CLOSING AND OPENING OF THE RNA POLYMERASE "TRIGGER LOOP" IN SOLUTION

by

MIAOXIN LIN

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Richard H. Ebright

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2017

ABSTRACT OF THE DISSERTATION

Closing and Opening of the RNA Polymerase “Trigger Loop” in Solution

by MIAOXIN LIN

Dissertation Director:

Richard H. Ebright

Previous crystal structures of transcription elongation complexes (TEC) suggested that each nucleotide-addition cycle is coupled to a closing/opening conformational cycle of the RNA polymerase (RNAP) active center, accomplished by folding (closing) of a structural element termed "trigger loop" (TL) upon binding of the incoming NTP, followed by unfolding (opening) of the RNAP TL upon formation of the phosphodiester bond. However, direct evidence for TL closing and opening in solution is lacking.

In this work, we have directly detected TL closing/opening in solution by measuring ensemble fluorescence resonance energy transfer (FRET) between two fluorescent probes in a labelled TEC without and with a bound NTP. As the FRET donor, a coumarin-derived fluorescent amino acid L-(7-hydroxycoumarin-4-yl)-ethylglycine (Cou) was incorporated by unnatural amino acid mutagenesis into TL. As the FRET acceptor, fluorescein was chemically coupled to each of eleven reference sites in DNA template strand. FRET measurements showed that in solution TL is open in the absence of bound NTP and closes after NTP binding.

In addition, we have analyzed the effects of substrate identities on TL conformation in solution by measuring ensemble FRET of labelled TEC with one of a series of NTPs or NTP analogues. The results indicated that: (i) TL closing depends on base complementarity; (ii) TL closing depends on tri-/di- /mono-phosphate identity; (iii) TL closing partially depends on 2'-ribose/deoxyribose identity, but does not depend on 3'-ribose/deoxyribose identity.

Moreover, we have mapped the positions of β'942 in open and closed TL in solution by combining systematic FRET and distance-restrained docking. Top-ranked model solutions placed the representative residue of TL in the secondary channel with a significant difference in position between the open and closed states. Modeled positions change from open state to closed state in a similar manner as suggested by crystal structures. Modeled position of β'942 in open TL is not compatible with the SI3 orientation accommodating closed TL, which suggested that the SI3 module varies in at least two orientational states in response to the conformational cycling of TL in each nucleotide addition step.

Besides, we have developed four procedures to solve specific problems encountered in this work and they may have further applications.

To improve the yield of Cou-labelled β'subunit, a procedure of incorporating Cou into β' with shortened distance from translation start by using split β' was developed. This procedure can probably improve the yield of β' or β (largest and most commonly used RNAP subunits in unnatural amino acid mutagenesis) mutants incorporated with other unnatural amino acids whose incorporation efficiencies have been found low.

To improve the transcriptional activity of Cou-labelled RNAP, a procedure of mutating the amino acid residues adjacent to Cou was developed. This procedure provides an option of introducing an unnatural amino acid into protein interior with minimized loss of protein functions.

To improve the monodispersity of Cou-labelled RNAP, a procedure of one-pot *in vivo* assembly for Cou-labelled RNAP was developed. This procedure can be easily applied to preparation of RNAP incorporated with other unnatural amino acids, and that can serve as an alternative method to RNAP reconstitution but with a higher protein yield, a higher enzyme activity, and an easier protocol.

To eliminate fluorescence background from non-functional or excess components, an electroelution-based purification procedure for elongation complexes was developed. This procedure can be readily applied to purification of other RNAP-promoter complexes for fluorescence, or non-fluorescence experiments simply by using a fluorescently labelled RNAP complex as marker.

ACKNOWLEDGEMENT

I would like thank my advisor, Dr. Richard Ebright, for his guidance and support in this work. He has been always steering me in the right direction when roadblocks were encountered in this work. His knowledgeability in his academic field, enthusiasm about science, and passion for his work has a far-reaching impact on me.

I would like to thank my committee members, Dr. Helen Berman, Dr. Wilma Olson and Dr. Bryce Nickels for their time and review on the work.

I would like to thank Dr. Claus Seidel and Dr. Mykola Dimura for providing the FPS program and their advice. I would like to thank Dr. Jennifer Knight, Dr. Jens Michaelis and Dr. Tobias Eilert for their advice and help in FRET-retrained positioning.

I would like to thank Dr. Yon Ebright for her guidance in organic synthesis and purification, as well as her help and encouragement in my academic and personal life.

I would like to thank Dr. Abhishek Mazumder for his enlightening discussions and suggestions.

I would like to thank each of the Ebright lab members both past and present who I have interacted with, especially Dr. Yu Feng, Dr. Juan Shen, Dr. Yu Zhang, Dr. Qumiao Xu, Ms. Carol Shuang Liu, Dr. David Degen, Dr. Libing Yu, Dr. Wei Lin, Dr. Anirban Chakraborty and Mr. Adam Hasemeyer for their help and advice.

I would like to thank my husband, Weiming Du, for everything he did for me. He is always there when I need help and comfort. I would like to express my thanks to my parents for their forever love and support, and to my lovely son. I could not have finished this work without them.

DEDICATION

Dedicated to my husband Weiming

Table of contents

List of figures

List of tables

1. Introduction

Transcription is the first step and a highly regulated step in gene expression. In transcription, gene coding for proteins is transcribed from DNA to RNA. The enzyme responsible for transcription is RNA polymerase (RNAP), which is the direct or indirect target of most regulations in transcription. Bacterial RNAP is the simplest member of the multisubunit RNAP family involved in bacteria, archea and eukarya. In contrast with other members of the protein family which contain additional subunits, bacterial RNAP comprises only the five subunits conserved through the protein family (β', β, α^{I}, α^{II}, and ω in bacterial RNAP; A, B, D, L, and K in archaeal RNAP; RPB1, RPB2, RPB3, RPB11, and RPB6 in eukaryotic RNAP II) (Ebright, et al., 2000; Cramer, et al., 2002, Werner, 2013). Bacterial RNAP has been studied for more than fifty years since its discovery in early 1960s (Hurwitz, 2005) and is the best characterized member of this protein family. Therefore, bacterial RNAP is the model system of choice for detailed structural and mechanistic studies.

1.1. Bacterial RNAP

Bacterial RNAP core enzyme, which consists of the five conserved subunits (β', β, α^{I}, α^{II}, and ω), is a large protein with molecular weight of ~400 KDa and dimensions of ~100 Å × ~100 Å × ~150 Å. It has a shape reminiscent of a crab claw (Zhang, et al., 1999; Ebright, et al., 2000; Darst, et al., 2002). The two pincers of the claw define the active-center cleft (main channel in RNAP). The active center cleft is ~40 Å in depth from the

protein surface to the active site, and ~25 Å in diameter in open state, large enough to allow double-stranded promoter DNA to be loaded into and unwound in the active center during initial transcription (Darst, et al., 2002). Nucleotide addition reaction occurs at the base of the active center cleft through a mechanism involved two Mg^{2+} ions, termed Mg^{2+} (I) and Mg^{2+} (II). Mg^{2+} (I) is tightly coordinated (K_d = ~ 100 μM) and thus retained in the active site, while Mg^{2+} (II) is recruited *ad hoc* for each cycle of nucleotide addition due to its lower affinity (K_d > 10 mM) (Sosunov, et al., 2003). Mg^{2+} (II) is likely to enter the active center along with the incoming NTP. In the active site, there are two conserved structural features: a long α-helical segment termed bridge helix (BH), and a flexible segment termed trigger loop (TL). They are believed to play key roles in RNAP catalysis and translocation through their conformational changes (Vassylyev, et al., 2007b; Zhang, et al., 2010; Epshtein, et al., 2002; Bar-Nahum, et al., 2005).

In addition to the active center cleft, RNAP has two other channels: the secondary channel and the RNA exit channel. The secondary channel is separated from the main channel by bridge helix in the active center, and connects the active center to the protein surface. It is the presumed entryway for nucleoside triphosphates (NTPs) to reach the catalytic site since the active center cleft is occupied by the loaded DNA (Zhang et al., 1999; Cramer et al., 2000). The RNA exit channel, as suggested by its name, is where the 5’ end of RNA transcript is mediated to leave the active center.

The largest subunit β' (1,407 amino acids in *E. coli*) of bacterial RNAP core makes up one of the two pincers and part of the base of the active-center cleft. The β' pincer, termed “clamp”, predominantly adapts an open conformation in free RNAP and early

intermediates in transcription initiation but changes to a closed conformation upon formation of initial transcribing complex and remains closed during transcription initiation and elongation (Chakraborty, et al., 2012). The second largest subunit β (1,342 amino acids in *E. coli*) of RNAP core, makes up the other pincer and part of the base of the active-center cleft. Bacterial RNAP contains two copies of α subunits (α^{I} and α^{II}), which are located distal to the active-center cleft. The α^{I} subunit is closer to the active-center cleft and interacts with β; the α^{II} subunit is located farther from the active-center cleft and interacts with β'. Each α subunit (329 amino acids in *E. coli*) consists of two independent domains, the N-terminal domain (NTD) and the C-terminal domain (CTD), connected by an unstructured flexible linker (Darst, et al., 1998). The α subunits initiate the assembly of RNAP by forming the αNTD dimer with which the β' and β subunits interact. The αCTD is important for the interaction with upstream promoter elements (Ross, et al., 1993) and is the target of some transcription regulators (Hudson, et al., 2009). The smallest subunit ω (91 amino acids in *E. coli*) of RNAP core is located distal to the active-center cleft. It is not essential in RNAP functions, but it helps stabilize the assembly of RNAP by "latching" the N-terminus and C-terminus of β'(Minakhin et al., 2001).

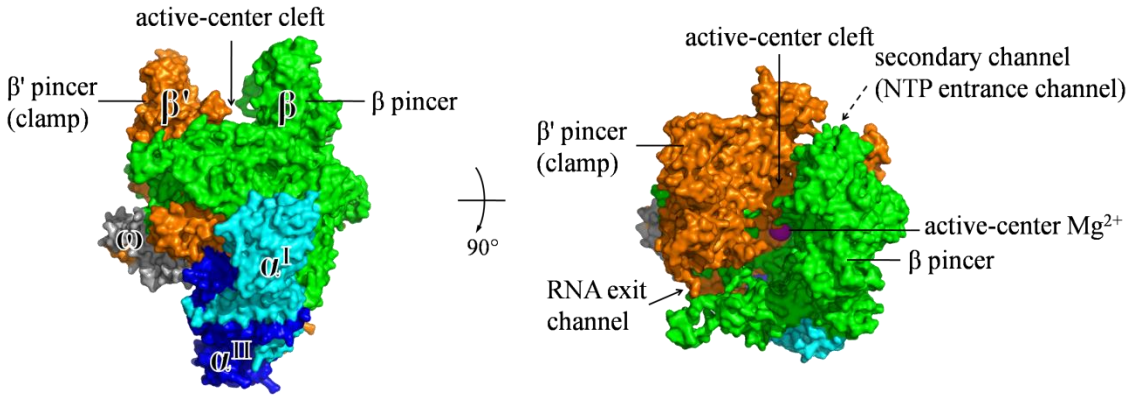


Figure 1. RNAP core. Crystal structure of bacterial RNAP core (Zhang, et al., 1999) is shown in two orthogonal views, upstream face (left) and top face (right), as a molecular surface representation (β' is in orange; β is in green; α^{I} is in cyan; α^{II} is in dark blue; ω is in gray; the active-center Mg^{2+} is in magentas.)

RNAP core enzyme is able to carry out non-specific, promoter-independent transcription initiation as well as all other steps of transcriptions, but is unable by itself to carry out promoter-specific transcription initiation. In order to carry out promoter-specific transcription initiation, RNAP core needs to associate with the initiation factor σ to form the RNAP holoenzyme (Darst, et al., 2001; Gross, et al., 1998). σ contains determinants for sequence-specific interaction with DNA and thus is capable of directing RNAP holoenzyme to promoters. However, σ by itself is incapable of recognizing promoter; binding to RNAP core unmasks the determinants in σ and creates a functional holoenzyme (Ruff, et al., 2015). Besides its role in promoter recognition, σ is required in promoter unwinding, promoter escape, early elongation, and interactions with some transcriptional regulators. The principle σ factor in *E. coli* is σ70 (termed the "housekeeping σ"), which contains five conserved regions: σR1, σR2, σR3, σR3/ σR4 linker, and σR4 (Murakami, et al., 2003). These regions all interact with RNAP core when forming the holoenzyme. σR2, σR3, and σR4 are independently folded domains

that contain determinants for sequence-specific interactions with the promoter -10 element, the promoter extended -10 element, and the promoter -35 element, respectively (Severinova, et al., 1996; Malhotra, et al., 1996; Haugen, et al., 2008; Ruff, et al., 2015). σR1 and σR3/ σR4 linker ("σ finger") are unstructured, negatively charged segments in the active-center cleft and in the RNA exit channel, respectively. σR1 was proposed to serve as a DNA mimic (Ruff, et al., 2015), and σR3/ σR4 linker was believed to function in 3'-initiating nucleotide binding and promoter escape (Kulbachinskiy, et al., 2006).

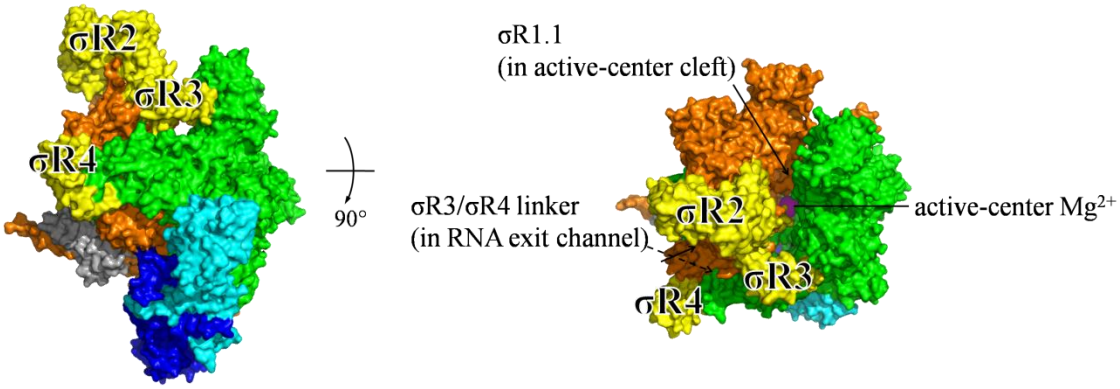


Figure 2. RNAP holoenzyme. Crystal structure of bacterial RNAP holoenzyme (Vassylyev, et al., 2002) is shown in two orthogonal views, upstream face (left) and top face (right), as a molecular surface representation (β' is in orange; β is in green; α^{I} is in cyan; α^{II} is in dark blue; ω is in gray; σ^{70} is in yellow; the active-center Mg^{2+} is in magentas.)

1.2. Transcription elongation

Transcription can be divided into three structurally and functionally distinct stages: initiation, elongation and termination. In transcription initiation and elongation, RNAP and promoter DNA experience a series of interactions (Ebright, 2000; Young, et al., 2002;

Ruff, et al., 2015;). First, RNAP binds to DNA and recognizes a promoter, forming an RNAP-promoter closed complex (RP_c). Second, RNAP unwinds DNA for ~13 base pairs near the transcription start site (from the -10 element to just beyond the start site, termed "transcription bubble") (Saecker, et al., 2011), forming an RNAP-promoter open complex (RP_o). Third, RNAP starts synthesizing RNA as an RNAP-promoter initial transcribing complex (RP_{itc}) using nucleoside triphosphates (NTPs). In this stage, RNAP moves relative to DNA using a "scrunching" mechanism, in which RNAP remains stationary on promoter DNA and unwinds and pulls downstream DNA into itself and past its active center in each nucleotide-addition cycle (Kapanidis, et al., 2006; Revyakin, et al., 2006). During initial transcription, RNAP can engage in off-pathway abortive cycles of synthesis and release of short RNA products. Once RNAP has synthesized an RNA product of a threshold length (typically 9-11 nucleotides at most promoters), it escapes the promoter, weakens its interactions with σ and begins transcription elongation as a transcription elongation complex (TEC; also referred as RDe) using the energy stored in the stressed intermediate generated by scrunching during initial transcription (Mukhopadhyay, et al., 2001). In this stages, RNAP moves relative to DNA with a "stepping" mechanism, in which RNAP advances along DNA without scrunching of DNA in each nucleotide-addition cycle (Abbondanzieri, et al., 2005). During transcription elongation, RNAP can enter into and return from off-pathway pausing states.

The structure of the *Thermus thermophilus* TEC was determined at 2.5Å and provided the first detailed view of the bacterial EC (Vassylyev, et al., 2007a). The complex consists of RNAP core and a synthetic nucleic acid scaffold. It was found that the arrangement of DNA and RNA in the complex has a high similarity with that of eukaryotic TECs

(Kettenberger, et al., 2004; Wang, et al., 2006), which suggests a conserved mechanism of transcriptional elongation throughout evolution. In the structure, the loaded DNA blocks the RNAP active center cleft. The 9-bp RNA/DNA hybrid resides in the base of active center cleft, and the 5'-end of nascent RNA transcript, which is displaced from the DNA template strand, is threaded through the RNA exit channel. Only the secondary channel is widely open (~15 Å × ~22 Å at the entrance to the active center), serving as the major entry route to the RNAP active center for substrate NTPs.

Transcription elongation complexes are highly stable and processive. Compared with initial transcription complexes, transcription elongation complexes are less various in protein content and complexity. The core enzyme that performs elongation is homologous in sequence and structure in prokaryotic and eukaryotic multisubunit RNAPs. In contrast with initial transcription complexes, which are characterized by abortive cycles of synthesis and release of short RNA products, TECs are able to carry out uninterrupted RNA synthesis of up to thousands of nucleotides.

1.3. Nucleotide addition cycle

RNA chain extension is accomplished by many cycles of incorporating a templated nucleotide to the nascent RNA product. Each nucleotide-addition cycle during transcription initiation and transcription elongation includes four steps (Erie, et al., 1992):

(i) translocation of the RNAP active center relative to DNA, (ii) binding of the incoming NTP, (iii) formation of the phosphodiester bond, and (iv) release of pyrophosphate.

Before NTP loading, the initial transcribing complex or elongation complex exists in equilibrium between pre-translocated and post-translocated states, in which the interconversion is driven through a Brownian ratchet mechanism. In pre-translocated state, the 3' end of the growing RNA chain occupies the nucleotide addition site (A site, also referred to as i+1 site), which is the state when a nucleotide has been just incorporated to the RNA transcript. To add a new nucleotide, RNAP must moves forward by one nucleotide so that the RNA 3' end transfers to the product site (P site, also referred to as i site) and the addition site is available for the incoming nucleotide. The process is termed translocation, and the resulting state is called the post-translocated state. NTP binding favors the post-translocated state and bias the complex towards forward translocation (Abbondanzieri, et al., 2005).

Substrate loading is thought to be a two-step process: first, incoming NTP initially binds in a catalytically incompetent conformation (pre-insertion state) to the complex with an open active center; second, bound NTP isomerizes into a catalytically competent conformation (insertion state), inducing closure of the active center. Two additional crystal structures of the *T. thermophilus* TECs with the non-hydrolyzable substrate analog or with the substrate analog and inhibitor streptolydigin revealed the inactive (pre-insertion) and active (insertion) substrate intermediates and provided an evidence for the two-step mechanism of substrate loading in bacterial RNAP (Vassylyev, et al., 2007b). The structures showed that isomerization from the inactive, open (pre-insertion) state to

the catalytically active closed (insertion) state occurs via substrate-induced folding of the trigger loop. The inhibitor streptolydigin prevents proper folding of the TL thereby freezing the substrate complex in the inactive, pre-insertion configuration.

