP	0.34	0.33	σ569Cy3B-RNAP	0.31	0.29

Steady-state fluorescence anisotropies were measured for probes in RNAP derivatives from experimental strategy I (left) and experimental strategy II (right), in context of RNAP holoenzyme and RPo.

4.7. Representative smFRET data

For each combination of labeling sites on RNAP core or σR2, σR3, σR4, and labeling sites on σR1.1 (48 in total, experimental strategy I and II), I used confocal optical microscopy with alternating laser excitation (ALEX) to measure the probe-probe smFRET efficiency (*E*)-both in the context of holoenzyme and in the context of RPo. Data collection was carried out in solution with freely diffusing sample at the level of single-molecule. For each measurement, *E* histograms were plotted and fitted with Gaussian curves. The resulting histograms provide equilibrium population distributions of *E*, define numbers of subpopulations with distinguishable *E*, and, for each subpopulation, define mean *E*.

Representative smFRET data (Fig. 15) are illustrated for experimental strategy I with reporter probes on σR1.1 residue 46 and reference probes on RNAP core. For each pair of probes, data measured in the context of holoenzyme were plotted, fitted with Gaussian curves, and compared with data in the context of RPo. For data fitted with a single Gaussian peak, the mean of the peak defines the mean FRET efficiency. About half of the data were fitted with two Gaussian peaks, having a predominant peak with area at least twice of that of the second peak. The second peak is substantially smaller, and might result from inactive RNAP species, a secondary position of σR1.1 or insufficient formation of holoenzyme and RPo. Single-molecule measurements permit identification of subpopulations and assignment of mean *E* values to each subpopulations, therefore avoiding incorrect *E* assignment averaged by ensemble FRET measurements.

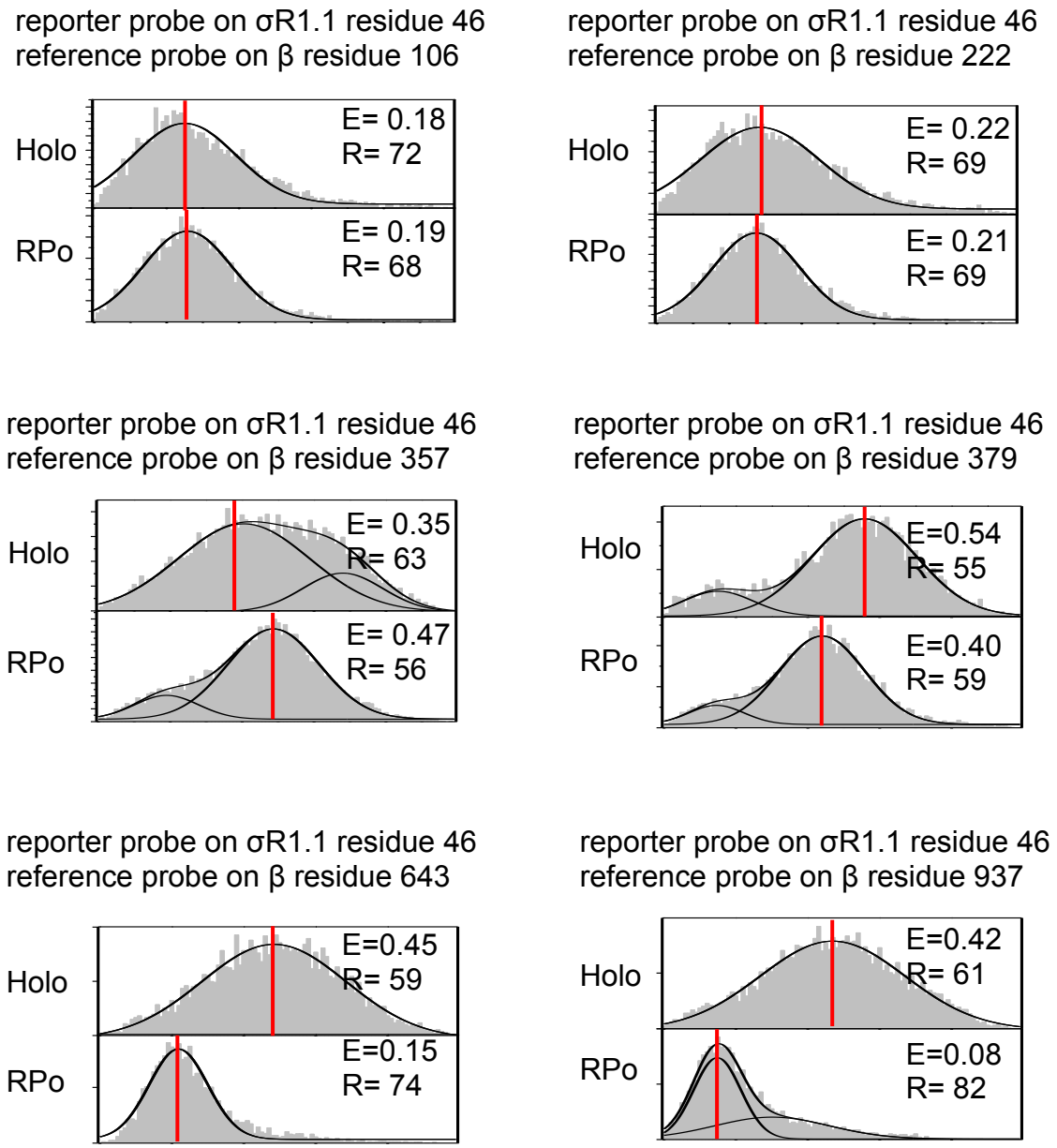


Fig. 15. Representative smFRET data.
Panels show histograms and Gaussian fits of observed probe-probe smFRET efficiencies. For each pair of probes, data for holoenzyme (above) and RPo (below) were compared. The vertical red lines denote mean *E* for the single peaks or the predominant Gaussian-fitted peaks in holoenzyme and RPo. The corresponding mean smFRET *E* values and distances were defined (right corner of the panels).

4.8. Systematic smFRET data

Based on the population distributions of *E*, the data can be classified into three groups (Table 6 for experimental strategy I and Table 7 for experimental strategy II):

I. Data that represent a single Gaussian peak, or data that represent two Gaussian peaks, with the area of the major peak at least twice of the second peak. The mean of the single peak or the major peak define the mean *E* (Fig. 15).

II. Data that represent two Gaussian peaks, with the area of the major peak less than twice of the second peak, or two comparable peaks. The mean values for both peaks are shown in parentheses, with the bold number indicating the major peak.

III. Data that represent broad distributions, with the averaged *E* values shown in square brackets.

The results (Table 6, 7) suggest that group I and II data account for 80% of the total data. Data in parentheses and square brackets were assigned with less reliability in analysis. Measurements were carried out in triplicates for each RNAP derivative. The probe-probe distance (*R*) was calculated from *E* and R_0 using Equation (2). Values of R have precision of $\pm$ 1% ~ 5%.

Table 6. Systematic smFRET data (Experimental strategy I)

Probe sites		Holo			RPo		
β	σR1.1	E	R_0 (Å)	R (Å)	E	R_0 (Å)	R (Å)
106	14	0.22	53	65	0.21	53	66
222	14	(**0.34**, 0.08)	54	(**60**, 81)	0.20	54	67
357	14	NA	54	NA	0.48	54	55
379	14	0.63	53	49	0.43	54	57
643	14	(**0.34**, 0.77)	54	(**60**, 44)	0.13	54	75
937	14	[0.44]	55	[57]	0.06	53	83
106	36	0.14	56	76	0.18	56	72
222	36	0.25	57	68	0.20	57	72
357	36	(**0.70**, 0.44)	57	(**50**, 59)	0.41	57	60
379	36	[0.64]	57	[52]	0.35	57	63
643	36	0.41	58	62	0.09	58	86
937	36	0.45	59	61	0.09	59	87
106	46	0.18	56	72	0.19	54	68
222	46	0.22	56	69	0.21	55	69
357	46	0.35	57	63	0.47	55	56
379	46	0.54	56	55	0.40	55	59
643	46	0.45	57	59	0.15	55	74
937	46	0.42	58	61	0.08	54	82
106	59	0.10	56	80	0.14	56	75
222	59	0.25	57	68	0.22	57	70
357	59	0.48	57	58	0.63	57	52
379	59	0.50	56	56	0.59	57	53
643	59	(**0.49**, 0.04)	58	(**58**, 82)	0.10	56	80
937	59	0.34	59	66	0.09	56	83

Table 7. Systematic smFRET data (Experimental strategy II)

Probe sites		Holo			RPo		
σ	σR1.1	E	R_0 (Å)	R (Å)	E	R_0 (Å)	R (Å)
118	14	0.37	61	67	0.58	63	59
304	14	0.17	61	80	0.34	62	69
366	14	(**0.25**, 0.62)	61	(**73**, 56)	0.29	62	72
396	14	[0.68]	64	[57]	0.15	65	86
459	14	0.50	60	60	0.08	60	90
569	14	(**0.23**,0.63)	61	(**75**, 56)	0.06	61	97
118	36	(0.29,0.63)	63	(73, 57)	(0.51, 0.31)	64	(64,73)
304	36	0.18	63	81	0.28	63	74
366	36	(0.58,0.29)	63	(60,73)	0.30	63	73
396	36	(0.57,0.89)	66	(63,46)	0.18	66	85
459	36	(**0.43**,0.70)	62	(**65**, 54)	0.12	62	86
569	36	0.26	63	75	(**0.21**, 0.05)	62	(**77**, 100)
118	46	(**0.27**,0.65)	65	(**77**, 59)	0.48	65	66
304	46	0.12	64	89	0.37	64	72
366	46	0.23	64	78	0.38	64	70
396	46	0.59	68	64	0.16	68	90
459	46	0.50	63	63	0.10	63	90
569	46	0.42	64	68	0.09	64	95
118	59	0.33	62	70	0.71	62	53
304	59	0.13	62	85	0.43	62	65
366	59	0.28	61	72	0.51	62	61
396	59	[0.63]	64	[58]	0.20	65	81
459	59	0.45	60	62	0.11	60	85
569	59	[0.39]	61	[66]	0.06	61	95

4.9. Distance-restrained docking

Distance restrained docking was performed (in collaboration with Dr. Jennifer Knight, Schrödinger) to generate FRET models of σR1.1 in RNAP holoenzyme and RPo, using smFRET distance restraints (Table 6,7). The structure of RNAP holoenzyme was modeled using *E. coli* RNAP holoenzyme (PDB accession code 4IGC, Murakami, 2013) with an open-clamp conformation as in Chakroborty et al. 2012 and σR1.1 removed. The structure of RPo was modeled by *E.coli* RNAP holoenzyme (Murakami, 2013) with the downstream DNA duplex incorporated from *T.thermophilus* RPo (PDB accession code 4G7H; Zhang et al., 2012) with a closed-clamp conformation and σR1.1 removed. σR1.1 was modeled using residues 35-87 of the NMR structure of *T.maritima* σR1.1 (PDB accession code 2K6X; Schwartz et al., 2008). Probes and linkers were modeled onto the structures and 500-2000 sterically-allowed conformations of probes and linkers were identified for each labeling site.

In each of RNAP RPo and RNAP holoenzyme models, Markov-chain Monte Carlo simulations (Metroplis et al., 1953) employing FRET restraints were performed. Four independent simulations were performed with different starting configurations, with each simulation involving 100,000 steps and generating 100,000 trial configurations. Modeled configurations of σR1.1-RNAP holoenzyme and σR1.1-RPo were assigned a penalty based on their deviation from experimental FRET distances (Materials and Methods, Distance-restrained docking). Specifically, experimental data for which two peaks were observed, the distance associated with the predominant peak was used and assigned 15% uncertainty of observed experimental distance ($\sigma_i = 0.15R_i^{obs}$); where both peaks were

comparable or could not be resolved, the average efficiency was used to compute the apparent distance and assigned with 30% uncertainty of observed experimental distance ($\sigma_i = 0.3R_i^{obs}$).

The simulation trajectories were post-processed using cut-off of σR1.1(Cα atom)-[RNAP(Cα atom) or DNA(P atom)] >2 Å and < 8 Å, so that only accepted configurations were retained that demonstrated some contact between σR1.1 and RNAP, but no severe clashes. The retained configurations were ranked by FRET penalties. Top 100 solutions with lowest FRET penalties were classified into distinct structural groups, based on the translational and rotational orientations of σR1.1 relative to RNAP core. The top solutions from each structural group were identified as representative optimal FRET models.

4.10. FRET models of σR1.1 in RNAP holoenzyme

The top two distinct FRET models (model M and model L) with lowest penalty scores are selected as representative models of σR1.1 in RNAP holoenzyme. They have much lower FRET penalty scores compared with the crystal structure of σR1.1 from Murakami and coworkers (Murakami, 2013; model W, Fig.16C). The segment used in FRET models of σR1.1 contains residue 9 to 59 (*E. coli* numbering, same for below), comprising three α helices (H1, residues 4-19; H2, residues 24-30; H3, residues 38-50) and part of the C-terminal tail (residues 51-59, Schwartz et al., 2008). The remaining part of C-terminal tail is shown in the models, but is not simulated by FRET distance restraints.