Phosphodiester-bond formation proceeds by S_N2-type nucleophilic attack of the RNA 3' OH group on the NTP α-phosphorus atom. Consequently, the NMP is covalently attached to the RNA 3' end, extending the RNA 3' end by one nucleotide and generating pyrophosphate (PPi). It is generally believed that the active site re-opens concurrently with or immediately after the release of PPi. Structures of T7 RNAP elongation complexes with and without PPi were found to have an open and closed conformation of the active site, respectively, which correspond to the pre-translocated and post-translocated states (Yin, et al., 2004). Translocation occurring shortly after or concurrently with PPi dissociation was also observed in multisubunit RNAP (Malinen, et al., 2012). In contrast, a recent structural analysis of σ^S-transcription initiation complex after PPi release reveals a closed active site with only early signs of opening (Liu, et al. 2016), suggesting that re-opening of the active site may occur later after a series of conformational changes induced by PPi release. Supporting this structural observation, study based on all-atom molecular dynamics simulations also proposes that PPi release precedes full opening of the TL due to faster PPi dynamics, although PPi facilitates the TL opening (Da, et al., 2012).

1.4. Closing and opening of TL

Based on crystal structures, it has been proposed that each nucleotide addition cycle is coupled to an RNAP active-center conformational cycle, comprising closing of the RNAP active center upon substrate loading, followed by opening of the RNAP active center upon or after release of the pyrophosphate. The closing and opening of the RNAP active-center is accomplished by the folding and unfolding, respectively, of an RNAP active-center structural element termed the "trigger loop" (TL, β'1236-1254 in *T. thermophilus* RNAP, β' 926-944 and 1133-1137 in *E. coli* RNAP). Therefore, the TL in folded state and the TL in unfolded state are also called closed TL and open TL, respectively.

The open and closed TL conformations with respect to RNAP are shown in Figure 3A. Structure of the streptolydigin-bound *T. thermophilus* RNAP holoenzyme with an open TL (PDB 1ZYR, Tuske, et al., 2005) was aligned with structure of the AMPcPP-bound *T. thermophilus* TEC with a closed TL (PDB 2O5J, Vassylyev, et al., 2007b) using β subunit, and the open TL segment from the holoenzyme structure was superimposed into the TEC structure. Closing and opening of TL occurs in the secondary channel, through which substrates NTPs enter the catalytic site. The distance change of the TL tip is about 25 Å in the two conformations (Figure 3B).

In the post-translocated substrate-free TEC, the TL is usually disordered due to its flexibility, as seen in the X-ray structure of *T. thermophilus* TEC (PDB 2O5I, Vassylyev, et al., 2007a). Structure of *T. thermophilus* TEC bound with AMPcPP and streptolydigin

(PDB 2PPB) shows a partially ordered TL, in which the N-terminus portion of the TL retains the α-helical configuration but is sharply bent, the C-terminus portion is in an extended rather than an α-helical configuration, and the TL tip is disordered. Crystal structure of *T. thermophilus* RNAP holoenzyme bound with antibiotic streptolydigin, reveals a completely ordered TL in open conformation (PDB 1ZYR, Figure 3B, Tuske, et al., 2005). The TL adopts an extended conformation, in which the N-terminus and C-terminus of the TL are sharply bent, forming a loop with the tip pointing away from the active site in the secondary channel. Crystal structure of inhibitor-free *E. coli* σ^{54}-RNAP holoenzyme (PDB 5BYH, Figure 4D) also contains a mostly ordered open TL with only four missing residues (β'937-940). Conformations of open TL in the *T. thermophilus* RNAP and *E. coli* RNAP basically agree with each other except some difference in detailed residue positions. Since the open TL is likely to adopt multiple conformations due to its flexibility, it is not surprising that there is some variance in observed positions of open TL.

In response to the insertion of cognate NTP, the TL folded into two α-helical segments that extend the α-helices flanking the TL to form two long anti-parallel trigger helices (TH, β'1221-1265 for *T. thermophilus*, β'915-938 and β'1135-1147 for *E. coli*), as shown in the structure of the AMPcPP-bound *T. thermophilus* TEC (PDB 2O5J, Figure 3). The two helices interact with the adjacent bridge helix (BH, β'1066-1103 for *T. thermophilus*, β'768-805 for *E. coli*), creating a metastable three-helical bundle. The folded TL generates a complex network of interactions with the loaded substrate. The TL residue β'Met1238 (β'Met932 in *E. coli*) directly stack on the substrate base, suggesting its importance in substrate positioning and selection. Residue β'Phe1241 (β'Phe935 in *E.*

coli) also interacts with the substrate base via van der Waals force. The TH helps stabilize the conformation of the substrate phosphates through hydrogen bounds with β'His1242 (β'His936 in *E. coli*) recognizing the α-phosphate and β'Arg1239 (β'Arg933 in *E. coli*) recognizing the β- and γ-phosphates. Notably, there is no observed direct interaction between the TH and the substrate ribose in this study. Superimposition of the *T. thermophilus* TEC with two *E. coli* transcription initiation complexes (PDB 4YLN, Zuo, et al., 2015; PDB 5IPM, Liu, et al. 2016) revealed high similarities between these structures in terms of the folded TL conformation and the network of interactions it forms, which implies that TL closing is a conserved process in different bacterial species (*T. thermophilus* versue *E. coli*) and in different transcription stages (elongation versus initiation). TL folding substantially reduces the dimensions of the secondary channel from ~15 Å × 22 Å in the substrate-free TEC (PDB 2O5I) to ~11 Å × 11 Å in the substrate-bound TEC (PDB 2O5J) at the entrance to the active center.

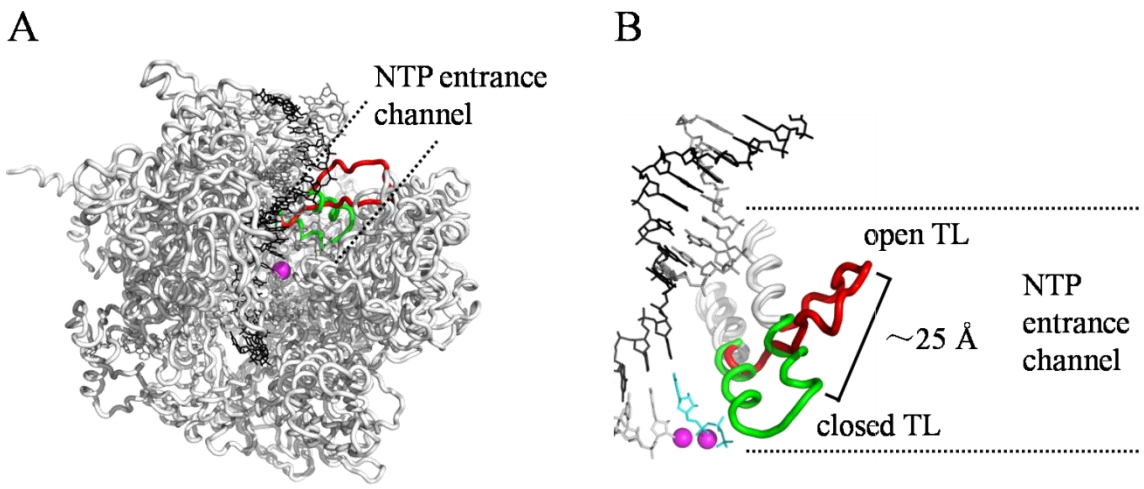


Figure 3. Open and closed TL.
(A) Structural overview of the open and closed TL in *T. Thermophilus* TEC. RNAP is indicated in ribbon representation (open TL in red, closed TL in green, the rest of RNAP in white), and the oligonucleotides are indicated in line representation (DNA template strand in black, DNA non-template strand in dark gray, and RNA chain in light gray). The active-center Mg^{2+} is indicated as a magentas sphere. NTP entrance channel is indicated with two dash lines.
(B) Crystal structure of *T. Thermophilus* open (PDB 1ZYR) and closed (PDB 2O5J) TL in the active center. The bound substrate is indicated in cyan, and the two Mg^{2+} ions are indicated as magentas spheres.

TL sequences are highly conserved in bacterial RNAPs and less highly conserved between prokaryotic and eukaryotic RNAPs, e. g. RNA polymerase II (Pol II) (Figure 4A). However, *E. coli* RNAP β' subunit contains a lineage-specific sequence insertion (termed SI3, β'945-1132) of 188 amino acids in length between the TL segments β' 926-944 and β' 1133-1137 (Artsimovitch, et al., 2003). The SI3 segment is in the middle of conserved region G in β', thus it is also known as β' subunit G region non-conserved domain (β'GNCD). The insertion is present in a wide range of bacterial species, including proteobacteria (e. g. *E. coli*), *Aquifex*, Chlamydia, Spirochaetes, *Cytophaga*,

Chlorobium, Planctomycetes, and *Fusobacterium* (termed Group I bacteria, Figure 4B, Chlenov, et al., 2005). It is absent in *Thermus* bacteria and in Eukaryotes, whose RNAPs have been determined by crystallography. In *E. coli* RNAP, the SI3 domain is located between the β' jaw and the β' rim helices (RH), at the entrance of the secondary channel (Figure 4C). X-ray structure of *E. coli* SI3 module reveals a rod-shape of the large domain (~ 60 Å × ~ 50 Å × ~ 40 Å) comprising two tandem repeats of a sandwich-barrel hybrid motif: SBHMa and SBHMb (Chlenov, et al., 2005). Based on the EM structure of *E. coli* activator-dependent transcription initiation complex, it has been proposed the SI3 domain varies in at least two positions in response to the conformational cycle of TL (Hudson, et al., 2009). This view was consistent with the interpretation of the structural model of *E. coli* RNAP from a hybrid approach (Opalka, et al., 2010). All available crystal structures of *E. coli* RNAP that have a determined SI3 domain as of October 1, 2016 are categorized in regard to their TL conformation (open or closed) in Table 1. Superimposition of these SI3 modules alone showed that the conformations are similar in SBHMa and SBHMb except minor differences in loops b1 and b2. However, alignment of the whole RNAP structures using β subunit identified three distinct orientations of the SI3 module, named groups A, B, and C. They account for 6, 8, and 1 structure(s) out of the 15 structures, respectively. Compared with group B, the SI3 module in group A is more "closed up" toward the β dispensable region 1 (β DR1, β 223–339) with the top of SBHMb ～25 Å closer to the β DR1, exhibiting a ～26 ° rotation of the whole SI3 domain about the bottom of SBHMa (Figure 4C, D). The SI3 orientation in group C is between groups A and B, suggesting it might be an intermediate state. Interestingly, all closed-TL structures exhibit an SI3 orientation of group A, and all open-TL structures

exhibit an SI3 orientation of group B or group C (which might be considered as a state in transition to group B). This observation was consistent with the conclusions based on the EM structure and the structural model of *E. coli* RNAP. We speculated that, the inter-conversion in orientation states of SI3 is coupled to the TL closing and opening in each nucleotide addition step.

A

```
E. coli            926  PGTQLTMRTFHIGGAASRA--(SI3)--DITGG 1137
T. thermophilus   1232  PGTQLTMRTFHTGGVAGAA---------DITQG 1255
S. cerevisiae     1075  PATQMTLNTFHFAGVASKK----------VTSG 1097
```

B

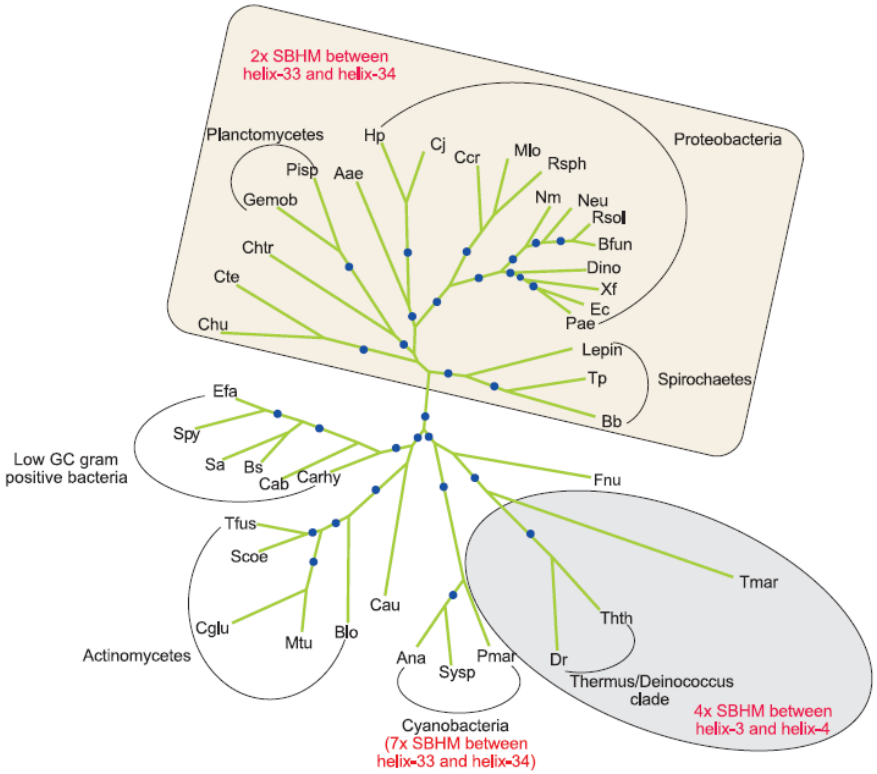


Figure 4. Conservation of TL in RNAPs and lineage-specific SI3 in *E. coli* RNAP.
(A) Conservation of TL residues in bacterial RNAP (*E. coli* and *T. thermophilus*) β' subunit and eukaryotic (*S. cerevisiae*) Pol II Rpb1.
(B) Phylogenetic tree of the β' subunit (adapted from Chlenov, et al., 2005). Group I bacteria is indicated in a pink rectangular. Species abbreviations are as follows: Ec: *E. coli*, Rru: *Rhodospirillum rubrum*, Neu: *Nitrosomonas europaea*, Dino: *Dichelobacter nodosus*, Nm: *Neisseria meningitidis*, Pae: *Pseudomonas aeruginosa*, Xf: *Xylella fastidiosa*, Vc: *Vibrio cholerae*, Rsol: *Ralstonia solanacearum*, Bfun: *Burkholderia fungorum*, Mlo: *Mesorhizobium loti*, Chu: *Cytophaga hutchinsonii*, Cte: *Chlorobium tepidum*, Ccr: *Caulobacter crescentus*, Rsph: *Rhodobacter sphaeroides*, Hp: *Helicobacter pylori*, Cj: *Campylobacter jejuni*, Aae: *A. aeolicus*, Gemob: *G. obscuriglobus*, Pisp: *Pirellula sp.*, Chtr: *Chlamydia trachomatis*, Tp: *Treponema pallidum*, Bb: *Borrelia burgdorferi*, Lepin: *Leptospira interrogans*, Fnu: *F. nucleatum*, Thth: *Thermus thermophilus*, Dr: *Deinococcus radiodurans*, Tmar: *Thermot. maritima*, Sysp: *Synechocystis sp.*, Ana: *Anabaena sp.*, Pmar: *Prochlorococcus marinus*, Mtu: *Mycobacterium tuberculosis*, Cglu: *Corynebacterium glutamicum*, Scoe: *Streptomyces coelicolor*, Tfus: *Thermobifida fusca*, Blo: *Bifidobacterium longum*, Bs: *Ba. subtilis*, Spy: *Streptococcus pyogenes*, Efa: *Enterococcus faecalis*, Sa: *Staphylococcus aureus*, Cab: *Clo. acetobutylicum*, Carhy: *Carboxydothermus hydrogenoformans*, Cau: *Chloroflexus aurantiacus*.

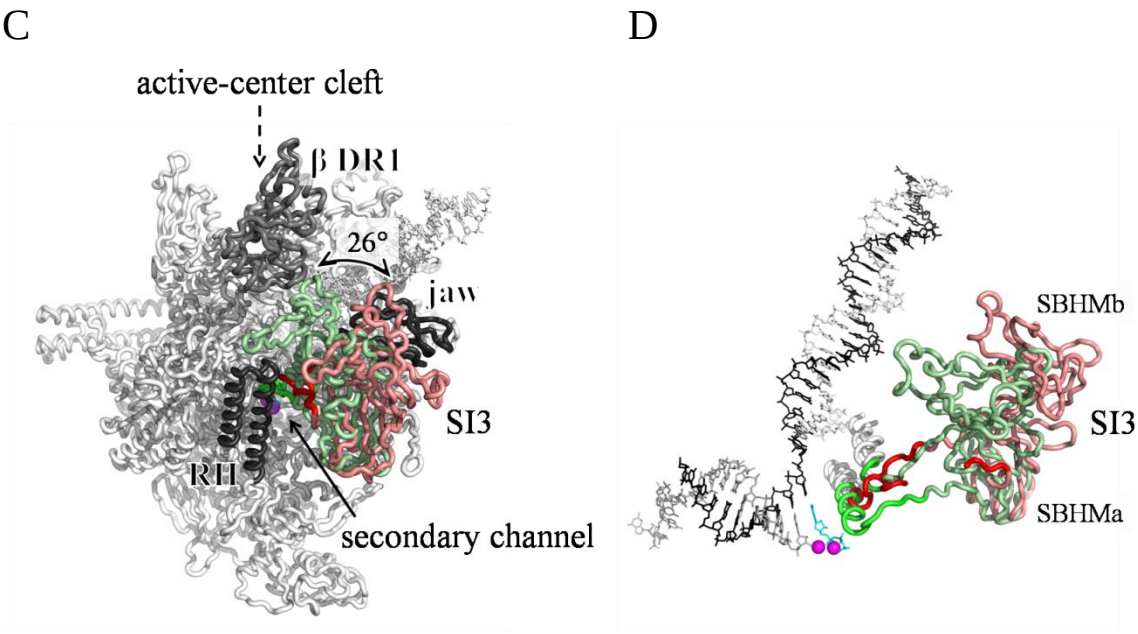


Figure 4. Conservation of TL in RNAPs and lineage-specific SI3 in E. coli RNAP. (continued)

(C) Overview of SI3 orientations with respect to RNAP. The SI3 orientation in group A accommodating the closed TL (green) is in light green (PDB 4YLN), the one in group B accommodating the open TL (red) is in pink (PDB 5BYH), and the one in group C is not shown. β' rim helices (RH) and jaw are in dark gray, β DR1 is in gray, substrate NTP is in cyan, Mg^{2+} ions are indicated as magentas spheres, and the rest part of the TEC is in white.

(D) Active-center view of the SI3 orientations accommodating the open and closed TL conformations.

Table 1. Correlation between TL conformations and SI3 orientations in *E. coli* RNAP.