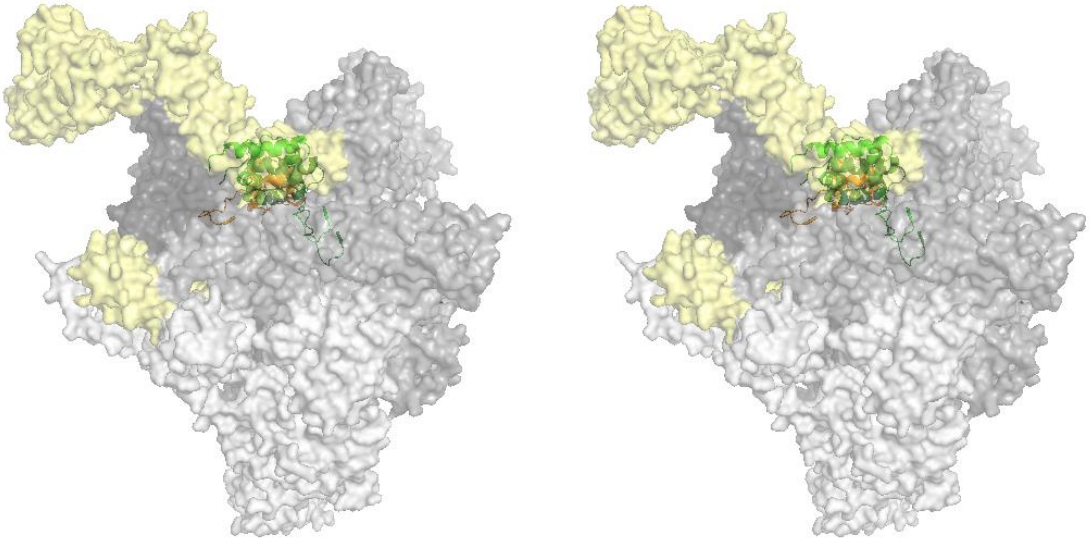
Model M and L are closely related to each other, with similar positions and rotational orientations (Fig.16 A, B). The two models differ in translational position by ~ 6 Å and rotational orientation by ~ 25 degrees. In both models, σR1.1 is located inside the active-center cleft, at similar positions as in the crystal structures of σR1.1 in holoenzyme from Murakami group and Darst group (Murakami 2013; Bae et al., 2013, in press). The orientations of σR1.1 in model M and L are close to that in the crystal structure of σR1.1 from Darst group. The model M is particular close, in terms of positions, rotational orientations and σR1.1-core interactions.

In model M (Fig.17), H1 bridges between β′ clamp and β lobe, with helices H3 and H2 close to β′ clamp and β lobe respectively. The N-terminal segment of H3 and a loop preceding H3 (residues 34-43) interact with part of β′ clamp (residues 118-122, 311-314). Notably, negatively charged residues 34-40 of σR1.1 are in close contact with positively charged residues 311-314 of β′. The C-terminal of H1, the loop preceding H2 and H2

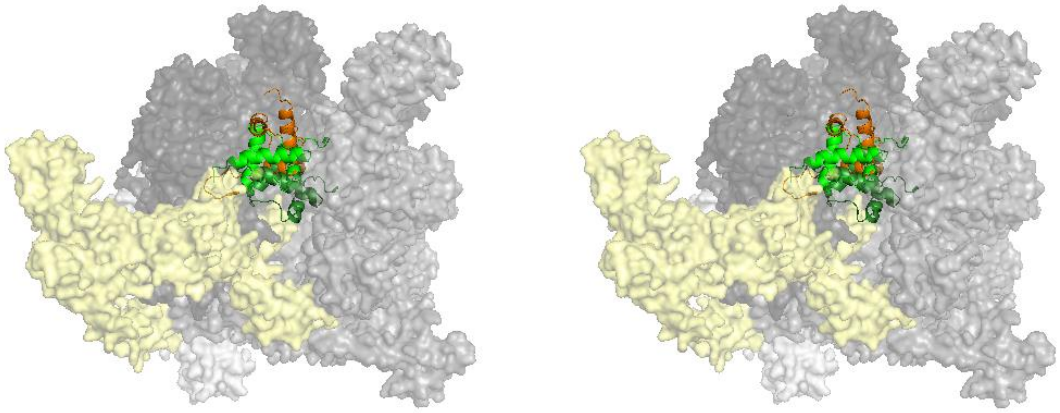
(residues 18-32) interact with the upper part of β lobe (residues 201-204, 365-373). The negatively charged residues on H1 and H2 are in close proximity with positively charged patch (residues 201-204) on β lobe.

In model L (Fig.18), σR1.1 is slightly tilted towards β lobe and is closer to the active-center compared to model M. The N-terminal residues (residue s3, 7) of H1, N-terminal segment (residues 38-40) of H3 and the loop (residues 31, 33-37) connecting H2 and H3 make contacts with part of β′ clamp (residues 310-314, 321-323, 332, 1327, 1329). The C-terminal segment (residues 14, 17-18) of H1, the C-terminal tail (residue 54, 55-59) and the loop (residues 20-22) connecting H1 and H2 interact with β lobe (residues 183, 197-201, 369-372, 510, 532-534).

A



B



C

	W	M	L
Penalty score	86.3	25.8	26.3

Fig. 16. FRET models of σR1.1 in RNAP holoenzyme.
Stereo diagrams are shown for FRET models of σR1.1 (cartoon represention) in RNAP holoenzyme (surface representation).
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A). αNTD^{I}, αNTD^{II} and ω are in light gray; β is in gray; β′ is in dark gray; σR2-4 is in yellow (subunits are shown as semi-transparent in surface representation to permit view of σR1.1 in the active-center cleft). FRET models of σR1.1 (model M in green; model L in dark green) are shown together with starting configuration of σR1.1 (model W in orange, the crystal structure of *E. coli* σR1.1 in holoenzyme from Murakami 2013).
(C) FRET penalty scores for model W, M, and L in RNAP holoenzyme.

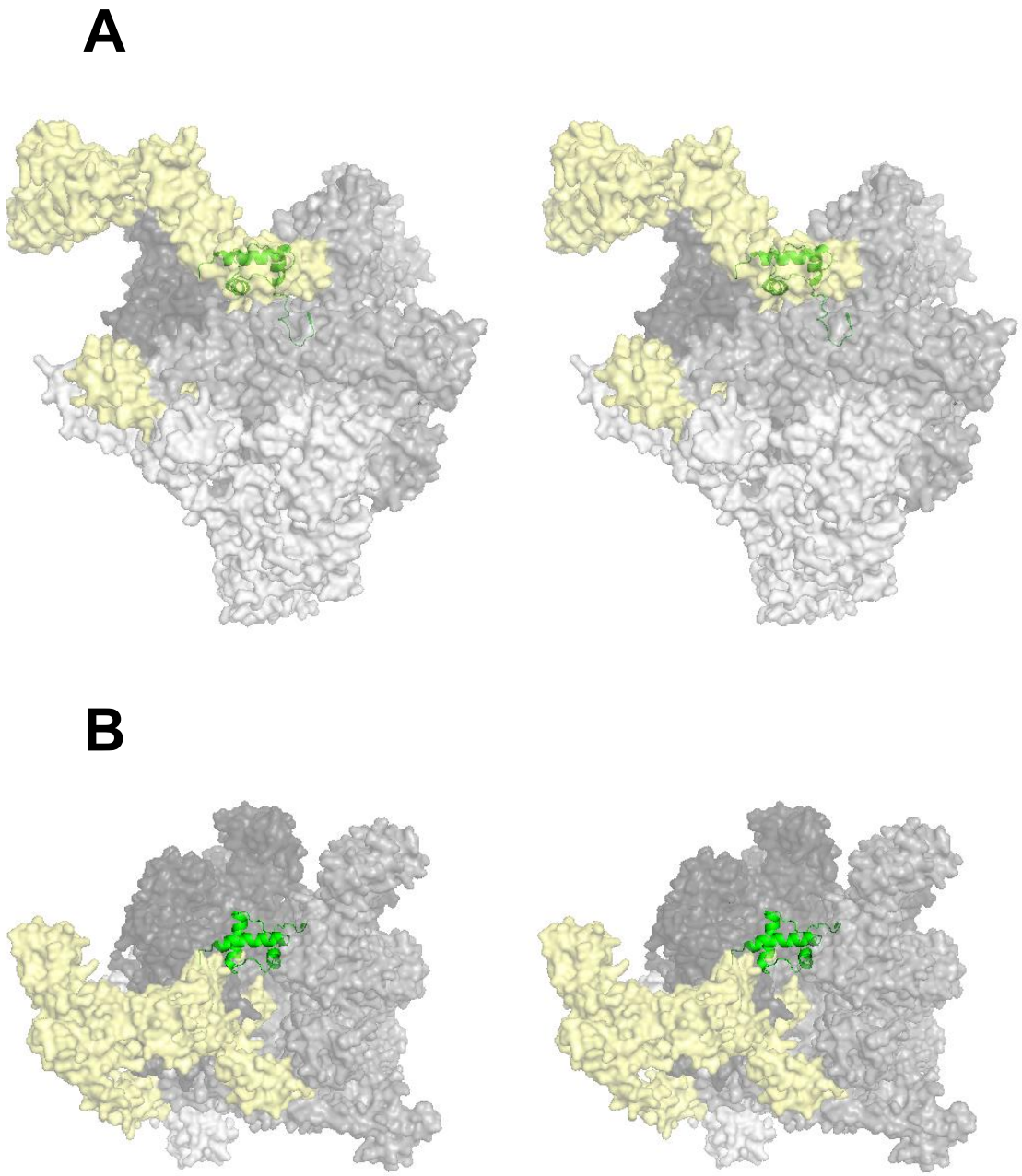


Fig. 17. FRET model M of σR1.1 in RNAP holoenzyme.
Stereo diagrams are shown for FRET model M of σR1.1 (cartoon representation) in RNAP holoenzyme (surface representation). Color scheme is as in Fig.16.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A).

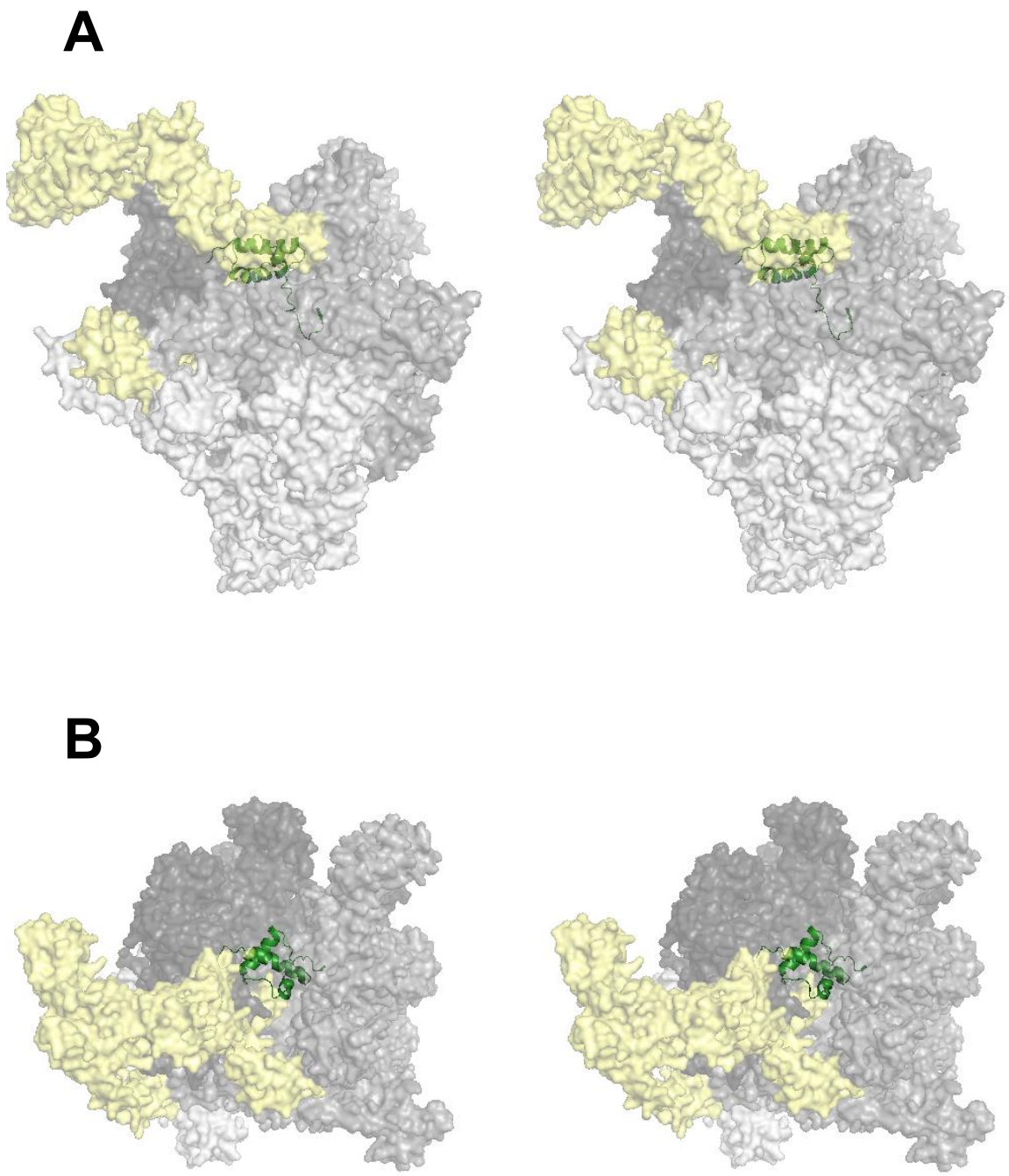


Fig. 18. FRET model L of σR1.1 in RNAP holoenzyme.
Stereo diagrams are shown for FRET model L of σR1.1 (cartoon representation) in RNAP holoenzyme (surface representation). Color scheme is as in Fig.16
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A).