TL conformation	PDB ID	Description	SI3 orientation	Ref.
closed	4YLN	σ^{70}-RP_{itc4} with a 17-nt -35/-10 spacer	A	Zuo, et al., 2015
closed	4YLO	σ^{70}-RP_{itc4} with a 16-nt -35/-10 spacer	A	Zuo, et al., 2015
closed	4YLP	σ^{70}-RP_{itc5} with a 16-nt -35/-10 spacer	A	Zuo, et al., 2015
closed	5IPL	σ^{S}-RP_{itc4} with a 14-nt bubble and PPi	A	Liu, et al., 2016
closed	5IPM	σ^{S} -RP_{itc4} with a 14-nt bubble	A	Liu, et al., 2016
closed	5IPN	σ^{S} -RP_{itc4} with a 14-nt bubble	A	Liu, et al., 2016
open	5BYH	σ^{54}-RNAP holoenzyme	B	Yang, et al., 2015
open	4ZH2	σ^{70}-RNAP holoenzyme with CBR703	B	Feng, et al., 2015
open	4ZH3	σ^{70}-RNAP holoenzyme with CBRH16-Br	B	Feng, et al., 2015
open	4ZH4	σ^{70}-RNAP holoenzyme with CBRP18	B	Feng, et al., 2015
open	4LJZ	Δ1.1 σ^{70}-RNAP holoenzyme	B	Bae, et al., 2013
open	4LK0	Δ1.1 σ^{70}-RNAP holoenzyme with Gp2	B	Bae, et al., 2013
open	4LK1	SeMet σ^{70}-RNAP holoenzyme	B	Bae, et al., 2013
open	4LLG	σ^{70}-RNAP holoenzyme with Gp2	B	Bae, et al., 2013
open	4JKR	σ^{70}-RNAP holoenzyme with ppGpp	C	Zuo, et al., 2013

1.5. Proposed roles of closing and opening of TL

Extensive structural and biochemical studies have shown that conformational change of TL plays key roles in each step of nucleotide addition cycle: substrate binding and selection, catalysis, and translocation. In post-translocated state of the RNAP complex, the TL is open to allow entrance of NTP to the active site. During substrate loading, transition from the pre-insertion state to the insertion state is believed to be coupled with the closing of TL. Upon binding of the cognate substrate, the TL folds into an α-helical hairpin thereby largely reducing the dimensions of the secondary channel at the entrance to the active site, which provides hindrance for substrate dissociation and competition (Vassylyev, et al., 2007b; Liu, et al. 2016). The folded TL facilitates substrate alignment probably via its contacts with the substrates in at least two ways: β'Met932 (maybe also β'Phe935) to the base; β'His936 and β'Arg933 to the phosphates (*E. coli* residue numbering is used here and afterwards in this work unless stated otherwise). β'Met932, together with the 3' end of the RNA/DNA hybrid on the other side, sandwiches the substrate base by stacking on it, likely contributing to the substrate alignment as well as selection. By interacting with the triphosphates, the closed TL aligns the incoming substrate with the RNA 3'OH and Mg^{2+} ions to facilitate the catalytic reaction (Vassylyev, et al., 2007b). It has been proposed that TL is critical for discrimination between complementary and non-complementary NTPs (Wang, et al., 2006; Vassylyve, et al., 2007b). TL is also proposed to make a contribution to the discrimination in favor of NTPs over 2'dNTPs in transcription (Wang, et al., 2006; Kaplan, et al., 2008), although its direct contact with the ribose 2'-OH is absent in crystal structures. In addition, the TL folding helps dehydrate the active center thereby facilitating optimum conditions for

phosphodiester bond synthesis (Seibold, et al., 2010). Single molecule experiments using TL mutants of RNAP have identified the conformational change of TL as one of the two rate-limiting steps in pause-free elongation (Mejia, et al., 2015). Substitutions of the TL change transcription rate as well as transcription fidelity (Bar-Nahum, et al., 2005). In contrast to the bridge-helix-centric structural model, a trigger-loop-centric model was supported by recent biochemical and structural studies of TECs (Toulokhonov, et al., 2007; Vasslyev, et al., 2007; Wang, et al., 2006). According to this model, a correct NTP binds to the i+1 site and sterically restricts the reversal of translocation, biasing the state equilibrium toward the post-translocated direction. On the other hand, the NTP binding induces refolding of the TL, which in turn further stabilizes the NTP binding. Therefore, the binding NTP and the folded TL act as two pawl locking the complex in the post-translocated state.

The TL has been proposed to be a checkpoint for base complementarity, a checkpoint for triphosphate/diphosphate /monophosphate identity, and checkpoint for ribose/deoxyribose identity. It is also believed to serve as a target for inhibitors and effectors, a “trigger" for catalysis, and “trigger" for pausing and termination. However, all of this is based on crystal structures, and direct evidence for TL closing/opening in solution is lacking.

In order to better understand the conformational cycling of TL in solution, this work specifically aims at: (i) determination of whether TL closing and opening occurs in solution, (ii) analysis of effects of substrate identities on TL conformation in solution (including base complementarity, triphosphate/diphosphate /monophosphate identity, and

ribose/deoxyribose identity), (iii) mapping TL positions in open and closed states in solution; and (iv) analysis of effect of TL conformation on SI3 orientation.

2. Experimental strategy

All of three aims in this work were addressed by performing experiments based on fluorescence resonance energy transfer (FRET), also called Förster resonance energy transfer (FRET). FRET is a mechanism by which energy is transferred between two chromophores (term "donor" and "acceptor") through non-radiative dipole-dipole coupling when the donor molecule (must be a fluorophore) is activated to its electronic excited state. There must be significant spectral overlap between the donor emission and the acceptor excitation to allow efficient FRET. Practical donor-acceptor distances are typically 10-100 Å, on the order of common macromolecular dimensions (Clegg, 1992). The FRET efficiency (*E*) is inversely proportional to the sixth power of donor-acceptor distance (*R*) as in

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

where R_0, termed Förster radius, is a parameter that defines the donor-acceptor distance at which the FRET efficiency is equal to 50%. R_0 can be determined as in Eq. (2)

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

where n is the refractive index of the medium between the donor and acceptor (used n = 1.4 in this work), κ is the orientation factor between the donor emission and acceptor excitation dipoles (assuming $\kappa^2 = 2/3$ and justified by fluorescence anisotropy measurements), Q_D is the fluorescent quantum yield of the donor in the absence of

acceptor, and $J_{(\lambda)}$ is the spectral overlap integral between the donor emission and acceptor excitation. The wavelength λ is expressed in cm and $J_{(\lambda)}$ is in units of M^{-1} cm^3 , and the resulting R_0 is in Å (Lakowicz, 2006). In turn, the distance R can be calculated using R_0 and the measurable E as in Eq. (3). Therefore, FRET has been termed "molecular ruler" (Stryer , et al., 1967, Figure 5A). To obtain accurate measurements of E, the distance to be determined is usually within the range of 0.5 R_0 – 2 R_0 (E = 98.5% when R = 0.5 R_0; E = 1.5% when R = 2 R_0).

$$R = R_0 [(1/E) - 1]^{1/6} \quad (3)$$

There are several experimental approaches to determine FRET efficiency: 1) measuring decreased fluorescence lifetime of donor; 2) measuring decreased fluorescence intensity of donor and increased fluorescence intensity of acceptor; 3) measuring decreased fluorescence quantum yield of donor; 4) measuring change in fluorescence anisotropy of donor and acceptor (Clegg, 1992). Fluorescence lifetime (τ) is a measure of time that a fluorophore spends in the excited state before returning to the ground state through radiative ways (e.g. emitting a photon) or non-radiative ways (e.g. FRET). It is an intrinsic molecular property and, within certain constraints, independent of concentration. When FRET occurs between the donor and acceptor, the fluorescence lifetime of donor would be shorter than that in the absence of the acceptor. The FRET efficiency can be determined directly from measurements of the fluorescence lifetime of Cou as in Eq. (4)

$$E = 1 - \tau_{DA} / \tau_D \quad (4)$$

where τ_D refers to the decay process of the donor species alone, and τ_{DA} refers to that of the donor in the presence of the acceptor (Clegg, 1992). This is the most direct way to measure *E*.

As FRET occurs or FRET efficiency increases due to a shorter probe-probe distance, the donor fluorescence intensity will decrease and the acceptor fluorescence intensity will increase. At the same time, the fluorescence lifetime of donor will decrease. Conversely, if FRET disappears or FRET efficiency decreases due to a longer distance, the donor fluorescence intensity will increase and the acceptor fluorescence intensity will decrease, accompanied by an increase in the donor lifetime. This is the theoretical basis of the experimental strategy in this work (Figure 5B).

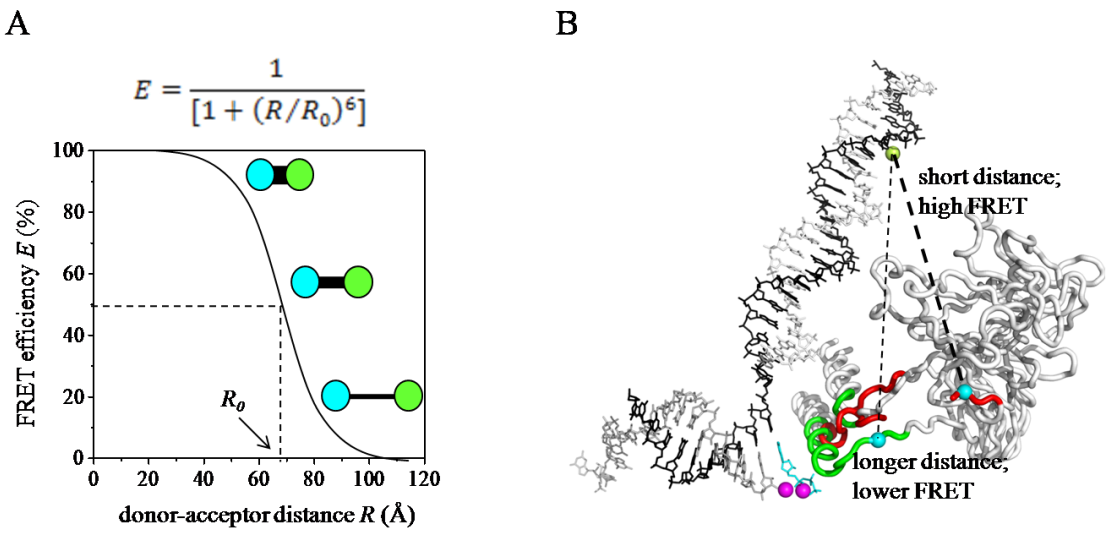


Figure 5. Experimental strategy: FRET as a molecular ruler.
(A) Dependence of FRET efficiency (*E*) on probe-probe distance I. The FRET donor and acceptor are indicated as cyan sphere and green sphere, respectively. The length of the black bar connecting donor and acceptor indicates the probe-probe distance, and the bar width indicates the FRET efficiency.
(B) Application of FRET on the TEC. The donor probe (cyan sphere) is placed on TL, and the acceptor probe (green sphere) is place on DNA template strand.

2.1. Aim (i): determination of whether TL closing and opening occurs in solution

Whether TL closing and opening occurs in solution was directly assessed by monitoring ensemble FRET efficiencies between a fluorescent probe incorporated into TL and a complementary probe incorporated at a reference site in the DNA template strand (Figure 5B). The reference site remains stationary in the TEC during the observation window of time in a nucleotide addition cycle, thus changes in interprobe distance implies movement of TL.

2.1.1. Labelling TL

Incorporating a probe into RNAP TL is challenging because: (i) RNAP contains more than 3400 residues in 5 subunits, thus the labelling must be highly selective; (ii) TL is in interior of RNAP, thus the label must not interfere with RNAP assembly; (iii) TL is in the active center of RNAP, thus the label must not interfere with RNAP activity. To selectively label TL, I applied the method of unnatural amino acid mutagenesis (Wang, et al., 2001), which can introduce a probe with ~100% incorporation specificity into any position of interest in the protein during translation by use of orthogonal tRNA/synthetase pair. To minimize the probe-caused interference in RNAP assembly and activity, the coumarin-derived fluorescent amino acid L-(7-hydroxycoumarin-4-yl)-ethylglycine (Cou, Figure 6A; excitation maximum = 365 nm; emission maximum = 457 nm; quantum yield = 0.63) was used for labelling TL (Wang, et al., 2006). Cou has the backbone identical

with that of standard amino acids, and its side chain is only slightly larger than that of Phe, Tyr, and Trp. By combining the unnatural amino acid mutagenesis with the RNAP-reconstitution procedure (Mukhopadhyay, et al., 2003), RNAP derivatives with a Cou-labelled TL were obtained. The TL segment covering residues β' 933-942 was chosen for incorporation because the segment meets two criteria: (i) it contains residues that exhibit large differences (15-30 Å) in positions between the open and closed state defined by crystal structures (Figure 6B); (ii) its sequence is conserved in *E. coli* RNAP (the experimental system in this work) and *T. thermophilus* RNAP and *S. cerevisiae* Pol II (the crystallographically defined systems) (Figure 4A). To identify positions within this segment that allow efficient incorporation of Cou and tolerate the incorporation without loss of transcriptional activity, Cou was incorporated into each of these positions by overexpression of β' subunit in the presence of Cou, prepared RNAP derivatives by reconstitution from RNAP subunits (α, β, labelled β', and ω) and analyzed the transcriptional activity of the RNAP derivative. Starting form this method, I further developed three additional preparation methods to improve incorporation efficiency of Cou, transcriptional activity and monodispersity of the RNAP derivatives.

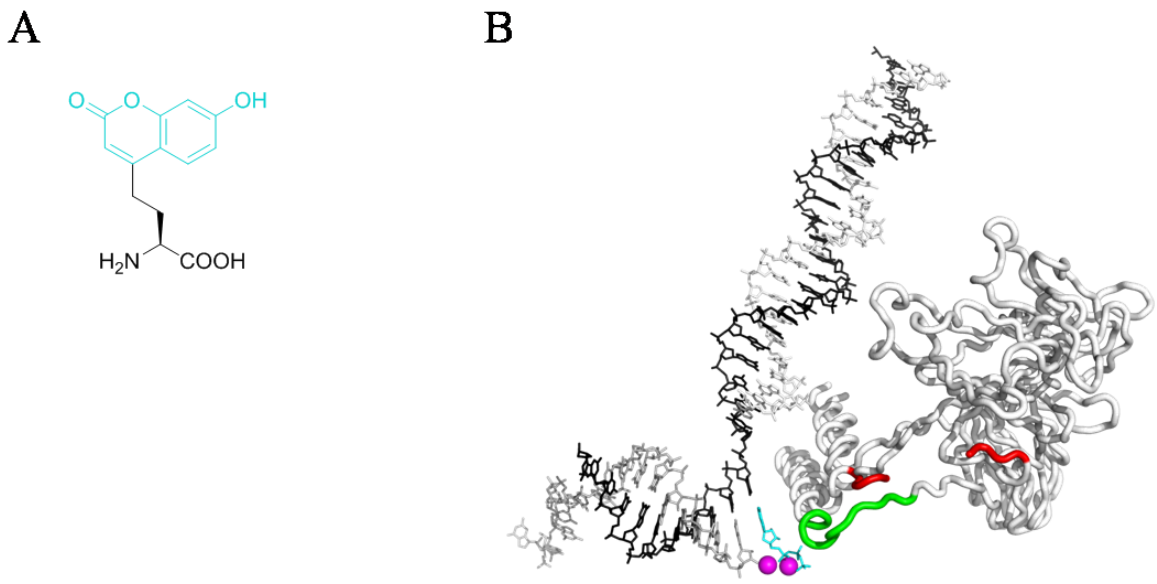


Figure 6. Labelling TL.
(A) Coumarin-derived fluorescent unnatural amino acid L-(7-hydroxycoumarin-4-yl)-ethylglycine. The Coumarin-derived motif is highlighted in cyan.
(B) *E. coli* residues β'933-942 for labelling TL. The segment β'933-942 is highlighted in red and green in the open and closed TL, respectively, with the rest of TL in white.

2.1.2. Labelling DNA

A complementary fluorescent probe, serving as a FRET acceptor, was incorporated at each of eleven reference sites in the DNA template strand. They are positions -6, -4, -2, +8, +10, +12, +14, +15, +16, +18, and +20 relative to the active center. The positions of first eight nucleotides (-6 to +15) have been crystallographically defined (Vassylyev, et al., 2007a; Vassylyev, et al., 2007b), and the last three were defined by extending the crystallographically determined downstream DNA duplex with a B-form DNA to position +25 (Naryshkin, et al., 2000; Mekler, et al., 2002). These nucleotides located ~ 20-70 Å, within the range of practical FRET distances (10-100 Å), from positions of

trigger-loop residues of interest in the crystallographically defined closed and open states. Three (-6, -4, and -2) of the sites are located upstream to the active center, so that the donor-acceptor distances are expected to decrease as the TL closes according to the proposal. The other eight sites (+8, +10, +12, +14, +15, +16, +18, and +20) are located downstream to the active center, so that the donor-acceptor distances are expected to increase as the TL closes.

Fluorescent probe fluorescein (excitation maximum = 495 nm; emission maximum = 520 nm) was used as the FRET acceptor. Its excitation spectrum overlaps with the emission spectrum of Cou, and its size is relatively small compared with probes of similar excitation wavelength. R_0 for Cou-fluorescein pair was reported to be 47 Å (Yi, et al., 2011), which should nicely cover the range of expected residue-nucleotide distances (～20-70 Å) based on crysal structures. Fluorescein was covalently attached to the nucleotide base through a flexible linker of ～13 Å (for purines), ～15 Å (for thymine) or ～16 Å (for cytosine) in length, providing flexibility and space for free orientation of the probe (Figure 7B). The probes on DNA are expected be accommodated on the major grove side rather than the minor grove side due to the geometry of the attachment point on the bases (C8 for purines, C7 for thymine and C5 for cytosine). Labelling specificity of ～100% and labelling efficiency of ～100% were ensured by use of a single reactive modified base incorporated in the synthetic oligos and by purification of the labelled oligos with HPLC. A labelled DNA oligo as the DNA template strand, together with complementary DNA non-template strand and complementary RNA, formed a labelled nucleic acid scaffold by annealing.

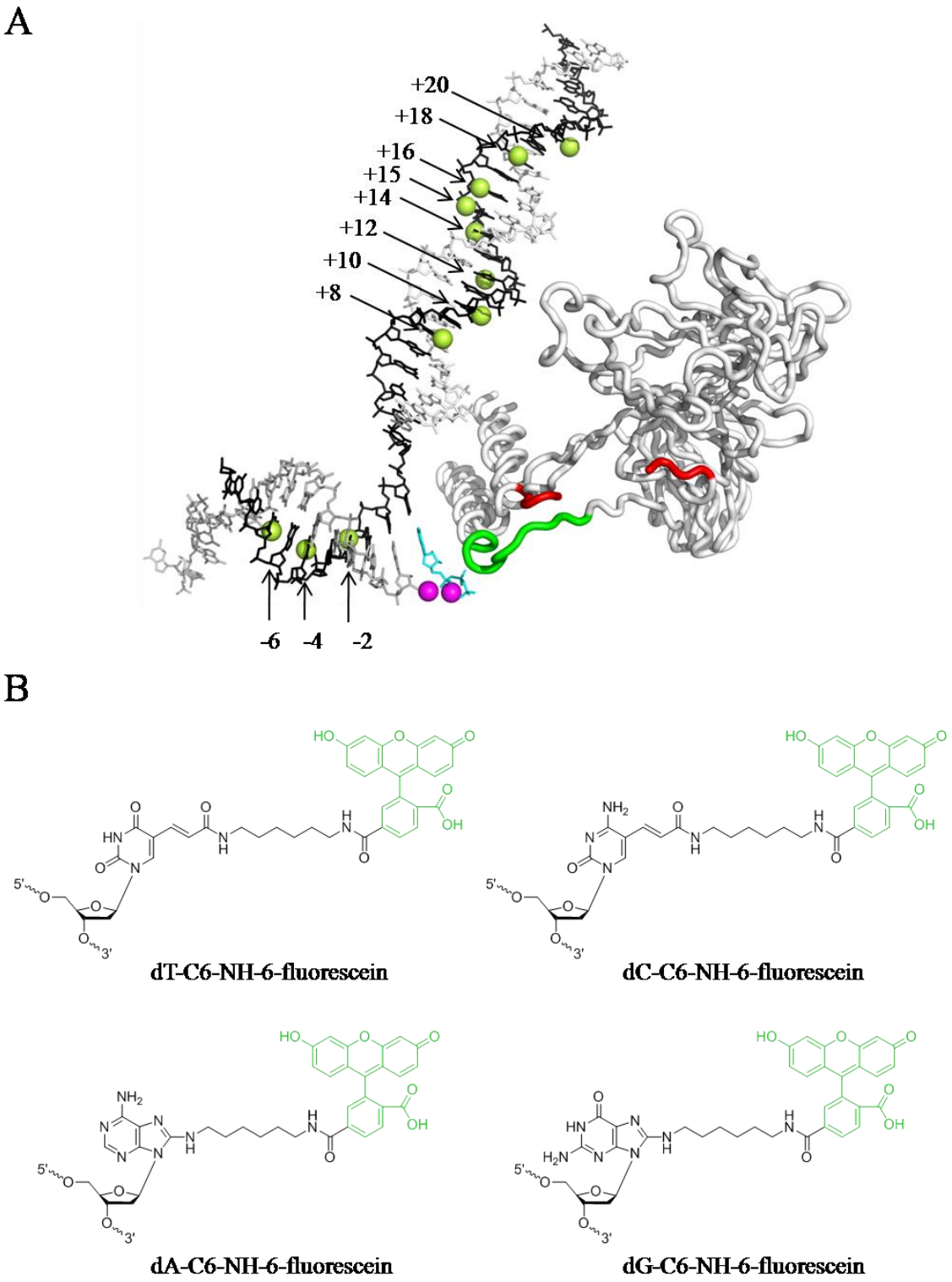


Figure 7. Labelling DNA.
(A) Labelling sites on the DNA template strand. Acceptor probes are indicated as green spheres at the base atom where the connecting linker is attached to (C8 for purines, C7 for thymine and C5 for cytosine), omitting the linker length.
(B) Fluorescein-labelled nucleotides. Fluorescein motif is highlighted in green.