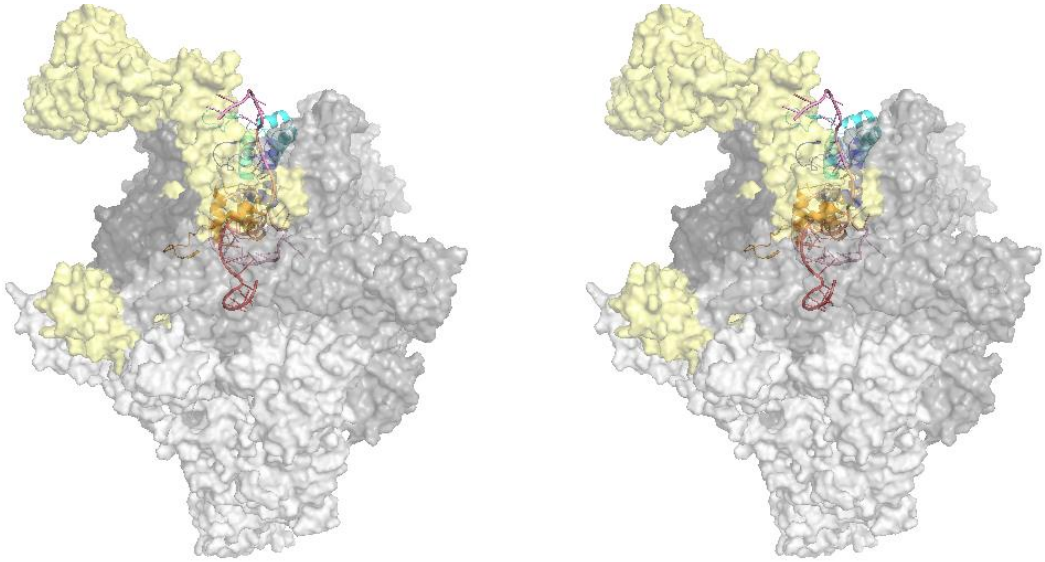
4.11. FRET models of σR1.1 in RPo

The top two distinct FRET models (model N and model K) with lowest penalty scores (Fig.19 C) were selected as representative models of σR1.1 in RPo. Both models clearly demonstrated displacement of σR1.1 from RNA active-center cleft upon formation of RPo. In comparison with model W which sits at a position equivalent to +8 bp of downstream DNA, σR1.1 in model N and model K is ejected > 40 Å away, to a position between β′ jaw and β lobe (Fig.19 A & B).

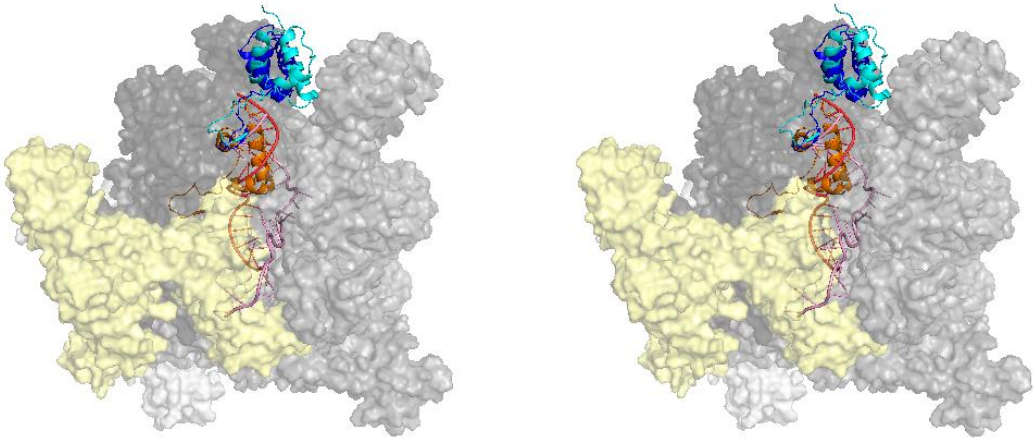
The models have placed σR1.1 in RPo at similar positions but with different orientations. The positions of σR1.1 in model N and model K partially overlap (Fig.19 A, B); model K is tilted towards βDR1 (β dispensable region 1, also referred to as βi4, β insertion 4; Lane and Darst, 2010a, b). In both models, σR1.1 is located above the downstream DNA channel; the N-terminus of σR1.1 points away from the active-center cleft and the C-terminal tail of σR1.1 swings towards the active-center cleft. In model K (Fig.20), the C-terminal segment of H1, part of the loop between H1 and H2 and the C-terminal residue of H2 (residues 15-21, 25-33) make contact with a patch of residues on βDR1 (residues 308-323, especially with residues 310-311, 315, 318, 319-320, 324).

In model N (Fig.21), the contact with β is lost; instead, the segment (residues 42-43, 46) of H3 interacts with a loop on β′ jaw (residues 1171-1172). The C-terminal end of H3 (residue 49) putatively contacts with the +12 nt (dC) of non-template strand.

A



B



C

	W	K	N
Penalty score	134.8	49.2	48.1

Fig. 19. FRET models of σR1.1 in RPo.
Stereo diagrams are shown for FRET models of σR1.1 in RPo. σR1.1 and DNA are in cartoon representation; RNAP subunits are in surface representation.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A). αNTD^{I}, αNTD^{II} and ω are in light gray; β is in gray; β′ is in dark gray; σR2-4 is in yellow (subunits are shown as semi-transparent in surface representation to permit view of σR1.1 in the active-center cleft). DNA template strand is in red; DNA nontemplate strand is in pink. FRET models of σR1.1 (model K in cyan; model N in blue) are shown together with starting configuration of σR1.1 (model W in orange, the crystal structure of *E. coli* σR1.1 in holoenzyme from Murakami 2013).
(C) FRET penalty scores for model W, K, and N in RPo.