C

```
                                Position      -8 -6 -4 -2  2  4  6  8  10 12 14 16 18 20 22 24
                                              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
scaffold 0                                    |  |  |  5'-ACGCCAGACAGGACCTCAGTCCG-3'
unlabelled                                 3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 1                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: -6                      3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 2                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: -4                      3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 3                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: -2                      3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 4                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +8                      3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 5                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +10                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 6                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +12                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 7                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +14                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 8                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +15                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 9                                                5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +16                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 10                                               5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +18                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'

scaffold 11                                               5'-ACGCCAGACAGGACCTCAGTCCG-3'
labelled position: +20                     3'-GCCGCGCGCTTGCGGTCTGTCCTGGAGTCAGGC-5'
substrate: ATP                          5'-GAGUCUGCGGCGCGCG-3'
```

Figure 7. Labelling DNA. (continued)
(C) Unlabelled scaffold (scaffold 0) and fluorescein-labelled nucleic-acid scaffolds (scaffolds 1-11). Each nucleic acid scaffold consists of a DNA non-template strand (top), a DNA template strand (middle) and a RNA (bottom).Acceptor template nucleotide is indicated in red. Fluorescein-labelled nucleotide is in green. Non-extendable RNA terminus 3'-deoxyguanosine is underlined.

2.1.3. Preparation of doubly-labelled TEC

FRET experiments were performed with transcription elongation complexes (TECs). The TECs contain a nucleic-acid scaffold (Figure 7C, Vassylyve, et al., 2007a, Kashkina, et al., 2006) that favors post-translocated state (due to the absence of non-template nucleotide complementary to the acceptor template base i+1), that precludes backtracking (due to the absence of upstream portion of DNA template strand and the absence of the whole DNA non-template strand in the transcription bubble), that precludes extension of the RNA (due to the presence of a 3'-deoxy-terminus, Vassylyev, et al., 2007b), that directs binding of ATP as the cognate NTP. The TECs were prepared by reconstitution from Cou-labelled RNAP and fluorescein-labelled scaffold. For ensemble FRET experiments with freely diffusing molecules, the TECs were purified by isolation with electrophoretic mobility shift assay from excess RNAP and scaffold, followed by recovery with electroelution technique, to remove undesired fluorescent species.

2.1.4. FRET measurements

FRET efficiencies between Cou in TL and fluorescein in DNA were monitored in two strategies: (i) semiquantitatively analyzing changes in the fluorescence intensities of donor and acceptor in doubly-labelled TEC without and with bound ATP; (ii) quantitatively determining fluorescence lifetimes of the donor in donor-only and doubly-labelled TECs without and with bound ATP.

In the absence and the presence of bound ATP, fluorescence emission spectra of a doubly-labelled TEC with the donor probe incorporated in TL and the acceptor probe incorporated in downstream DNA (position +20) were recorded, and differences in fluorescence intensities of the donor and acceptor were individually analyzed. According to the proposal, the probe-probe distance is expected to increase on transition from the open TL conformational state to the closed state, thus the FRET efficiency is expected to decrease. As a result, an increased fluorescence intensity of the donor and a decreased fluorescence intensity of the acceptor would be observed in the spectra after addition of ATP. This is a straightforward visual aid to understand changes of FRET efficiency.

FRET efficiencies were quantitatively determined from the donor fluorescence lifetime in the absence and the presence of the acceptor. First, Cou-only TEC which contains a Cou-labelled RNAP and an unlabelled scaffold was used as a control for τ measurements. Fluorescence decays were measured with the ATP-unbound TEC (proposed to adopt an open trigger-loop conformational state) and with the ATP-bound TEC (proposed to adopt a closed trigger-loop conformational state), generating τ_D (open) and τ_D (closed), respectively. Second, doubly-labelled TEC which contains a Cou-labelled RNAP and a fluorescein-labelled scaffold (DNA position +20) was used for τ measurements in a analogous procedure, generating $\tau_{DA,+20}$ (open) and $\tau_{DA,+20}$ (closed). Consequently, FRET efficiency in the ATP-unbound doubly-labelled TEC can be determined as

$$E_{+20}\,(\text{open}) = 1 - \tau_{DA,+20}\,(\text{open}) / \tau_D\,(\text{open}) \qquad (5)$$

And FRET efficiency in the ATP-bound doubly-labelled TEC can be determined as

$$E_{+20}\ (\text{closed}) = 1 - \tau_{DA,+20}\ (\text{closed}) / \tau_{D}\ (\text{closed}) \quad (6)$$

From E_{+20} (open) and E_{+20} (closed), the corresponding probe-probe distances R_{+20} (open) and R_{+20} (closed) were determined as in Eq. (3). The distance change from R_{+20} (open) to R_{+20} (closed) infers the movement of TL. Two sets of parallel experiments were performed: (i) using acceptor incorporated in seven additional sites on DNA downstream to the active center (positions +8, +10, +12, +14, +15, +16, and +18) in which FRET efficiencies were expected to change in the same direction, and (ii) using acceptor incorporated in three sites on DNA upstream to the active center (positions -6, -4, and -2) in which FRET efficiencies were expected to change in the opposite direction.

2.2. Aim (ii): analysis of effects of substrate identities on TL conformation in solution.

To assess effects of complementary and non-complementary NTPs on trigger-loop conformation in solution, I performed experiments analogous to those in the preceding section, but using four nucleic-acid scaffolds (Figure 8)--directing binding of ATP, GTP, UTP, and CTP as the incoming complementary NTPs--and, for each nucleic-acid scaffold, performing experiments with no NTP, with complementary NTP, and with each non-complementary NTP.

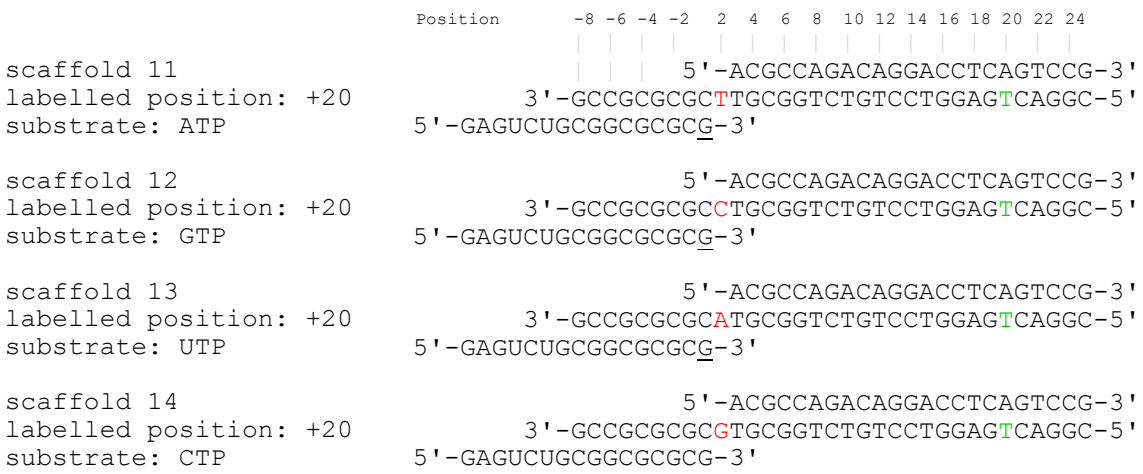


Figure 8. Nucleic-acid scaffolds used for analyzing effects of base complementarity on TL conformation.

To assess effects of dNTPs on trigger-loop conformation in solution, I performed experiments analogous to those in the preceding section, but using 2'-dATP and 3'-dATP instead of NTP, and using one representative nucleic-acid scaffold (scaffold 11, Figure 7C, Figure 8).

To assess effects of nucleoside diphosphate (NDP) and nucleoside monophosphate (NMP) on trigger-loop conformation in solution, I performed experiments analogous to those in the preceding section, but using ADP and AMP instead of NTP, and using one representative nucleic-acid scaffold (scaffold 11, Figure 7C, Figure 8).

2.3. Aim (iii): mapping TL positions in open and closed states in solution

Positions of TL in open and closed conformational states with respect to the TEC were mapped by use of FRET-derived distance-restrained docking. The experiments addressing aim (i) generated two sets of systematic FRET-derived distances: one with NTP-free TEC and the other with NTP-bound TEC. The results were used as distance restrains to direct the positioning of TL in open and closed stated, respectively. Recently, several methods have been developed to derive structural information from single molecule FRET (smFRET) measurements (Beckers, et al., 2015; Kalinin, et al., 2012; Choi, et al., 2010). Seidel et al. presented a comprehensive toolkit for FRET-restrained structural modeling, termed FRET-restrained positioning and screening (FPS), which is able to carry out rigid body docking as well as screening of structural ensembles, for instance, conformations generated from molecular dynamics simulations (Kalinin, et al., 2012). They demonstrated the high accuracy of the approach by docking a DNA primer-template to HIV-1 reverse transcriptase and comparing the derived model with the known crystal structure. FPS was applied to map the positions of open and closed TL using the systematic FRET measurements addressing aim (i). In FPS, spatial distributions of the dyes were approximately modeled by a geometric accessible volume (AV) algorithm (Cai, et al., 2007; Muschielok, et al., 2008) based on the dye dimensions, the linker length and width, and the local structure of the biomolecules where the dyes were attached. It was assumed that a dye molecules is free to occupy every position within its AV and spends an equal amount of time in each position, thus its position can be represented by a mean

position. The modeled donor-acceptor distances were calculated from their mean positions. To find the position of labelled TL residues with respect to the TEC, rigid body docking was accomplished by minimizing the weighted data-model deviation (χ_E^2) for the set of n distance restrains

$$\chi_E^2 = \sum_{i=1}^{n} \frac{(R_{DA(i)} - R_{model(i)})^2}{(\Delta R_{DA})^2} \tag{7}$$

where R_{DA} is the experimentally measured donor-acceptor distance, ΔR_{DA} is the uncertainty of R_{DA}, and R_{model} is the modeled donor-acceptor distance. The subscript i refers to the i^{th} pair of donor-acceptor. In addition, steric clashes between docking bodies were prevented by introducing strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii.

$$\chi^2_{\text{clash}} = \sum_{i,j} \begin{cases} 0 & , r_{ij} \geq r_{wi} + r_{wj} \\ (r_{wi} + r_{wj} - r_{ij})^2 / r^2_{\text{ctol}} & , r_{ij} < r_{wi} + r_{wj} \end{cases} \tag{8}$$

where r_{ij} is the distance between atoms i and j which belong to different subunits, r_{wi} and r_{wj} are their van der Waals radii, and r_{ctol} is the pre-defined clash tolerance. In this way, positioning of TL residues relative to the TEC was guided by minimizing an overall quality parameter (reduced χ_r^2) that accounts for violation of FRET restrains and of van der Waals radii.

$$\chi_r^2 = (\chi_E^2 + \chi^2_{\text{clash}})/(n - p) \rightarrow \min \tag{9}$$

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

2.4. Aim (iv): analysis of effect of TL conformation on SI3 orientation in solution

Effect of TL conformation on SI3 orientation was analyzed based on the distance-restrained docking for determination of TL positions in the open and closed states. Docking simulations were performed in two strategies characterized by usage of different TEC structures as docking body. One strategy used a TEC structure without SI3 (named SI3-absent docking), and the other used two TEC structures with SI3(B) or SI3(A) (named SI3-present docking) for open TL and closed TL, respectively. Resulting structures by SI3-absent docking were superimposed with SI3(B) or SI3(A). Compatibility between TL conformational states (open or closed) and SI3 orientational states (A or B) was determined by checking for steric clashes.

3. Materials and methods

3.1. Nucleic acid fragments and scaffolds

DNA oligo used as non-template strand was purchased from IDT. DNA oligos used as template strands without fluorescein (unlabelled), or with fluorescein labelled at positions +8, +10, +15 or +20 were purchased from IDT. DNA oligos used as template strands with fluorescein labelled at positions -6, -4, -2, +12, +14, +16 or +18 were purchased from TriLink BioTechnologies. Non-extendible RNA oligo with terminal 3'-deoxyguanosine was purchased from TriLink BioTechnologies.

Nucleic acid scaffolds consisting of a DNA non-template strand, a DNA template strand and an RNA strand were generated by annealing a mixture solution of DNA non-template strand, DNA template strand and RNA at ratio of 1.2: 1: 4 in 10 mM Tris-HCl pH8.0 and 200 mM NaCl. At this ratio, DNA template strand (labelled or unlabelled) was annealed to scaffold near completion. Concentration of RNA was determined by titrating pre-annealed double-strand DNA with RNA so that almost all DNA was used to form DNA-RNA scaffold. Resulting scaffold and excess oligos were identified by applying reaction mixture to non-denaturing polyacrylamide gel electrophoresis, followed by gel staining and imaging with SYBR Gold (Thermo Fisher Scientific Inc.). Unlabelled scaffold was prepared with an unlabelled DNA template strand, and fluorescein-labelled scaffold was prepared using a DNA template strand with fluorescein labelled at positions -6, -4, -2, +8, +10, +12, +14, +15, +16, +18, or +20.

3.2. Plasmids

Plasmid pET28a-NF-αNTD^{I}-αNTD^{II} encodes an N-terminally Flag-tagged *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{I}), followed by GlySerGlyGlySerGly, followed by a second *E. coli* RNAP α subunit N-terminal domain (α residues 1-235; αNTD^{II}), that is, Flag-αNTD^{I}-GSGGSG-αNTD^{II}.

Plasmids pMKSe2 and pT7ω encode *E. coli* β subunit and ω subunit, respectively.

Plasmid pET21a-rpoC-CH6 (Chakraborty, et al.k, 2012) encodes C-terminally hexahistidine-tagged *E. coli* RNAP β' subunit. Plasmids pET21a-rpoCXXXTAG-CH6 (XXX = 933, 934, 935, 936, 937, 938, 939, 940, 941, or 942) were constructed by replacing one of the rpoC codons 933-942 in template pET21a-rpoC-CH6 with an amber codon (TAG) by use of site-directed mutagenesis (QuikChange II; Agilent Technologies, Inc.) (S. Mandal and R. Ebright, unpublished).

Plasmid pEVOL-CouA (Young, et al., 2010) encodes a mutant *Methanococcus jannaschii* tyrosyl amber suppressor tRNA (*Mj*tRNA$^{Tyr}_{CUA}$)/tyrosyl tRNA synthetase (*Mj*TyrRS) pair that selectively incorporates L-(7-hydroxycoumarin-4-ly)ethylglycine in response to the TAG codon. The plasmid harbors one copy of suppressor tRNA and two copies of the RS gene. The tRNA was optimized (tRNA$^{opt}_{CUA}$) by modifying the T-stem sequence for higher unnatural amino acid incorporation efficiency. The *Mj*TyrRS specifically evolved for Cou (*Mj*CouRS) has nine mutations: Tyr32Glu, Leu65His, Ala67Gly, His70Gly, Phe108Tyr, Gln109His, Asp158Gly, Leu162Gly, Asp286Arg. The plasmid was kindly provided by Dr. Peter Schultz, The Scripps Research Institute, La Jolla CA.

Plasmids pβ'$_{1\text{-}878}$ and pβ'$_{821\text{-}1407}$ encode *E. coli* RNAP β' subunit N-terminal residues 1-878 and C-terminal residues 821-1407, respectively.

Plasmid pEcABC(1-878)Z, encoding *E. coli* RNAP α subunit, β subunit, β' subunit N-terminal residues 1-878, and ω subunit under the control of T7 promoter, were constructed from pEcABC-H6 and pVS6. pEcABC-H6 (Hudson, et al., 2009) was digested by XhoI and XbaI, and the fragment containing subunits α, β, and β' was isolated. pVS6 were digested by XhoI and XbaI and the fragment containing ω subunit was isolated. These two fragments were joined to generate plasmid pEcABCZ. β' residues 606-878 (residues 610-612 contain a BsmI site) were replicated from pET21a-rpoC-H6 by add-on PCR to add an ochre stop codon (TAA) and an XhoI site following the rpoC gene. The PCR product was digested by BsmI and XhoI, and then joined with the large fragment of pEcABCZ pre-treated by BsmI and XhoI, generating pEcABC(1-878)Z.

pCDF-rpoC(821-1407)-CH6 was constructed by replacing the NcoI-HindIII small fragment of vector pCDFDuet-1 with the BspHI-HindIII fragment (encoding β' subunit residues 821-1407 followed by a C-terminal hexa histidine-tag) replicated from pET21a-rpoC-CH6.

pCDF-rpoC(821-1407, 941TAG942GCG943GCG)-CH6 and pCDF-rpoC(821-1407, 942TAG943GCG)-CH6 were constructed from pCDF-rpoC(821-1407)-CH6 by changing the 942 and 943 codons to TAA and GCG, respectively, with site-directed mutagenesis using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies).

Plasmids used in this work were summarized in Table 2.

Table 2. Plasmids

Plasmids	Characteristics	Source or Ref.	Used in Method*
pEVOL-Cou	CmR; ori-p15A; *araBADP-CouRS, glnS'P-CouRS, proKP- tRNA$^{opt}_{CUA}$*	Young et al. 2010	1, 2, 3, 4
pET28a NF α^{I}NTD- α^{II}NTD	ApR; ori-pBR322; ϕ10P-rpoA(1-235)-rpoA(1-235)(NFLAG)	Wang et al., 2008	1, 2, 3
pMKSe2	ApR; ori-pBR322; lacP-rpoB	Severinov et al., 1993	1, 2, 3
pT7ω	ApR; ori-pBR322; ϕ10P-rpoZ	Naryshkin et al., 2001	1, 2, 3
pET21a-rpoC-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC*	Chakraborty et al., 2012	1
pET21a-rpoC933TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC933amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC934TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC934amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC935TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC935amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC936TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC936amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC937TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC937amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC938TAG-CH6	Ap^{R}; ori-pBR322; ori-f1; *φ10P-rpoC938amber*	S. Mandal and R. Ebright, unpublished	1

Table 2. Plasmids (continued)

Plasmids	Characteristics	Source or Ref.	Used in Method*
pET21a-rpoC939TAG-CH6	Ap^R; ori-pBR322; ori-f1; *φ10P-rpoC939amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC940TAG-CH6	Ap^R; ori-pBR322; ori-f1; *φ10P-rpoC940amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC941TAG-CH6	Ap^R; ori-pBR322; ori-f1; *φ10P-rpoC941amber*	S. Mandal and R. Ebright, unpublished	1
pET21a-rpoC942TAG-CH6	Ap^R; ori-pBR322; ori-f1; *φ10P-rpoC942amber*	S. Mandal and R. Ebright, unpublished	1
pβ'$_{1-878}$	Km^R; ori-pBR322; ori-f1; *lacP-φ10P-rpoC(1–878)*	Severinov et al. 1996	2, 3
pβ'$_{821-1407}$	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)*	Severinov et al. 1996	2, 3
pβ'$_{821-1407}$935TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)935amber*	this work	2
pβ'$_{821-1407}$936TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)936amber*	this work	2
pβ'$_{821-1407}$937TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)937amber*	this work	2
pβ'$_{821-1407}$938TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)938amber*	this work	2
pβ'$_{821-1407}$939TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)939amber*	this work	2

Table 2. Plasmids (continued)

Plasmids	Characteristics	Source or Ref.	Used in Method*
pβ'$_{821\text{-}1407}$940TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)940amber*	this work	2
pβ'$_{821\text{-}1407}$941TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)941amber*	this work	2
pβ'$_{821\text{-}1407}$942TAG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407)942amber*	this work	2
pβ'$_{821\text{-}1407}$ 941TAG942GCG943GCG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407) 941amber942gcg943gcg*	this work	3
pβ'$_{821\text{-}1407}$ 942TAG943GCG	Km^R; ori-pBR322; ori-f1; *φ10P-rpoC(821-1407) 942amber943gcg*	this work	3
pEcABC(1-878)Z	Ap^R; ori-pMB1; *φ10P-rpoA-rpoB-rpoC(1-878)-rpoZ*	this work	4
pCDF-rpoC(821-1407)-CH6	Sm^R; ori-CDF; *φ10P-rpoC(821-1407)(CH6)*	this work	4
pCDF-rpoC(821-1407) 942TAG943GCG-CH6	Sm^R; ori-CDF; *φ10P-rpoC(821-1407)942amber943gcg(CH6)*	this work	4

* Methods: 1, first method: *in vitro* reconstitution of RNAP with labelled full-length β'. 2, revised method: *in vitro* reconstitution of RNAP with labelled split β'. 3, further revised method: *in vitro* reconstitution of RNAP with labelled split β' containing Ala substitution. 4, final method: *in vivo* assembly of RNAP with labelled split β' containing Ala substitution.