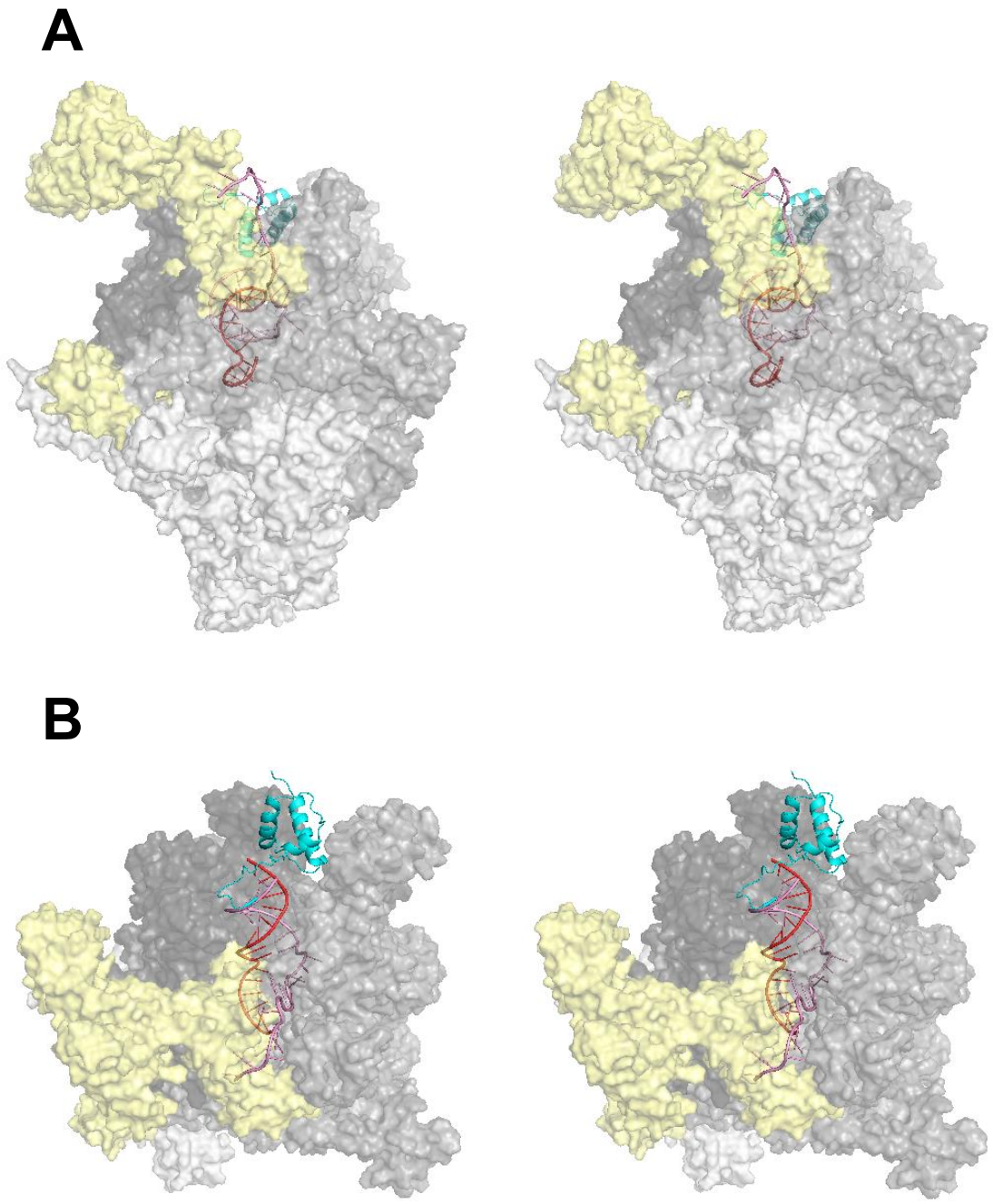


Fig. 20. FRET model K of σR1.1 in RPo.
Stereo diagrams are shown for FRET model K of σR1.1 in RPo. Color scheme is as in Fig.20.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A).

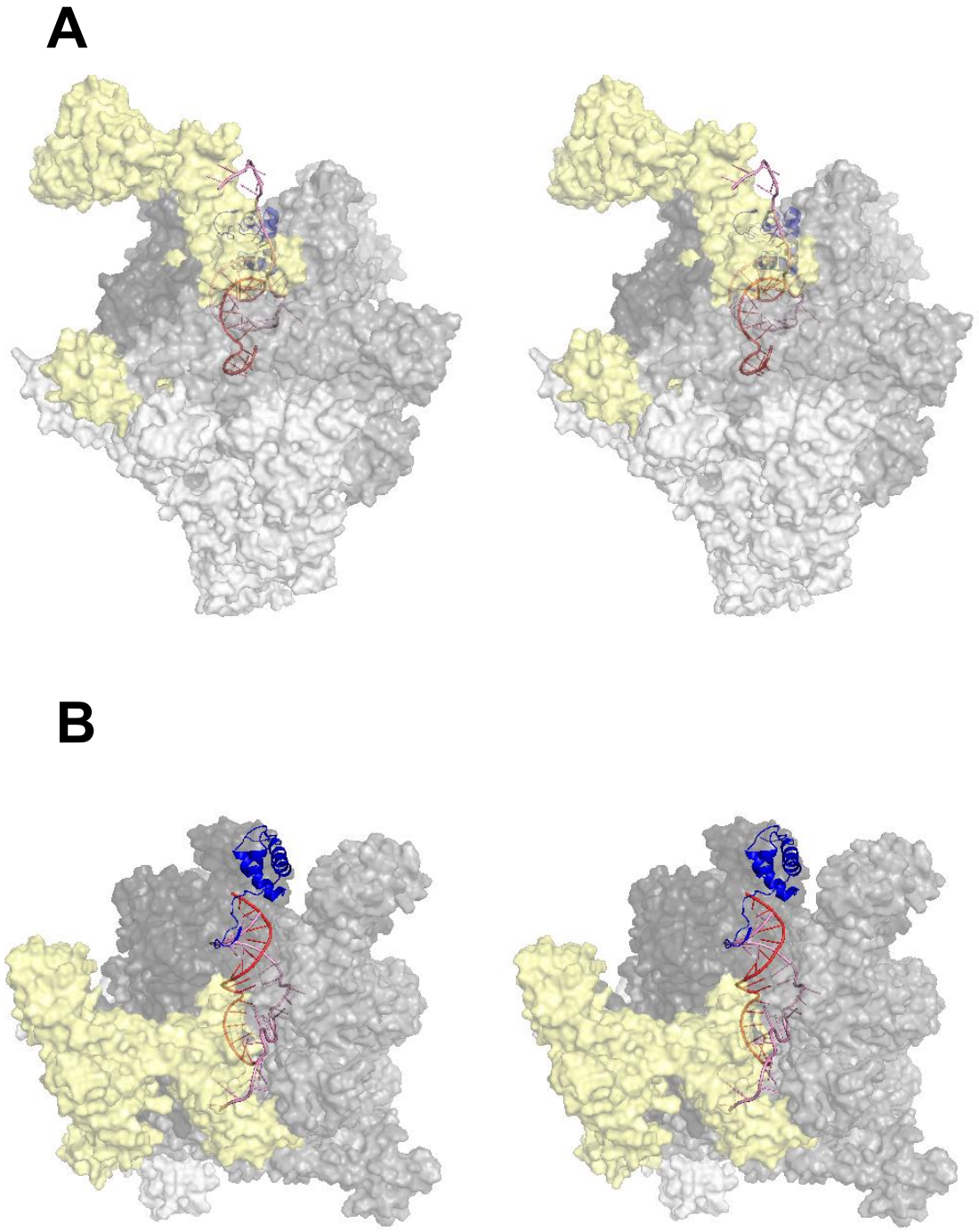


Fig. 21. FRET model N of σR1.1 in RPo.
Stereo diagrams are shown for FRET model N of σR1.1 in RPo. Color scheme is as in Fig.20.
(A) "Upstream" face.
(B) "Top" face (view into active-center cleft; -90 °rotation about x-axis relative to A).

5. Conclusions

The N-terminal conserved region of primary σ factors-σR1.1 was for long regarded as a disordered segment, and was missing from the crystal structures of free σ, RNAP holoenzyme, and RPo (Campbell et al., 2002; Malhotra et al., 1996; Murakami et al., 2002b; Vassylyev et al., 2002; Zhang et al., 2012). Solution structure of *T. maritima* σ^A σR1.1 has been determined by NMR, in which σR1.1 comprises a core folded domain (Fig.4 B) with three α-helices (Schwartz et al., 2008). The folded domain of σR1.1 presents highly negative electrostatic surface potential. Systematic ensemble FRET analysis indicated that in RNAP holoenzyme σR1.1 is located inside the RNAP active-center cleft, serving as "molecular-mimic" of DNA, and must be displaced of the RNAP active-center channel to allow the entry of DNA upon formation of RPo (Mekler et al., 2002). However, the results from ensemble FRET analysis could not provide precise positions or rotational orientations of σR1.1 in RNAP holoenzyme and in RPo.

Recent crystal structures of *E. coli* σ^{70} σR1.1 in RNAP holoenzyme from two different groups—Murakami and co-workers and Darst and co-workers—provided new information on σR1.1 in holoenzyme (Murakami, 2013; Bae et al., 2013, in press). The two groups place σR1.1 inside the RNAP active-center cleft, consistent with ensemble FRET analysis, however, the folds and the rotational orientations of σR1.1 in the crystal structures from the two groups are different. In Murakami's crystal structure, σR1.1 contains four α-helices, and the assignment of each helix is distinct from that in the solution structure of σR1.1 (Schwartz et al., 2008). Darst's crystal structure of σR1.1 comprises a core folded domain of three α-helices, with essentially similar topology to

the solution structure of σR1.1. The rotational orientations of σR1.1 relative to RNAP core are also different in the crystal structures from the two groups, therefore the predicted protein-protein interactions between σR1.1 and RNAP core subunits deviate.

In this work, I have used systematic single-molecule FRET and distance-restrained docking to define the positions and the rotational orientations of σR1.1 in RNAP holoenzyme in solution and in RPo in solution. The most prominent advantage of single-molecule measurements over ensemble measurements is to permit differentiation of subpopulations, so that the FRET efficiency for each subpopulation could be defined instead of being averaged. More probe sites were chosen (4 reporter probe sites on σR1.1, 6 reference probe sites on RNAP core, and 6 reference probe sites on σR2, σR3 and σR4) to allow more accurate mapping of σR1.1 in RNAP holoenzyme and in RPo.

Top solutions for σR1.1 in holoenzyme (Fig. 16) from this work place σR1.1 inside the RNAP active-center cleft, consistent with ensemble FRET analysis and the crystal structures of σR1.1 in RNAP holoenzyme (Mekler et al., 2002; Murakami, 2013; Bae et al., 2013, in press). The top solutions are essentially similar in positions and rotational orientations. My results suggest that the rotational orientation of σR1.1 corresponds to the crystal structure of σR1.1 from Darst group, with similar orientations relative to RNAP core, and same predicted protein-protein interactions between σR1.1 and RNAP core subunits. Back-calculation of FRET penalty for the crystal structure of σR1.1 in RNAP holoenzyme from Murakami group yielded unacceptable high penalty scores.

Top solutions of σR1.1 in RPo (Fig. 19) from this work place σR1.1 outside the RNAP active-center cleft—at least 40 Å away from its position in holoenzyme—between the

RNAP β’ jaw and β dispensable region 1 (βDR1). They have similar positions but different rotational orientations. One of the top models places σR1.1 in direct contact with βDR1, suggesting protein-protein interactions between σR1.1(residues 15-21, 25-33) and βDR1(residues 310-324). Therefore, σR1.1 in RPo in solution is displaced out of the RNAP active-center cleft, to a position between β′ jaw and βDR1, potentially interacting with βDR1.

6. Implications and Prospects

This work provided detailed structural frameworks for σR1.1 in RNAP holoenzyme in solution and in RPo in solution. The FRET models can be further refined (e.g., strengthen the post-processing steric parameters, rank the FRET models by steric penalty) and compared with models simulated from different starting configurations (e.g., different clamp conformation states in holoenzyme, and different clamp conformation states in RPo). The refinement of models might help differentiate the top solutions of σR1.1 with comparable FRET penalty scores.

My results for RPo provided first information on identifying regions of σR1.1 and regions of RNAP core that interact in RPo. It has been reported that deletion of β dispensable region 1 (βDR1) affects the rate and temperature-dependence of formation of RPo and alters the footprint of RPo, with unknown mechanisms (Severinov and Darst, 1997). My results suggest that, upon formation of RPo, σR1.1 is displaced from the RNAP active-center cleft to a binding site on βDR1, and, as such, provide an explanation for effects of deleting βDR1 on the rate and temperature-dependence of formation of RPo and on the footprint of RPo. Single-molecule FRET analysis with RNAP derivatives containing deletion or mutations of βDR1 near the binding site of σR1.1 in RPo can be employed to further assess the interactions between σR1.1 and βDR1.

Single-molecule FRET data in this work suggested structural heterogeneity exist in samples of RNAP holoenzyme and RPo. The heterogeneity indicates that, σR1.1 in holoenzyme and RPo could sample different positions over the detection time, and that might open the door for DNA to enter the RNAP active-center cleft. To analyze the

positional dynamics of σR1.1 in RNAP holoenzyme and RPo, single-molecule analysis can be employed over a function of time using immobilized RNAP derivatives (Roy et al., 2008). Other statistical methods, e.g., burst variance analysis, can also be used to determine whether the heterogeneity arises from dynamic processes or from the coexistence of several static structures (Torella et al., 2011).

The efforts of mapping the positions and the rotational orientations of σR1.1 relative to RNAP core can be further extended to various conformational states of RNAP at each step of the transcription pathway. Using the same set of techniques and samples prepared in this work, the positions and the rotational orientations of σR1.1 could be defined in the context of RNAP-promoter initial transcribing complex (RPitc) with 2-8 nt of RNA and in the context of RNAP-DNA elongation complex (RDe) (Vassylyev et al., 2007; Zhang et al., 2012).

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.

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.

Bae, B., Davis, E., Brown, D., Campbell, E.A., Wigneshweraraj, S., Darst, S.A. (2013). Phage T7 gp2 inhibits Escherichia coli RNA polymerase by misappropriation of σ70 domain 1.1. (in press)

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994). Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell *78*, 889-896.