3.3. Labelled RNAP

3.3.1. First method: *in vitro* reconstitution of RNAP with labelled full-length β'

Flag-αNTD^{I}-GSGGSG-αNTD^{II}. A fusion protein comprising an N-terminally Flag-tagged *E. coli* RNAP α subunit N-terminal domain (residues 1-235; αNTD^{I}), followed by GlySerGlyGlySerGly, followed by a second *E. coli* RNAP α subunit N-terminal domain (residues 1-235; αNTD^{II}) was prepared as in Chakraborty et al., 2012.

β subunit. Inclusion bodies containing *E. coli* RANP wild type β subunit were prepared as in Naryshkin et al. 2001.

β' subunit and Cou-labelled β'derivative. Inclusion bodies containing *E. coli* RANP wild type β' subunit were prepared as in Naryshkin et al. 2001. Inclusion bodies containing *E. coli* RANP Cou-labelled β' subunit were prepared as follow: *E. coli* strain BL21(DE3) was transformed with plasmids pEVOL-CouA and pET21a-rpoCXXXTAG-CH6 (XXX = 933, 934, 935, 936, 937, 938, 939, 940, 941, or 942). A single colony of the resulting transformants was inoculated into 50 ml LB broth containing 40 μg/mL chloramphenicol and 100 μg/mL ampicillin. Cultures were incubated 12 h at 37°C with shaking. Overnight culture was inoculated at ratio of 1:100 to 2 L LB broth containing 40 μg/mL chloramphenicol, 100 μg/mL ampicillin and 2 mM Cou. Cells were grown in dark at 37°C with shaking until OD600 reached 0.6-0.7, and arabinose was added to 0.02%. When OD600 reached 1, IPTG was added to 1 mM into the culture. Protein production was induced at 37°C with shaking for 12 h. Cells were collected and Cou-containing β' subunit was prepared using the same procedure as in preparation of wild type β' subunit. Cou-labelled protein was protected from ambient light.

ω subunit. Inclusion bodies containing *E. coli* RANP wild type ω subunit were prepared as in Chakraborty, et al., 2015.

RNAP core enzyme and Cou-labelled RNAP core enzyme derivative. Unlabelled RNAP core enzyme was prepared by reconstitution from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, β' and ω. Cou-labelled RNAP core enzyme derivatives were reconstituted from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, Cou-labelld β' and ω. Reaction mixtures (50 mL) containing 1.7 mg (30 nmol) Flag-αNTD^{I}-GSGGSG-αNTD^{II} , 4.5 mg (30 nmol) β, 2.2 mg (15 nmol) unlabelled or Cou-labelld β', and 1.5 mg (150 nmol) ω in denaturation buffer (50 mM Tris-HCl pH 7.9, 6 M guanidine-HCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM EDTA, 10 mM DTT, and 10% glycerol), were dialyzed at 4°C for 12 h against 2 L reconstitution buffer (50 mM Tris-HCl pH 7.9, 200 mM KCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1mM EDTA, 5 mM β-mercaptoethanol, and 20% glycerol), further dialyzed at 4°C for 12 h against 2 L reconstitution buffer, and further dialyzed at 4°C for 12 h against 2 L reconstitution buffer without β-mercaptoethanol. Following dialysis, insoluble material was removed by centrifugation (20,000×g; 30 min at 4°C). Supernatant was applied to column packed with ANTI-FLAG M2 affinity gel (3 mL solid, Sigma-Aldrich) and pre-equilibrated with buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1 mM EDTA and 5% glycerol as per manufacturer's instructions. Columns were washed with 30 ml above buffer and eluted with 15 ml of the same buffer containing 0.1 mg/mL FLAG peptide (Sigma-Aldrich). Eluted sample was further purified by fast protein liquid chromatography (FPLC) on Mono-Q column according to procedures as in Mukhopadhyay, et al., 2003.

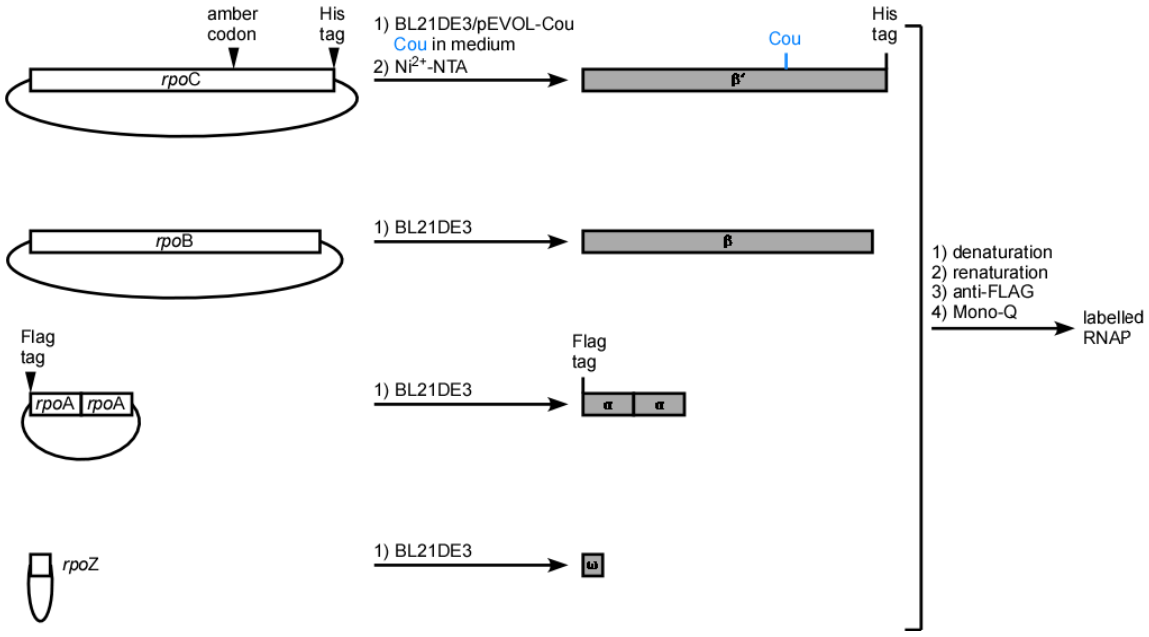


Figure 9. First method of preparing labelled RNAP: *in vitro* reconstitution of RNAP with labelled full-length β'

3.3.2. Revised method: *in vitro* reconstitution of RNAP with labelled split β'

Split β' subunit fragments. Inclusion bodies containing unlabelled *E. coli* RNAP split β' subunit fragments $\beta'_{1\text{-}878}$ and $\beta'_{821\text{-}1407}$ were prepared as in Nikolai, et al., 2001. Inclusion bodies containing Cou-labelled fragments $\beta'_{821\text{-}1407}$ were prepared as follows: *E. coli* strain BL21(DE3) was transformed with plasmids pEVOL-CouA and $p\beta'_{821\text{-}1407}$XXXTAG (XXX = 935, 936, 937, 938, 939,940, 941, or 942). A single colony of the resulting transformants was inoculated into 50 ml LB broth containing 40 μg/mL chloramphenicol and 40 μg/mL kanamycin. Cultures were incubated 12 h at 37°C with

shaking. Overnight culture was inoculated at ratio of 1:100 to 2 L LB broth containing 40 μg/mL chloramphenicol, 40 μg/mL kanamycin and 2 mM Cou. Cells were grown in dark at 37°C with shaking. When OD600 reached 1, arabinose was added to 0.02%, IPTG was added to 1 mM into the culture. Protein production was induced at 37°C with shaking for 12 h. Cells were collected and Cou-containing $\beta'_{821\text{-}1407}$ fragment was prepared using the same procedure as in preparation of unlabelled $\beta'_{821\text{-}1407}$ fragment. Cou-labelled protein was protected from ambient light.

Other *E. coli* RNAP subunits. Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, and ω were prepared according to procedures as in section 3.3.1.

RNAP core enzyme with split β' and Cou-labelled RNAP core enzyme derivative with split β'. Unlabelled RNAP core enzyme with split β' was prepared by reconstitution from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, $\beta'_{1\text{-}878}$, $\beta'_{821\text{-}1407}$ and ω. Cou-labelled RNAP core enzyme derivatives with split β' were reconstituted from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, $\beta'_{1\text{-}878}$, Cou-labelld $\beta'_{821\text{-}1407}$ and ω. Reaction mixtures (60 mL) contained: 0.9 mg (15 nmol) Flag-αNTD^{I}-GSGGSG-αNTD^{II} , 4.5 mg (30 nmol) β, 5.9 mg (60 nmol) $\beta'_{1\text{-}878}$, 3.8 mg (60 nmol) unlabelled $\beta'_{821\text{-}1407}$ or Cou-labelld $\beta'_{821\text{-}1407}$, and 1.5 mg (150 nmol) ω in denaturation buffer (50 mM Tris-HCl pH 7.9, 6 M guanidine-HCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM EDTA, 10 mM DTT, and 10% glycerol). Reconstitution and following purification were carried out as described in section 3.3.1.

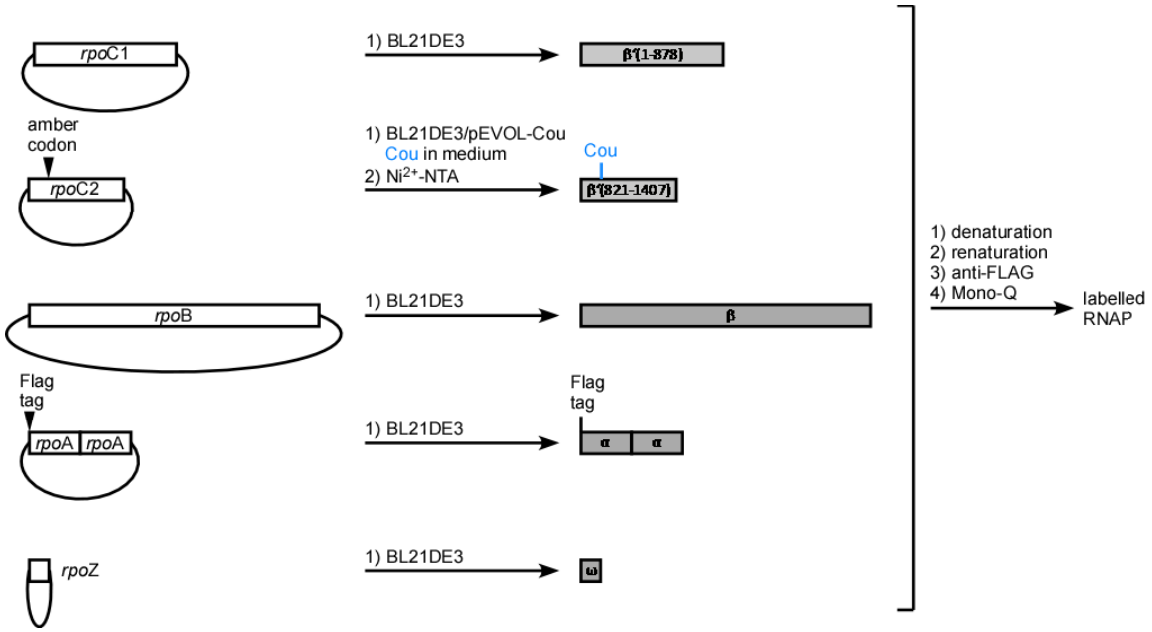


Figure 10. Revised method of preparing labelled RNAP: *in vitro* reconstitution of RNAP with labelled split β'.

3.3.3. Further revised method: *in vitro* reconstitution of RNAP with labelled split β' containing Ala substitution

Split β' subunit fragments. Inclusion bodies containing unlabelled *E. coli* RNAP split β' subunit fragments $\beta'_{1\text{-}878}$ and $\beta'_{821\text{-}1407}$ were prepared as in Nikolai, et al., 2001. Inclusion bodies containing Cou-labelled $\beta'_{821\text{-}1407}$ fragment with mutations were prepared as in section 3.3.2 except using plasmid $p\beta'_{821\text{-}1407}$ 941TAG942GCG943GCG or $p\beta'_{821\text{-}1407}$ 942TAG943GCG in place of $p\beta'_{821\text{-}1407}$XXXTAG (XXX = 935, 936, 937, 938, 939, 940, 941, or 942).

Other *E. coli* RNAP subunits. Flag-αNTD^{I}-GSGGSG-αNTD^{II} , β, and ω were prepared according to procedures as in section 3.3.1.

RNAP core enzyme with split β' and Cou-labelled RNAP core enzyme derivative with split β'. Unlabelled RNAP core enzyme with split β' was prepared by reconstitution from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, β'$_{1-878}$, β'$_{821-1407}$ and ω. Cou-labelled RNAP core enzyme derivatives with split β' were reconstituted from Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, β'$_{1-878}$, β'$_{821-1407}$ 941Cou942Ala943Ala or β'$_{821-1407}$ 942Cou943Ala and ω. Reaction mixtures (60 mL) contained: 0.9 mg (15 nmol) Flag-αNTD^{I}-GSGGSG-αNTD^{II} , 4.5 mg (30 nmol) β, 5.9 mg (60 nmol) β'$_{1-878}$, 3.8 mg (60 nmol) unlabelled β'$_{821-1407}$ or Cou-labelld β'$_{821-1407}$, and 1.5 mg (150 nmol) ω in denaturation buffer (50 mM Tris-HCl pH 7.9, 6 M guanidine-HCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM EDTA, 10 mM DTT, and 10% glycerol). Reconstitution and following purification were carried out as described in section 3.3.1.

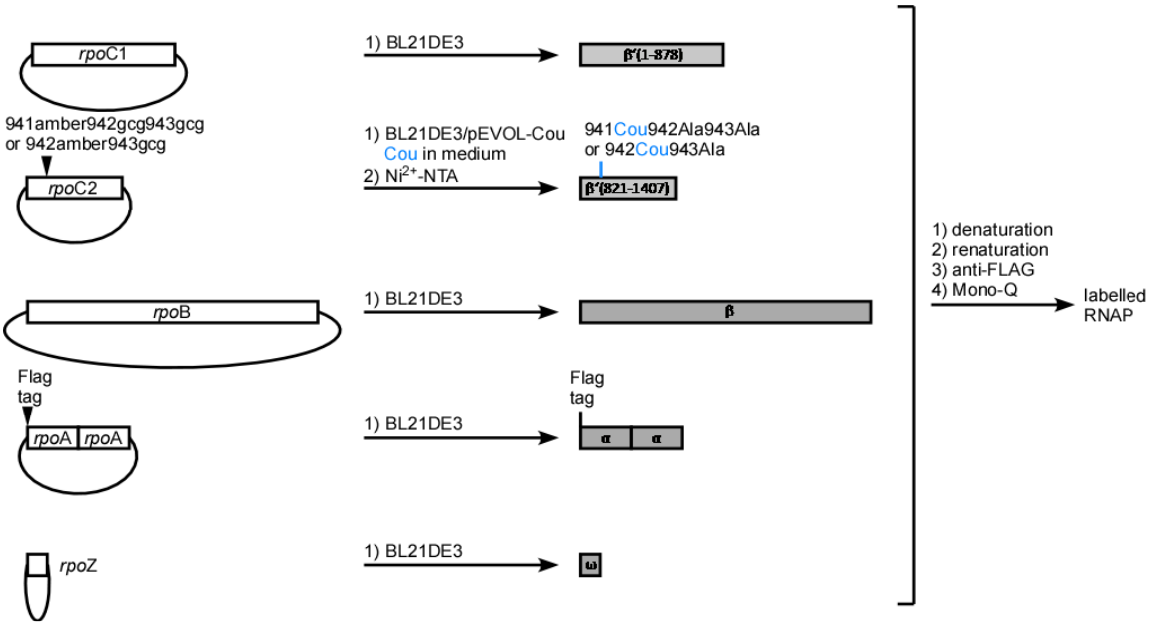


Figure 11. Further revised method of preparing labelled RNAP: *in vitro* reconstitution of RNAP with labelled split β' containing Ala substitution.

3.3.4. Final method: *in vivo* assembly of RNAP with labelled split β' containing Ala substitution

Unlabelled *E. coli* RNAP core enzyme with split β'. *E. coli* strain BL21(DE3) was transformed with plasmids pEcABC(1-878)Z and pCDF-C(821-1407)-H6. A single colony of the resulting transformants was inoculated into 50 ml LB broth containing 100 μg/mL ampicillin and 50 μg/mL streptomycin. Cultures were incubated 12 h at 37°C with shaking. Overnight culture was inoculated at ratio of 1:100 to 2 L LB broth containing 100 μg/mL ampicillin and 50 μg/mL streptomycin. Cells were grown at 37°C with shaking until OD600 reached 0.6, and IPTG was added to 1 mM into the culture. Protein

production was induced at 37°C with shaking for 3 h followed by incubation at 16°C with shaking for 12 h. Cells were collected by centrifugation at 5000 × g and 4°C for 20 min , and resuspended in 50 mL of lysis buffer (50 mM Tris pH 7.9, 5% glycerol, 2 mM EDTA, 5 mM DTT, 200 mM NaCl). A cOmplete™ EDTA-free protease inhibitor cocktail tablet (Sigma Aldrich) was added as recommended by the manufacturer, and the cells were disrupted with a homogenizer (EmulsiFlex-c5, AVESTIN, Canada). The crude lysate was cleared by centrifugation at 20,000 × g and 4 °C for 20 min. The supernatant was transferred to a clean beaker in which Polymin P (10% in water, pH 7.8) was added to 0.6% (add 6.4 mL / 100 mL lysate) with stirring at 4 °C for 30 min. Clear lysate turned cloudy upon addition of Polymin P. The RNAP-containing precipitate was collected by centrifugation at 20,000 × g and 4 °C for 15 min. The precipitate pellet was resuspended by a dispenser in 50 mL washing buffer I (20 mM Tris pH 7.9, 5% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 500 mM NaCl) and collected by centrifugation at 20,000 × g and 4 °C for 15 min. The washing step was repeated in 50 mL washing buffer II (20 mM Tris pH 7.9, 5% glycerol, 500 mM NaCl) and solid particles were collected by centrifugation at 20,000 × g and 4 °C for 15 min. The pellet was resuspended in extracting buffer (20 mM Tris pH 7.9, 5% glycerol, 1 M NaCl) by a dispenser and incubated in ice for 15 min. Insoluble material was removed by centrifugation at 20,000 × g and 4 °C for 15 min. Protein was precipitated by addition of $(NH_4)_2SO_4$ (29.1 g / 100 mL liquid) to the supernatant. RNAP-containing precipitate was collected by centrifugation at 20,000 × g and 4 °C for 30 min and supernatant was removed. The precipitate can be stored temporarily at -80 °C if further purification was not carried out immediately. The protein precipitate was dissolved with 20 mL binding buffer (20 mM Tris-HCl pH 7.9, 250 mM

NaCl, 5% glycerol, 5 mM 2-mercaptoethanol) and insoluble material was removed by centrifugation at 20,000 × g and 4 °C for 15 min. The protein solution was incubated with pre-equilibrated Ni-NTA resin (5 mL solid volume) at 4 °C for 12 h in a rolling capped tube. The suspension was re-packed into an Econo-Pac chromatography column (Bio-Rad # 7321010) and flowed through the column. Column was washed with 25 mL binding buffer followed by 25 mL binding buffer with 10 mM imidazole. Protein was eluted with 25 mL binding buffer with 250 imidazole. RNAP was further purified by FPLC on Mono-Q column according to procedures in Minakhin, et al., 2001, and stored in 10 mM Tris-HCl pH 7.9, 0.2 M NaCl, 0.05 mM EDTA, 0.5 mM DTT, and 50% glycerol at -20°C.