Borukhov, S., and Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. Curr Opin Microbiol *6*, 93-100.

Bown, J.A., Owens, J.T., Meares, C.F., Fujita, N., Ishihama, A., Busby, S.J., and Minchin, S.D. (1999). Organization of open complexes at Escherichia coli promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem *274*, 2263-2270.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat Rev Microbiol *2*, 57-65.

Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y., and Gralla, J.D. (2000). The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J Bacteriol *182*, 4129-4136.

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

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., Heyduk, E., and Heyduk, T. (1999). Core RNA polymerase from E. coli induces a major change in the domain arrangement of the sigma 70 subunit. Mol Cell *3*, 229-238.

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.

Camarero, J.A., Shekhtman, A., Campbell, E.A., Chlenov, M., Gruber, T.M., Bryant, D.A., Darst, S.A., Cowburn, D., and Muir, T.W. (2002). Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. Proc Natl Acad Sci U S A *99*, 8536-8541.

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.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol *11*, 121-127.

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. J Am Chem Soc *124*, 9026-9027.

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

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

Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr Opin Genet Dev *14*, 218-226.

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.

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.

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.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol *57*, 441-466.

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.

Hawley, D.K., and McClure, W.R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res *11*, 2237-2255.

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

Hinton, D.M., Vuthoori, S., and Mulamba, R. (2006). The bacteriophage T4 inhibitor and coactivator AsiA inhibits Escherichia coli RNA Polymerase more rapidly in the absence of sigma70 region 1.1: evidence that region 1.1 stabilizes the interaction between sigma70 and core. J Bacteriol *188*, 1279-1285.

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.

James, E., Liu, M., Sheppard, C., Mekler, V., Camara, B., Liu, B., Simpson, P., Cota, E., Severinov, K., Matthews, S., *et al.* (2012). Structural and mechanistic basis for the inhibition of Escherichia coli RNA polymerase by T7 Gp2. Mol Cell *47*, 755-766.

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.

Keilty, S., and Rosenberg, M. (1987). Constitutive function of a positively regulated promoter reveals new sequences essential for activity. J Biol Chem *262*, 6389-6395.

Kim, Y., Ho, S.O., Gassman, N.R., Korlann, Y., Landorf, E.V., Collart, F.R., and Weiss, S. (2008). Efficient site-specific labeling of proteins via cysteines. Bioconjug Chem *19*, 786-791.

Lane, W.J., and Darst, S.A. (2010a). Molecular evolution of multisubunit RNA polymerases: sequence analysis. J Mol Biol *395*, 671-685.

Lane, W.J., and Darst, S.A. (2010b). Molecular evolution of multisubunit RNA polymerases: structural analysis. J Mol Biol *395*, 686-704.

Lee, D.J., Minchin, S.D., and Busby, S.J. (2012). Activating transcription in bacteria. Annu Rev Microbiol *66*, 125-152.

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.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol *174*, 3843-3849.

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.

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.

Mujumdar, R.B., Ernst, L.A., Mujumdar, S.R., Lewis, C.J., and Waggoner, A.S. (1993). Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. Bioconjug Chem *4*, 105-111.

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.

Murakami, K.S. (2013). The X-ray Crystal Structure of Escherichia Coli RNA Polymerase Sigma70 Holoenzyme. J Biol Chem.

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.

Nagai, H., and Shimamoto, N. (1997). Regions of the Escherichia coli primary sigma factor sigma70 that are involved in interaction with RNA polymerase core enzyme. Genes Cells *2*, 725-734.

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.

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

Opalka, N., Brown, J., Lane, W.J., Twist, K.A., Landick, R., Asturias, F.J., and Darst, S.A. (2010). Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. PLoS Biol *8*.

Owens, J.T., Miyake, R., Murakami, K., Chmura, A.J., Fujita, N., Ishihama, A., and Meares, C.F. (1998). Mapping the sigma70 subunit contact sites on Escherichia coli RNA polymerase with a sigma70-conjugated chemical protease. Proc Natl Acad Sci U S A *95*, 6021-6026.

Paget, M.S., and Helmann, J.D. (2003). The sigma70 family of sigma factors. Genome Biol *4*, 203.

Perdue, S.A., and Roberts, J.W. (2011). Sigma(70)-dependent transcription pausing in Escherichia coli. J Mol Biol *412*, 782-792.

Peters, J.M., Vangeloff, A.D., and Landick, R. (2011). Bacterial transcription terminators: the RNA 3'-end chronicles. J Mol Biol *412*, 793-813.

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.

Roberts, C.W., and Roberts, J.W. (1996). Base-specific recognition of the nontemplate strand of promoter DNA by E. coli RNA polymerase. Cell *86*, 495-501.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nat Methods *5*, 507-516.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol *412*, 754-771.

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.

Schwartz, E.C., Shekhtman, A., Dutta, K., Pratt, M.R., Cowburn, D., Darst, S., and Muir, T.W. (2008). A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. Chem Biol *15*, 1091-1103.

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

Severinov, K., and Darst, S.A. (1997). A mutant RNA polymerase that forms unusual open promoter complexes. Proc Natl Acad Sci U S A *94*, 13481-13486.

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.

Sheppard, C., Camara, B., Shadrin, A., Akulenko, N., Liu, M., Baldwin, G., Severinov, K., Cota, E., Matthews, S., and Wigneshweraraj, S.R. (2011). Inhibition of Escherichia coli RNAp by T7 Gp2 protein: role of negatively charged strip of amino acid residues in Gp2. J Mol Biol *407*, 623-632.

Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980). E. coli RNA polymerase interacts homologously with two different promoters. Cell *20*, 269-281.

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.

Torella, J.P., Holden, S.J., Santoso, Y., Hohlbein, J., and Kapanidis, A.N. (2011). Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. Biophys J *100*, 1568-1577.

Vassylyev, D.G. (2009). Elongation by RNA polymerase: a race through roadblocks. Curr Opin Struct Biol *19*, 691-700.

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., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. Nature *448*, 157-162.

Vuthoori, S., Bowers, C.W., McCracken, A., Dombroski, A.J., and Hinton, D.M. (2001). Domain 1.1 of the sigma(70) subunit of Escherichia coli RNA polymerase modulates the formation of stable polymerase/promoter complexes. J Mol Biol *309*, 561-572.

Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the genetic code of Escherichia coli. Science *292*, 498-500.

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.

Wheeler, A.R., Woody, A.Y., and Woody, R.W. (1987). Salt-dependent binding of Escherichia coli RNA polymerase to DNA and specific transcription by the core enzyme and holoenzyme. Biochemistry *26*, 3322-3330.

Wilson, C., and Dombroski, A.J. (1997). Region 1 of sigma70 is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase. J Mol Biol *267*, 60-74.

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

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.