Cou-labelled *E. coli* RNAP core enzyme with split β'. To prepare Cou-labelled *E. coli* RNAP core enzyme, *E. coli* strain BL21(DE3) was transformed with plasmids pEcABC(1-878)Z, pCDF-C(821-1407, 941TAG942GCG943GCG)-H6 or pCDF-C(821-1407, 942TAG943GCG)-H6, and pEVOL-Cou. A single colony of the resulting transformants was inoculated into 50 ml LB broth containing 100 μg/mL ampicillin, 50 μg/mL streptomycin, and 40 μg/mL chloramphenicol. Cultures were incubated 12 h at 37°C with shaking. Overnight culture was inoculated at a ratio of 1:100 to 2 L LB broth containing 100 μg/mL ampicillin, 50 μg/mL streptomycin, and 40 μg/mL chloramphenicol. Cells were grown at 37°C with shaking until OD600 reached 0.3-0.4, at which point L-(+)-arabinose (20% in water) was added to 0.2% into the culture. Cells were incubated at 37°C with shaking until OD600 reached 0.6, and IPTG was added to 1 mM into the culture. Protein production was induced at 37°C with shaking for 3 h followed by incubation at 16°C with shaking for 12 h. Cells were collected and protein

was purified by using the same procedure in preparation of unlabelled *E. coli* RNAP with split β'. Cou-labelled protein was protected from ambient light.

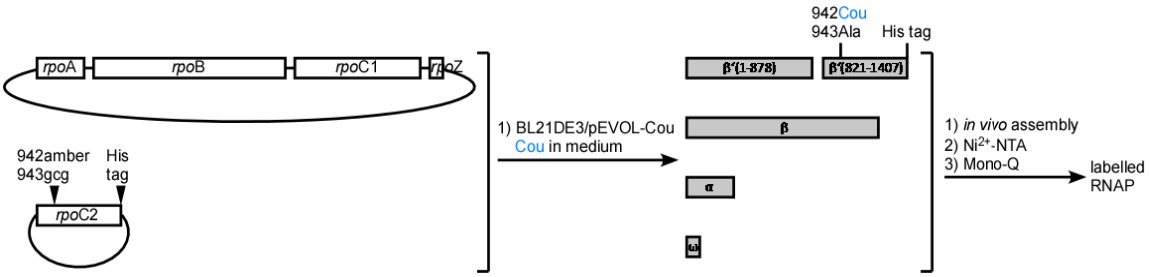


Figure 12. Final method of preparing labelled RNAP: *in vivo* assembly of RNAP with labelled split β' containing Ala substitution.

3.4. Transcription assay

Transcriptional activities of unlabelled RNAP and of Cou-labelled RNAP derivatives were determined with RiboGreen fluorescence-based transcription assay (Kuhlman, et al., 2004). Unlabelled RNAP or Cou-labelled RNAP core were incubated with *E. coli* principal σ factor σ^{70}, on ice for 10 min to generate RNAP holoenzyme in transcription buffer (50 mM Tris-HCl, pH8.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 10 μg/mL bovine serum albumin or BSA). DNA fragment (0.4 μM, 1 μL) containing the bacteriophage T4 N25 promoter (positions -72 to +367 PCR amplified from pARTaqN25-340-tR2, Liu, et al., 2007) was added to the protein and incubated at 37 ºC for 10 min to form RP_o. A mixture solution (2 μL) of nucleoside triphosphates (NTP) containing 1 mM of each NTP was added to RP_o and incubated at 37 ºC for 60 min. Each reaction mixture contained: 75 nM RNAP core enzyme, 300 nM σ^{70}, 20 nM DNA fragment, 100 μM ATP, 100 μM GTP, 100 μM UTP, and 100 μM CTP. Negative control has all the components except RNAP core enzyme and σ^{70}. Transcription was quenched by addition of $CaCl_2$ (5 μM, 1μL) and DNase I (RNase-free, 2 U/μL, 1 μL, Thermo Fisher Scientific Inc.), followed by incubation at 37 ºC for 90 min. Product RNA was quantified by addition of 100 μl Quant-iT™ RiboGreen® RNA Reagent (ThermoFisher Scientific; 1:500 dilution in 10 mM Tris-HCl pH 8.0, 1 mM EDTA), followed by incubation at 22°C for 10 min, followed by measurement of fluorescence intensity [excitation wavelength = 485 nm and emission wavelength = 535 nm; GENios Pro microplate reader (Tecan)].

3.5. Labelled TEC

Formation of TEC

E. coli RNAP core enzyme (unlabelled or Cou-labelled) was incubated with nucleic acid scaffold (unlabelled or fluorescein-labelled) 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, or BSA) at 37°C for 15 min.

Isolation of TEC

Resulting TEC was isolated by applying the mixture solution to a 4–20% Mini-PROTEAN® TBE Gel (10 well, 50 μL, Bio-Rad #4565094), followed by electrophoresis in TBE at 8 V/cm and 22°C for 1 h. Fluorescently-labelled TEC was visualized with fluorescence imager (Typhoon 9400 Variable Mode Imager, excitation at 532 nm) and TEC-containing gel regions were excised. Gel regions containing unlabelled TEC were identified by alignment with gel regions containing labelled TEC in the next running lane on the same gel, and were excised.

Recovery of TEC

TEC was recovered from gel slices by electroelution, which involves four steps:

(i) Place the TEC-containing gel slices in D-Tube™ Dialyzer (EMD Millipore), fill the tube with TBE and cap the tube.
(ii) Mount the tube in a rack (D-Tube™ Electroelution Accessory Kit, Novagen), place the rack in a horizontal electrophoresis gel box, and carry out electrophoresis in TBE at 8

V/cm and 22°C for 40 min. Switch connections from cathode and anode to the power supply and run for additional 1 min to prevent TEC stick on membrane of the dialyzer.
(iii) Transfer the liquid from dialyzer to a microcentrifuge tube and remove gel residues by centrifugation. Apply the supernatant to an Amicon® Ultra 0.5 mL Centrifugal Filters (30 KDa) and concentrate TEC by centrifugation.
(iv) Change buffer to transcription buffer by diluting concentrated TEC solution with transcription buffer and concentrating again. Repeat dilution and concentration three times.

3.6. Measurements of fluorescence spectrum

Fluorescence emission spectra of TEC were measured in the absence of and the presence of corresponding substrate NTP or corresponding NTP analogue. A saturating concentration of 1 M was used for all NTPs and NTP analogues to permit ~100% binding efficiency in all fluorescence measurements unless stated otherwise. Purified TEC was contained in a submicro fluorometer cuvette (Starna Inc.; catalog number 16.40F-Q-10). After incubation at 37°C for 5 min, fluorescence emission spectra of TEC were measured upon excitation at 330 nm (QuantaMaster, Photon Technology International, Inc.). Emission spectra were scanned at range of 400-600 nm and step size of 1 nm with FeliX software (Photon Technology International, Inc.). Corresponding substrate NTP or corresponding NTP analogues were added to TEC and incubated at 37°C for 10 min.

Fluorescence emission spectra of TEC bound with NTP or analogue were measured at the same excitation wavelength and emission wavelength range.

3.7. Measurements of fluorescence lifetime

Fluorescence lifetime was measured by a time-domain method, time correlated single photon counting (TCSPC). Fluorescence decays of TEC were measured in the absence of and in the presence of corresponding substrate NTP or corresponding NTP analogue. Samples were prepared with an absorbance at 375 nm below 0.1 to avoid inner filter effects. Purified TEC was contained in a submicro fluorometer cuvette (Starna Inc.; catalog number 16.40F-Q-10). Fluorescence decays (NanoLED pulsed diode excitation wavelength 375 nm, emission wavelength 450 nm, FluoroCube, IBH, Inc.) of Cou-labelled TEC were measured after incubation for 5 min at 37°C. Excitation polarizer was vertically oriented, and emission polarizer was set at the magic angle (54.7° to the vertical) to remove polarization effects. A long pass filter (KV418, Schott, Inc.) was used in the emission filter holder to filter out short wavelength background signal. Corresponding substrate NTP or corresponding NTP analogues were added to TEC and incubated at 37°C for 10 min. Fluorescence decays of TEC bound with NTP or analogue were measured at the same conditions.

3.8. Measurements of fluorescence anisotropy

Steady-state fluorescence anisotropies were measured in solution containing donor-only or acceptor-only TEC at 37 ºC using QuantaMaster equipped with excitation emission polarizers (Photon Technology International, Inc.). Excitation and emission wavelengths were 375 nm and 450 nm for coumarin, and 495 nm and 520 nm for fluorescein. Monochromators were used minimize light scattering. Anisotropy (r) was calculated as in (Chen and Bowman, 1965):

$$r = (I_{VV} - GI_{VH})/(I_{VV} + 2GI_{VH}) \quad (10)$$

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:

$$G = I_{HV}/I_{HH} \quad (11)$$

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.

3.9. Measurements of fluorescence quantum yield

Quantum yield of Cou incorporated in RNAP was determined by comparison to a reference fluorophore with a well-known quantum yield. Quinine sulfate (Q = 0.577 in 0.1 M H_2SO_4 at 22 °C when λex = 350 nm, Eastman, et al., 1967) was used as the reference fluorophore for Cou. The quantum yield (Q) of a given sample was calculated using (Lakowicz, 1999):

$$Q = Q_R \frac{I}{I_R} \frac{OD_R}{OD} \frac{n^2}{n_R^2} \quad (12)$$

where I is the integrated fluorescence intensity, n is the refractive index, and OD is the optical density. The subscript R refers to the reference fluorophore of known quantum yield. Low concentrations of samples were used so that the absorbance at 375 nm and above was less than 0.05 to avoid inner filter effect. In this work, 10 μM quinine sulfate in 0.1 M H_2SO_4 and 2 μM TEC in TB were used. Sample solutions were contained in a sub-micro fluorometer cuvette with 10-mm path length. OD was determined at absorption wavelength 375 nm using Lamda 25 UV/VIS spectrometer (PerkinElmer, Inc.) Fluorescence intensity curves were recorded upon excitation at 375 nm with QuantaMaster (Photon Technology International, Inc.), and emission curves were integrated using Microsoft Excel. Since both the reference fluorophore and Cou-containing sample were measured in aqueous solution, it is assumed that $n^2/n_R^2 = 1$.

3.10. Distance-restrained rigid body docking

Distance-restrained rigid body docking was performed individually for the open TL and closed TL using the FPS program (Kalinin, et al., 2012). Each run of docking needed the following data: (i) structures of individual molecules; (ii) labelling position data; (iii) distance restraints.

Docking for either open TL or closed TL involved two docking partners: a TEC with deletion of TL, and a single amino acid residue of Ala representing position β'942. As the first docking partner, the TEC consists of RNAP core and a nucleic-acid scaffold, in which the TL (residues β'926-944 and β'1133-1137) was removed to make space for potential positions of the second docking partner. In SI3-absent docking, RNAP core from crystal structure 4YLN with deletion of TL and deletion of SI3 was used to construct the TEC (SI3-absent 4YLN-derived TEC) for both open and closed TL. In SI3-present docking, RNAP core from crystal structure 4YLN with deletion of only TL was used to construct the TEC (SI3-present 4YLN-derived TEC) for closed TL, and RNAP core from crystal structure 5BYH with deletion of only TL was used to construct the TEC (SI3-present 5BYH-derived TEC) for open TL. The nucleic acid scaffold (Figure 7C) was constructed by merging two parts: the upstream portion (position -15 to +14) of the scaffold was a copy of the scaffold in the AMPcPP-bound TEC structure with a deletion of the last pair of downstream DNA nucleotides (Vassylyev, et al., 2007b, PDB 2O5J, numbering of nucleotides is different from that in this work), and the downstream portion of the scaffold (positions +15 to +25) was obtained by extending the downstream DNA duplex by a B-form DNA whose sequence was designed for the preliminary biochemical characterization of the TEC structure (Kashkina, et al., 2006). The second docking

partner was a single Ala residue representing residue β'942 omitting the dye and linker. Each docking structure was prepared into a .pdb file.

In the labelling position file, all labelling positions in either TL or DNA were specified with the following data: labelling position name; molecule name; donor or acceptor or unspecified; linker length; linker width; AV3 (or AV1); dye radius 1; dye radius 2; dye radius 3; attachment point in corresponding molecule (atom ID). Dyes were approximated by a sphere with defined radii, which were estimated from the physical dimensions of the dye molecules with Chem3D. The connecting linker was modeled as a flexible cylinder with length (L_{link}) measured from the attachment point (the beta carbon for TL labelling, and the base atom for DNA labelling) to the dye center and width (W_{link}) using typical value of 4.5 Å (Muschielok, et al., 2008; Kalinin, et al., 2012). Choice of AV1 or AV3 means accessible volume simulation with on dye radius or with three dye radii, and AV3 was used in this work. Atom IDs were manually extracted from .pdb files. The main dye and linker parameters were shown in Table 3. The results of systematic FRET distance measurements addressing aim (i) were used as distance restrains. The uncertainties of restrains (ΔR_{DA} in Eq. (7)) were estimated to 15% of the corresponding measured distance (Knight, et al., 2005).

Model generation was performed in a two-step procedure: search and refinement. In the search step, a large number (e.g. 1,000 in this work) of model solutions were generated from random configurations of the docking bodies, excluding those with steric clashes (used clash tolerance of 2 Å in this work, Eq. (8)). Model solutions resulted in this step were clustered into groups of solutions with similar χ_r^2 values and low r.m.s. deviation. Top 50 solutions in the groups of interest were refined in the second step with clash

tolerance of 0.5 Å. In the refinement step, AVs were recalculated to account for possible steric clashes between the dyes and parts of the biomolecule structure that the dyes were not attached to, and the new mean positions were used to optimize the structure.

Table 3. Dye and linker parameters used for distance-restrained docking

LP NAME*	**Dye**	**L_{link} (Å)**	**W_{link} (Å)**	**$R_{dye(1)}$ (Å)**	**$R_{dye(2)}$ (Å)**	**$R_{dye(3)}$ (Å)**
942Ala_CB	Cou	4.4	4.5	3.9	2.5	0.8
DNA_-6_C_C5	fluorescein	17.2	4.5	5.2	4.6	1.5
DNA_-4_C_C5	fluorescein	17.2	4.5	5.2	4.6	1.5
DNA_-2_C_C5	fluorescein	17.2	4.5	5.2	4.6	1.5
DNA_+8_T_C7	fluorescein	15.9	4.5	5.2	4.6	1.5
DNA_+10_T_C7	fluorescein	15.9	4.5	5.2	4.6	1.5
DNA_+12_T_C7	fluorescein	15.9	4.5	5.2	4.6	1.5
DNA_+14_C_C5	fluorescein	17.2	4.5	5.2	4.6	1.5
DNA_+15_T_C7	fluorescein	15.9	4.5	5.2	4.6	1.5
DNA_+16_G_C8	fluorescein	13.6	4.5	5.2	4.6	1.5
DNA_+18_A_C8	fluorescein	13.6	4.5	5.2	4.6	1.5
DNA_+20_T_C7	fluorescein	15.9	4.5	5.2	4.6	1.5

***** LP NAME: name of labelling position. CB: beta carbon. Labelling positions in DNA were named by the position, nucleobase, and the attachment atom in order, e. g. DNA_-6_C_C5**.**

4. Results

4.1. Preparation of Cou-labelled RNAP derivatives

In attempts to prepare Cou-labelled RNAP, four methods have been used: the first method, the revised method, the further revised method, and the final method. They were adapted based on previous method, or subsequently developed in this work to improve incorporation efficiency of Cou, transcriptional activity of labelled RNAP, and monodispersity of labelled RNAP.

4.1.1. First method: incorporation of Cou into full-length β' and *in vitro* reconstitution of RNAP

A method analogous to this method but using a different unnatural amino acid was successfully applied to prepare fluorescently labelled RNAP in our laboratory (Chakraborty, et al., 2012). Thus this method was adapted based on the previous one and served as the starting point for preparation of Cou-labelled RNAP in this work.

Incorporation efficiency of Cou

Site-specific incorporation of the unnatural amino acid Cou into full-length β' subunit was tested at each of ten amino acid positions on TL (β' residues 933-942). The results show that the unnatural amino acid was able to been incorporated into each of these positions. Fluorescence of Cou was found only in full-length β', but not in any other proteins including truncated β' fragment in cell lysate, which indicates that the

incorporation is selectively coupled to suppression of the amber stop codon at target position. Undesired incorporation of endogenous amino acids by the *Mj*tRNA$^{opt}_{CUA}$ / *Mj*CouRS was also assessed by inducing protein expression in the absence of Cou. Results showed that truncated β' fragment was found, but no full-length β' was observed. It suggested that almost all of the resulting full-length β' was labelled at the target position with ~100% labelling specificity and ~100% labelling efficiency. However, the yield of full-length β' was unacceptably low. A large amount of truncated β' fragment was found due to suppression failure. Suppression efficiency, defined as the ratio of the amount of full-length protein to the total amount of full-length and truncated protein, is only ~10-20%.

Transcriptional activity of Cou-labelled RNAP derivatives

Unlabelled RNAP core enzyme and Cou-labelled RNAP core enzyme derivatives were produced by *in vitro* reconstitution from isolated subunits. Unlabelled RNAP core enzyme consists of Flag-αNTD^{I}-GSGGSG-αNTD^{II} , β, β' and ω, while Cou-labelled RNAP core enzyme consists of Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, Cou-labelld β' and ω. Their transcription activities were determined by RiboGreen fluorescence-based transcription assay. Compared with wild type RNAP core enzyme (Epicentre) generated from *in vivo* assembly, reconstituted unlabelled RNAP core enzyme has a comparable transcriptional activity (~100%), while reconstituted Cou-labelled RNAP derivatives have ~10-15% activity. The results indicated the incorporation of Cou into TL impairs the RNAP enzyme activity.

The ability of forming elongation complex was not tested with the resulting RNAPs due to their extremely low transcriptional activities.

In summary, the first method resulted in unacceptably low incorporation efficiencies of Cou and unacceptably low activities of resulting RNAP derivatives. Therefore, this method must be revised.

4.1.2. Revised method: incorporation of Cou into split β' and *in vitro* reconstitution of RNAP

This method was tried because we had observed that incorporation efficiency depends on distance from translation start, with short distances giving higher efficiency. To shorten the distance, split β' fragment was used for TL labelling. Split sites in β' and β, at which the subunits are split into smaller polypeptides that are encoded in multiple genes, naturally exist in some archaebacteria and chloroplasts, and have been demonstrated not interfere with the catalytic activity of *E. coli* RNAP (Severinov, et al., 1996). Herein split fragments β' 1-878 and β' 821-1407 were used to replace the full-length β', and Cou was introduced into the target positions in β' 821-1407 so that the distance from translation start to the labelling site was shortened by 820 amino acids. This split site in β' is the one precedes the TL region within the shortest distance (~110 amino acids), so that the corresponding truncated β' due to amber-suppression failure is the smallest in size.

Incorporation efficiency of Cou

Site-specific incorporation of Cou into C-terminal split β' fragment was tested at each of eight amino acid positions on TL (β' residues 935-942). Consistent with the results of the preceding method, the unnatural amino acid was able to been incorporated into each of these positions with ~100% labelling specificity and ~100% labelling efficiency. In contrast with the preceding method, the revised method showed a substantial increase in incorporation efficiencies of Cou (~40-50% versus ~10-20%) in all tested labelling positions. Consistent with the results, better incorporation of the unnatural amino acid *p*-azido-L-phenylalanine was observed at amber codon closer to the N-terminus of the eDHFR protein using a mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase and $tRNA^{Tyr}$ pair (Goerke, et al., 2009).

Transcriptional activity of Cou-labelled RNAP derivatives

Unlabelled RNAP core enzyme consisting of Flag-αNTDI-GSGGSG-αNTDII , β, β'_{1-878} , $\beta'_{821-1407}$ and ω, and Cou-labelled RNAP core enzyme consisting of Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, β'_{1-878} , Cou-labelld $\beta'_{821-1407}$ and ω, were prepared from reconstitution *in vitro*. The results of transcription assay showed that reconstituted wild-type RNAP containing unlabelled split β' is indistinguishable from wild-type RNAP prepared by *in vivo* assembly, but the reconstituted Cou-labelled RNAP derivatives with split β' remain ~10-25% transcriptional activity. Consistent with the preceding method, the presence of Cou prevents RNAP from acting properly. The activities of labelled RNAP derivatives containing split β' are slightly higher than that of labelled RNAP

derivatives containing full-length β', which is probably due to their slightly higher purities.

Ability of forming elongation complex

The abilities of forming TEC of the unlabelled and labelled RNAPs were tested by incubating the RNAP with a nucleic acid scaffold. The mixture was isolated by electrophoretic mobility shift assay in non-denaturing polyacrylamide gel, and was visualized by fluorescence imaging and by Coomassie blue staining to evaluate formation efficiencies. The results showed that Cou-labelled RNAP derivatives formed TEC with an extremely poor efficiency (<5%, shown as a smeared fade band on the gel) with most of the protein staying in the loading well of gel in an aggregated state. As positive controls, wild type RNAP (Epicentre) by *in vivo* assembly formed TEC with ~100% efficiency (shown as a sharp band on the gel), while reconstituted unlabelled RNAP formed TEC with ~50% efficiency. Coomassie blue staining revealed that some of unlabelled reconstituted RNAP stayed in the loading well of gel as aggregates, which suggested that the RNAP has a low monodispersity. The results were consistent with the observation during protein purification by anion exchange chromatography that the protein came out as an asymmetric broad peak in chromatogram. The low monodispersity is likely resulted from improper protein folding during reconstitution *in vitro*.

In summary, the revised method resulted in significantly higher incorporation efficiencies of Cou, however, yet unacceptably low activities and unacceptably low conformational monodispersities of the RNAPs urged the method to be further revised.

4.1.3. Further revised method: incorporation of Cou into split β' containing Ala substitution and *in vitro* reconstitution of RNAP

This method was tried because we had observed that activity of Cou-labelled RNAP depends on identity of the amino acid following Cou, with amino acids having short side chains giving higher activity.

Incorporation efficiency of Cou

Site-specific incorporation of Cou into split β' at position 941 with two following amino acids substituted with Ala (S942A and R943A), or at position 942 with one following amino acid substituted with Ala (R943A) was tested. The results showed that the incorporation specificities and efficiencies were similar with those of incorporation at the corresponding positions in split β' without Ala substitution.

Transcriptional activity of Cou-labelled RNAP derivatives

Cou-labelled RNAP core enzyme derivative consisting of Flag-αNTD^{I}-GSGGSG-αNTD^{II}, β, β'$_{1-878}$, β'$_{821-1407}$ 941Cou942Ala943Ala or β'$_{821-1407}$ 942Cou943Ala, and ω, were prepared by reconstitution *in vitro*. Transcription assay results showed that the RNAP derivative containing mutations 941Cou942Ala943 remains 58% activity and the RNAP derivative containing mutation 942Cou943Ala remains 85% activity compared with wild type RNAP. In both cases, the values are substantially higher than those of the corresponding RNAP derivatives without Ala substitutions, which are 22% and 18%, respectively.

Ability of forming elongation complex

The abilities of forming TEC by the two labelled RNAPs were tested, and the results showed both derivatives were able to form TEC with a fair efficiency (30-40%, shown as a clear band on the gel), which is dramatically higher than that of the corresponding labelled RNAPs without Ala substitution (<5%). Consistent with the results, these two RNAP derivatives came out as less asymmetric and less broad peaks in chromatogram during FPLC purification. Unlabelled RNAP with split β' formed TEC in a slightly higher efficiency (~50%), which supports the results of transcription assay that the Cou-labelled RNAP derivatives maintain most of the functions of unlabelled RNAP. In cases of both unlabelled and labelled RNAPs, a large portion of the protein formed aggregates likely due to improper folding during reconstitution.

In summary, the further revised method resulted in significantly higher incorporation efficiencies of Cou and significantly higher activities of labelled RNAP, however, the relatively low conformational monodispersities of labelled RNAP prevented this method from being used in this work.

4.1.4. Final method: incorporation of Cou into split β' containing Ala substitution and *in vivo* assembly of RNAP

This method was tried because we had observed that RNAP prepared from *in vivo* assembly is more likely to have a higher transcriptional activity.

Incorporation efficiency of Cou

Unlabelled RNAP core enzyme and Cou-labelled RNAP derivative with mutation R943A were produced by co-overexpression of subunits and *in vivo* assembly of RNAP. Yield of unlabelled RNAP is 3 mg/L culture, and yield of Cou-labelled RNAP mutant is 1.2 mg/L culture, which was substantially higher than the yield by a typical scale of reconstitution preparation (0.5 mg/L). Suppression efficiency is estimated to be 40%, which is similar with that of incorporation into $\beta'_{821\text{-}1407}$ when overexpressed individually as inclusion bodies. Denaturing polyacrylamide gel electrophoresis of purified Cou-labelled RNAP followed by gel fluorescence imaging verified that the fluorescent unnatural amino acid has been incorporated into $\beta'_{821\text{-}1407}$ fragment, and that the unnatural amino acid is present only in $\beta'_{821\text{-}1407}$, not in $\beta'_{1\text{-}878}$ or other RNAP subunits, which implied that the incorporation is site-specific.

Transcriptional activity of Cou-labelled RNAP derivatives

Cou-labelled RNAP core enzyme derivative containing mutations $\beta'_{821\text{-}1407}$ 942Cou943Ala remains 95% transcriptional activity compared with wild type RNAP, which implied that the labelled RNAP derivative and the unlabelled RNAP are almost indistinguishable in catalytic activity.

Ability of forming elongation complex

The Cou-labelled RNAP was found to form TEC in a good efficiency (80-90%, shown as a clear sharp band on the gel). Only a small portion (10-20%) of the protein aggregated and accumulated in the loading well. Unlabelled RNAP prepared by the same method exhibits a similar efficiency in forming TEC, which supported the conclusion from

transcription assay. Consistent with the result, this RNAP derivative came out as a symmetric and sharp peak in chromatogram during FPLC purification.

In summary, final method is better than the other methods in five ways: (i) higher yield of RNAPs; (ii) higher transcriptional activity of RNAPs; (iii) higher TEC-formation efficiency of RNAPs; (iv) higher monodispersity of RNAPs; (v) easier procedure. Instead of doing multiple transformations for individual subunits, preparations for individual subunits, dialysis for days, and purification for the reconstituted product, one just needs a single cycle of transformation, preparation and purification for the assembled protein. Therefore, this method was used to prepare Cou-free and Cou-labelled RNAPs for further experiments.

4.2. Formation and purification of TEC

RNAPs obtained from the final method were used to form TECs by incubating with a nucleic acid scaffold. The presence of non-functional Cou-labelled RNAP and excess labelled nucleic acid scaffold could disturb fluorescence measurements. Therefore, the formed TEC was isolated by electrophoretic mobility shift assay for following fluorescence analysis. Applying the incubated mixture of complexes to non-denaturing polyacrylamide gel electrophoresis allows good separation of the TEC from non-functional RNAP aggregates and from excess scaffold. Isolated TEC was recovered from the gel by electroelution, in which TEC molecules migrate in and eventually leave the gel slice by electrophoresis in non-denaturing conditions. Recovery efficiency was typically

higher than 80%. Quality control of the recovered TEC by denaturing polyacrylamide gel electrophoresis, followed by SYBR Gold nucleic acid staining and coomassie blue staining, indicated that the TEC exhibits a correct stoichiometry of individual RNAP subunits (or subunit fragments), DNA template and non-template strands, and RNA.

4.3. Determination of Förster radii of Cou-fluorescein pairs in TECs

Measured fluorescence anisotropies of Cou in TEC without and with bound NTP were 0.19 and 0.28, respectively (Table 4). Measured anisotropies of fluorescein linked to DNA ranged from 0.08 to 0.27, which were lower than the fundamental anisotropy of fluorescein (0.08 - 0.27 versus 0.38, Chen, et al., 1965) and lower than calculated anisotropy of fluorescein linked to molecules of ~0.5MDa and restricted in its local motion (0.08 - 0.27 versus 0.38, Cantor and Schimmel, 1980). The results indicate that the probes rotate on the time scale of the probe fluorescence lifetime, justifying the assumption of $\kappa^2 = 2/3$ in Eq. (6) for determination of Förster radii in follows.

Measured quantum yield of Cou in TEC without and with bound ATP were 0.67 and 0.87, respectively, which were comparable to the reported value of 0.63 for the free unnatural amino acid (Wang, et al., 2006). The difference in measured values is likely due to different microenvironment in open TL and closed TL.

Spectral overlap $J(\lambda)$ between Cou emission and fluorescein excitation was determined for each Cou-flourescein pair, and the results are shown in Table 5. Förster radius (R_0)

was obtained as in Eq. (6) between Cou in TL and fluorescein conjugated in each labelling site in DNA. Calculated R_0 are 48-49 Å for TEC and 49-51 for TEC with bound ATP, with slightly higher values for closed TL due to the higher value of corresponding quantum yield of Cou (0.87 verses 0.67). In both cases, the measured values consistent with the reported value of 47 Å (Yi, et al., 2011).

Table 4. Results of fluorescence anisotropy of the dye molecules in the TECs

Labelling sites	**Dye**	**λ_{Ex} (nm)**	**λ_{Em} (nm)**	***r***
β'942 *	Cou	375	450	0.19
β'942 **	Cou	375	450	0.28
DNA (-6)	fluorescein	495	520	0.21
DNA (-4)	fluorescein	495	520	0.27
DNA (-2)	fluorescein	495	520	0.25
DNA (+8)	fluorescein	495	520	0.18
DNA (+10)	fluorescein	495	520	0.19
DNA (+12)	fluorescein	495	520	0.19
DNA (+14)	fluorescein	495	520	0.09
DNA (+15)	fluorescein	495	520	0.11
DNA (+16)	fluorescein	495	520	0.13
DNA (+18)	fluorescein	495	520	0.08
DNA (+20)	fluorescein	495	520	0.08

* Measured in NTP-free TEC. ** Measured in NTP-bound TEC.

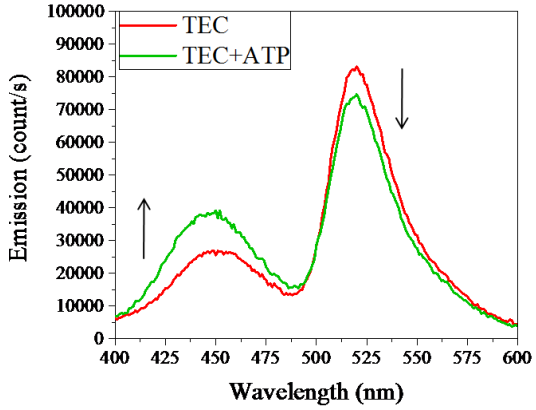
Table 5. Results of integrated spectral overlaps and Förster radii of Cou-fluorescein pairs in the TECs

Labelling site	TEC		TEC+ATP	
	$J(\lambda)$ ($M^{-1}cm^3$)	R_0 (Å)	$J(\lambda)$ ($M^{-1}cm^3$)	R_0 (Å)
DNA (-6)	1.4E-13	49	1.3E-13	51
DNA (-4)	1.4E-13	49	1.3E-13	51
DNA (-2)	1.4E-13	49	1.2E-13	50
DNA (+8)	1.4E-13	49	1.3E-13	51
DNA (+10)	1.3E-13	49	1.2E-13	50
DNA (+12)	1.3E-13	49	1.2E-13	50
DNA (+14)	1.3E-13	49	1.2E-13	50
DNA (+15)	1.4E-13	49	1.3E-13	51
DNA (+16)	1.3E-13	49	1.2E-13	50
DNA (+18)	1.3E-13	48	1.2E-13	50
DNA (+20)	1.2E-13	48	1.1E-13	49

4.4. Detection of TL closing/opening in solution

To directly determine whether TL closing and opening occurs in solution, fluorescence spectra were recorded before and after adding ATP to a doubly-labelled TEC with Cou incorporated at β'942 and fluorescein incorporated in DNA template strand position +20. A saturating concentration of ATP was used to allow ~100% conversion from ATP-free TEC to ATP-bound TEC. As shown in Figure 13, addition of ATP resulted in a significantly increased fluorescence intensity of donor and a significantly decreased fluorescence intensity of acceptor, which semiquantitatively indicated that FRET efficiency significantly decreases.

A



B

	TEC (open TL)	TEC+ATP (closed TL)
τ_D (ns)	6.22 ± 0.06	5.51 ± 0.03
τ_{DA_2+20} (ns)	3.83 ± 0.07	4.52 ± 0.09
E (%)	38.4 ± 2.4	18.0 ± 2.2
R (Å)	52 ± 1	63 ± 2

τ : fluorescence lifetime
E : FRET efficiency
R: donor-acceptor distance

Figure 13. TL closing and opening in solution.
(A) Increased fluorescence intensity of donor and decreased fluorescence intensity of acceptor caused by addition of ATP.
(B) Decreased FRET efficiency and increased probe-probe distance caused by substrate binding caused by addition of ATP.

The quantitative change in *E* was assessed by analyzing fluorescence lifetime of Cou. Measured fluorescence lifetime of Cou in Cou-only TEC without and with bound NTP,

τ_D (open) and τ_D (closed), were 6.22 ns and 5.51 ns, respectively, which are close to the reported lifetime of 5.70 ns for the same unnatural amino acid (Liu, et al., 2015). In the presence of acceptor incorporated in DNA position +20, the Cou lifetime dropped to 3.83 ns and 4.52 ns, respectively. Averaged FRET efficiencies obtained from the lifetime values were 38.4% for the ATP-free TEC and 18.0% for the ATP-bound TEC, which dramatically decreased by ~20%. The quantitative decrease in *E* corresponds to an increase in *R* of ~11 Å from 52 Å to 63 Å. Parallel experiments performed using an acceptor incorporated at a different site (positions 8, +10, +12, +14, +15, +16, or +18) on DNA downstream to the active center also shown increases in averaged donor-acceptor distance (increases of ~1 Å to ~ 12 Å). On the other hand, parallel experiments performed using acceptor incorporated at a different site (positions -6, -4 and -2) on DNA upstream to the active center shown decreases in averaged donor-acceptor distance (decreases of ~2 Å to ~6 Å). In all cases, the results clearly support the proposal that TL closing and opening occurs in solution, and that NTP binding causes the closing of TL. It is worth pointing out that the relative sizes of distance changes (ΔR) with positions +8 to +20 are basically consistent with our expectations based on crystal structures, with ΔR exhibiting an increasing trend and large changes being more likely occur at position +10, +15 and +20.

Table 6. Results of FRET-derived distance determination

Labelling site	τ_{DA} (open) (ns)	τ_{DA} (closed) (ns)	*E* (open) (%)	*E* (closed) (%)	*R* (open) (Å)	*R* (closed) (Å)	Δ*R* (Å)
DNA (-6)	5.03	4.13	19.2	25	63	61	-2
DNA (-4)	5.34	4.37	14.1	20.7	67	64	-3
DNA (-2)	4.77	3.14	23.4	43	59	53	-6
DNA (+8)	2.27	2.14	63.4	61.2	46	47	1
DNA (+10)	1.56	2.57	75	53.4	41	49	8
DNA (+12)	2.79	2.77	55.2	49.7	47	50	3
DNA (+14)	4.25	4.27	31.7	22.7	56	62	6
DNA (+15)	4.65	4.86	25.3	11.9	59	71	12
DNA (+16)	3.89	4.16	37.4	24.6	53	61	8
DNA (+18)	4.22	4.66	32.2	15.5	55	66	11
DNA (+20)	3.83	4.52	38.4	18	52	63	11

4.5. Analysis of effects of substrate identities on TL closing in solution

The second objective of this work is to determine whether TL conformation in solution depends on substrate identities, including base complementarity, triphosphate/diphosphate /monophosphate identity, and ribose/deoxyribose identity.

4.5.1. Effects of base complementarity

The effects of base complementarity on TL conformation in solution was assessed by analyzing fluorescence spectra of TEC with Cou incorporated at β'942 and fluorescein labelled in DNA template position +20 (scaffold 11 in Figure 7C) in the presence of non-complementary NTPs (GTP, UTP, or CTP), and compared to that with complementary ATP (Figure 14A, also Figure 15). The results showed that, in contrast with ATP, non-complementary GTP, UTP, and CTP caused only very small increases in fluorescence intensity of Cou, corresponding respectively to 4.8%, 4.0, and 1.5% of the increase caused by ATP (set as 100%), which are incomparable with that caused by ATP. Parallel experiments were performed using three different nucleic-acid scaffolds with the same labelling site (+20) but with the acceptor template nucleotide changed from T to C, A, or G, so that the complementary substrate NTP was changed from ATP to GTP, UTP or CTP respectively (Figure 14B, C, D). The results of all sets of experiments showed that, the complementary NTP causes a much larger increase (individually set as 100%) in the fluorescence intensity of Cou than the non-complementary NTPs do (most within range of -9.1% to 5.2% with one exception of 44.5% for ATP versus complementary GTP), which suggested that binding of non-complementary NTPs do not cause TL to fully close.

The non-specific effect of ATP was consistent with the observation based on RNET-seq that G-to-A error rates at the 3’ RNA ends increased in the absence of Gre factors (Imashimizu, et al., 2015). Consistent with the results of fluorescence intensity, measurements of fluorescence decays of Cou indicated that complementary substrate ATP causes a significant increase of 0.69 ns (from 3.83 ns to 4.52 ns) in Cou lifetime, while non-complementary NTPs do not cause significant increases (changes of 0.00 ns, -0.07 ns, and 0.03 ns with GTP, UTP, and CTP respectively) (Figure 15). These results all implied that TL closing occurs only when complementary substrate binds, in agreement with the proposals based on crystal structures of elongation complexes (Vassylyev, et al., 2007b; Wang, et al., 2006). Therefore, TL closing depends on base complementarity.

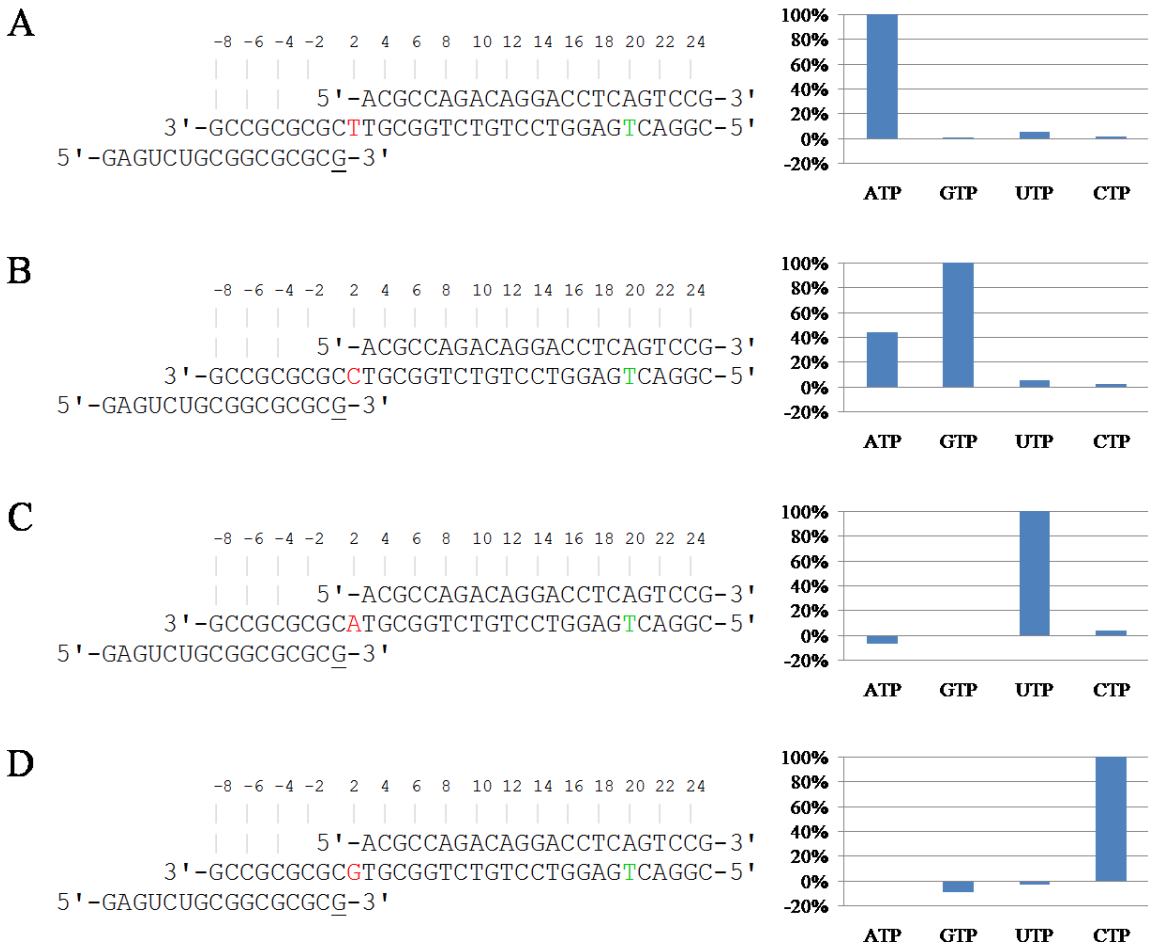


Figure 14. Effects of base complementarity on TL closing. In each panel, left: nucleic-acid scaffolds used to direct binding of ATP, GTP, UTP, or CTP, right: NTP-caused changes in fluorescence intensity of Cou in TEC formed with the scaffold on the left.

4.5.2. Effects of triphosphate / diphosphate / monophosphate identity

The effects of substrate phosphates on TL closing were also assessed by measurements of fluorescence intensity analogous to the previous section but using adenosine diphosphate

(ADP) or adenosine monophosphate (AMP) in place of non-complementary NTPs. Analysis of the fluorescence spectra indicated that ADP causes TL closing to an almost full extend compared with that by ATP (80.1% versus 100%), and that AMP does not cause TL closing (-12.0%, small negative value due to fluorescence fluctuation). Consistent with the results of fluorescence intensity, ADP induces an increase in fluorescence lifetime of Cou close to that by ATP (0.60 ns versus 0.69 ns), while AMP induces a much smaller increase (0.22 ns) (Figure 15). Therefore, TL closing depends on tri-/di- /mono-phosphate identity.

Consistent with the results herein, it was observed that the TL residue β'Arg933 counterpart forms hydrogen bonds with both β- and γ-phosphates in crystal structure of elongation complex with a substrate analogue AMPcPP (Vassylyev, et al., 2007b). The absence of γ-phosphates as in ADP might weaken but does not completely remove the interactions between the residue and the substrate. However, the absence of both phosphates as in AMP eliminates the interactions which may play an important role in stabilizing the folded conformation of TL. Also consistent with the results herein but inconsistent with the *T. thermophilus* structure, the *S. cerevisiae* elongation complex with substrate GTP shown that the TL residue β'His936 counterpart forms a hydrogen bond with only β-phosphate but not γ-phosphate (no direct interactions between the β'Arg933 counterpart and substrate phosphates were observed). In this case, it is not difficult to understand why the absence γ-phosphate as in ADP does not prevent TL closing but the absence of β-phosphate does.

4.5.3. Effects of ribose/deoxyribose identity

The effects of substrate ribose/deoxyribose identity on TL closing were also assessed by measurements of fluorescence intensity analogous to the previous two sections but using 2'-deoxyadenosine triphosphate (2’-dATP, or dATP) or 3'-deoxyadenosine triphosphate (3’-dATP). The results showed that 2’-dATP has a significantly smaller (62.8% versus 100%) effect on TL closing than that of ATP, while 3’-dATP has an effect comparable (83.7% versue 100%) with that of ATP (Figure 15). The results of fluorescence decays shown that both 2’-dATP and 3’-dATP cause an increase in Cou lifetime but 2’-dATP has a slightly smaller effect than ATP (Figure 15). Therefore, TL closing partially depends on 2’-ribose/deoxyribose identity, but does not depend on 3’-ribose/deoxyribose identity.

In consistence with the results herein, single-nucleotide addition assay using wild type and ΔTL *E. coli* RNAP demonstrated that TL makes a 100-fold discrimination in favor of NTPs over 2’NTPs (Zhang, et al., 2010), compared to an overall 1,000-fold discrimination by RNAP as a whole (Svetlov, et al., 2004). It was proposed that TL closing is not the sole factor for selection between NTP and 2’dNTP (Zhang, et al., 2010). There might be other components of the RNAP active center, such as β’Asn458 and β’Arg425 in *E. coli* RNAP, playing an essential role in discrimination of substrate ribose (Svetlov, et al., 2004; Zhang, et al., 2010). The proposal was strongly supported by the observation that, in the X-ray structure of *T. thermophilus* AMPcPP-bound elongation complex, the counterpart residues makes key contacts with the substrate 2’-OH (Vassylyev, et al., 2007b).

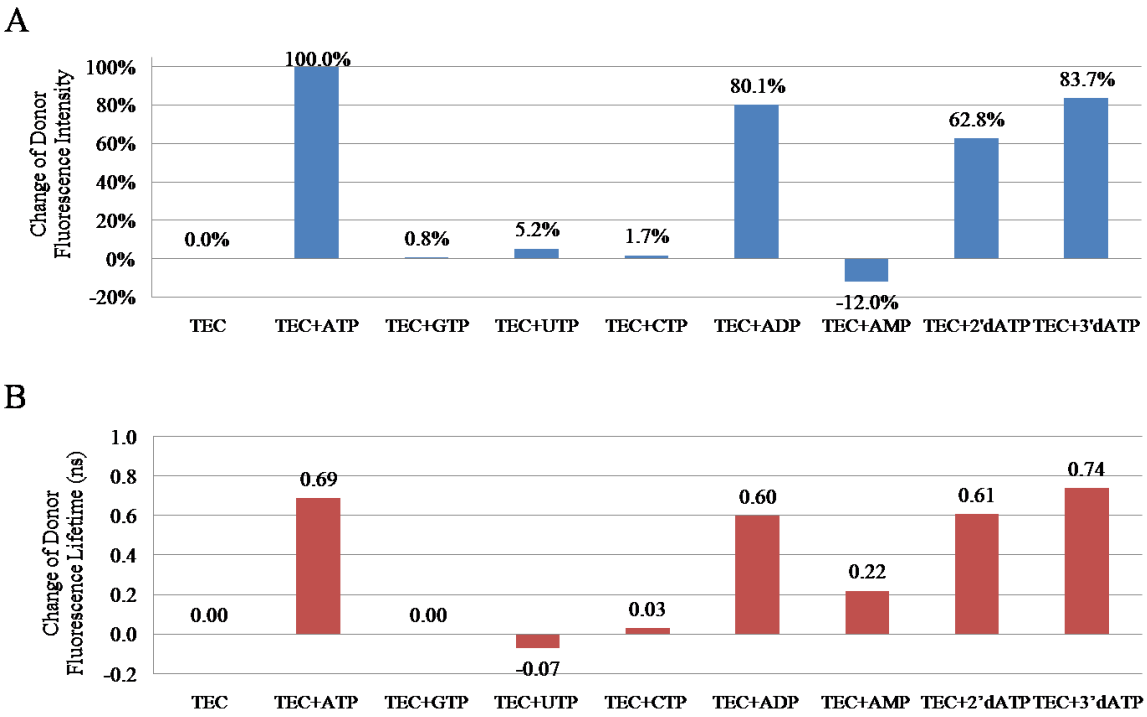


Figure 15. Effects of substrate identities on TL conformation using TEC with acceptor probe in DNA template position +20.
(A) Changes of Cou fluorescence intensity caused by NTPs or NTP analogues.
(B) Changes of Cou fluorescence lifetime caused by NTPs or NTP analogues.

4.6. Determination of TL positions in open and closed states in solution

Positions of TL residue β'942 in open and closed states were determined by distance-restrained rigid body docking using the FPS program. Dye distribution was modeled by the accessible-volume (AV) approach (Muschielok, et al., 2008), and dye position was approximated by a fixed mean position calculated from the AV. AV clouds and mean positions of the acceptor at position +20 (Figure 16A) and at all positions (-6, -4, -2, +8, +10, +12, +14, +15, +16, +18 and +20, Figure 16B) on DNA template strand were shown in Figure 16.

Measured distances with ATP-free TEC (proposed to adopt an open TL conformation) or ATP-bound TEC (proposed to adopt a closed TL conformation) in section 4.4. were used as restrains to map the positions of β'942 in the open and closed TL, respectively. There were 11 restrains for open TL, and 11 for closed TL. The first docking body was a 4YLN-derived TEC structure with deletion of TL (ΔTL) and deletion of SI3 domain (ΔSI3), and the second body was a single Ala residue representing β'942 omitting the dye and linker.

In results of residue β'942 in the open TL, the 1,000 structures obtained by a coarse search step with 2-Å clash tolerance were ranked by their χ_r^2 and clustered into groups of solutions with similar χ_r^2 and low r.m.s. deviation. The results showed that 999 solutions (99.9% population, Table 7) belong to one predominant group, followed by one structure with a large χ_r^2 value. The top fifty structures were refined using clash tolerance of 0.5 Å. The refined solutions mapped the residue β'942 in positions adjacent to each other, between the rim helices (β'650-703, border of the secondary channel) and the SI3 domain

at the entrance of the secondary channel (Figure 17A, top). By superimposing the FRET-derived solutions with X-ray structure 5BYH using β for alignment, we found that the cluster of FRET-defined positions of β'942 is ~10 Å to its crystallographically defined position in an open TL conformation (Figure 17A, bottom).

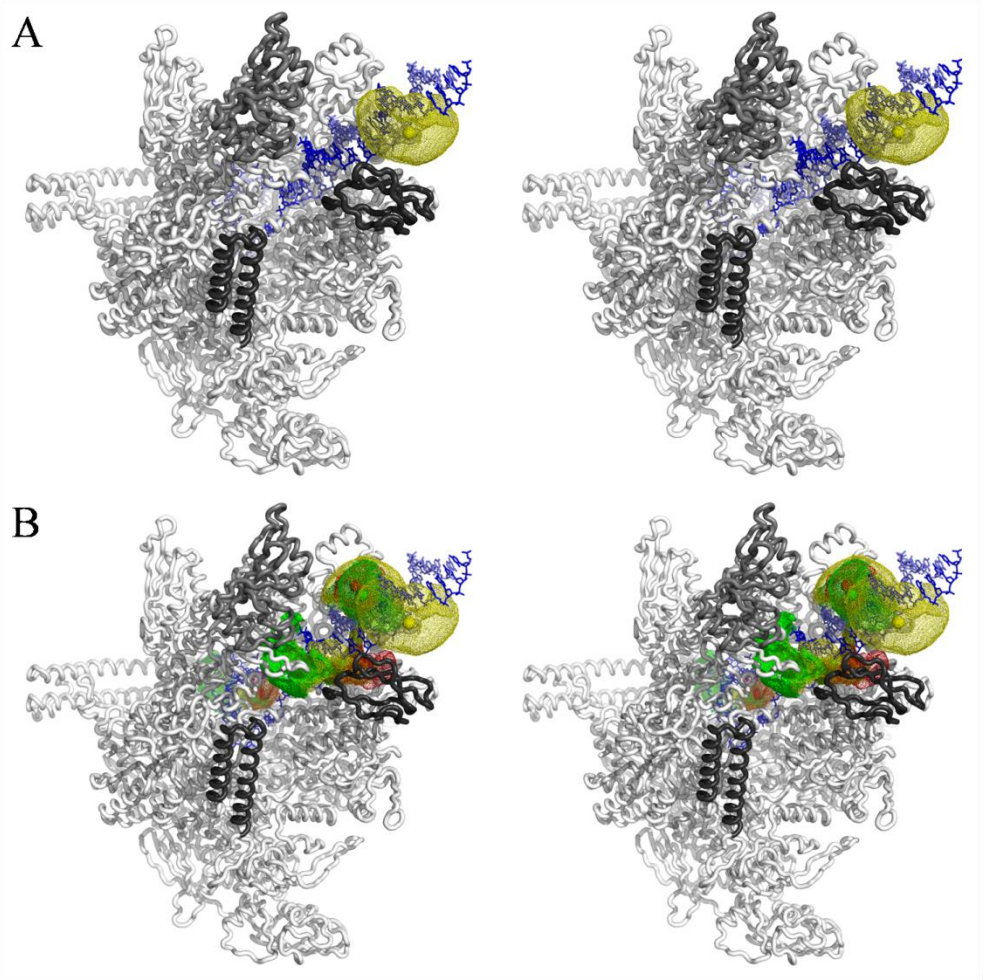


Figure 16. AV cloud and mean position of the acceptor.
(A) AV cloud (yellow mesh) and mean position (yellow sphere) of the acceptor at position +20.
(B) AV clouds (meshes in alternating green, yellow, and red) and mean positions (spheres in alternating green, yellow, and red) of the acceptor at all positions (-6, -4, -2, +8, +10, +12, +14, +15, +16, +18 and +20) on DNA template strand position.

In results of residue β'942 in the closed TL, clustering the 1,000 obtained solutions identified three distinct groups: a group accounting for 53.2% population, followed by two groups accounting for 22.5% and 24.0% population, followed by two ungrouped structures with large χ_r^2 (Table 7). These three groups had significantly different χ_r^2 and mapped the residue β'942 in distinct positions. The first group was the major group with lowest χ_r^2. It was the only one placing the residue β'942 in the secondary channel between the rim helices and the SI3 domain, with the second group placing it close to but outside the secondary channel (between the rim helices and β subunit), and the third group placing it distal to the secondary channel. Therefore, the major group is more likely to represent the real position of the residue in solution. Fifty refined solutions from this group mapped the residue in positions close to each other, forming a compact cluster (Figure 17B, top). Superimposition of the solutions with the crystal structure 4YLN revealed that the cluster of FRET-defined positions of β'942 was ~18 Å to its crystallographically defined position in an open TL conformation (Figure 17B, bottom). Compared with the modeled residue position in open TL, the one in closed TL was closer to the center of the channel entrance, which is more effective in blocking the path to the active center (Figure 17C, top). Superimposition of these two clusters shown that they had a distance of ~20 Å from each other, which is close to the crystallographically defined distance of ~21.3 Å (Figure 17C, bottom).

Interestingly, by superimposing the SI3-absent docking results with SI3 domain, we found that: (i) FRET-defined position of β'942 in open TL is compatible with SI3(B) (Figure 18A); (ii) FRET-defined position of β'942 in closed TL is compatible with SI3(A)

(Figure 18B); (iii) but FRET-defined position of β'942 in open TL is incompatible with SI3(A) (Figure 18C).

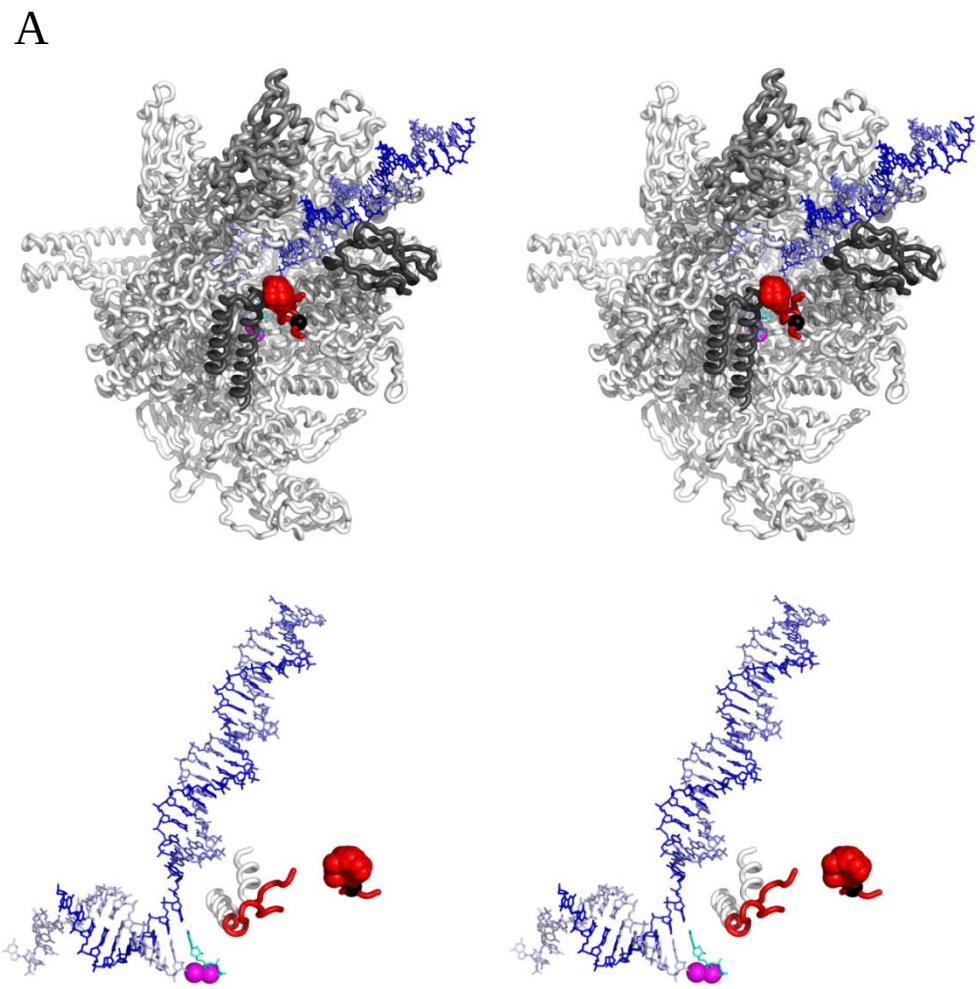


Figure 17. Positions of residue β'942 in open and closed TL by SI3-absent docking. FRET-defined and crystallographically-defined positions of β'942 in open TL are indicated as red and black spheres, respectively. FRET-defined and crystallographically-defined positions of β'942 in closed TL are indicated as green and gray spheres, respectively. Open TL is in red (ribbon); closed TL is in green (ribbon); DNA template strand is in blue (lines); DNA non-template strand is in slate (lines); RNA is in light blue (lines).
(A) Positions of β'942 in open TL (stereo view). Top: β'942 positions with respect to the TEC. Bottom: β'942 positions in the active center.

B

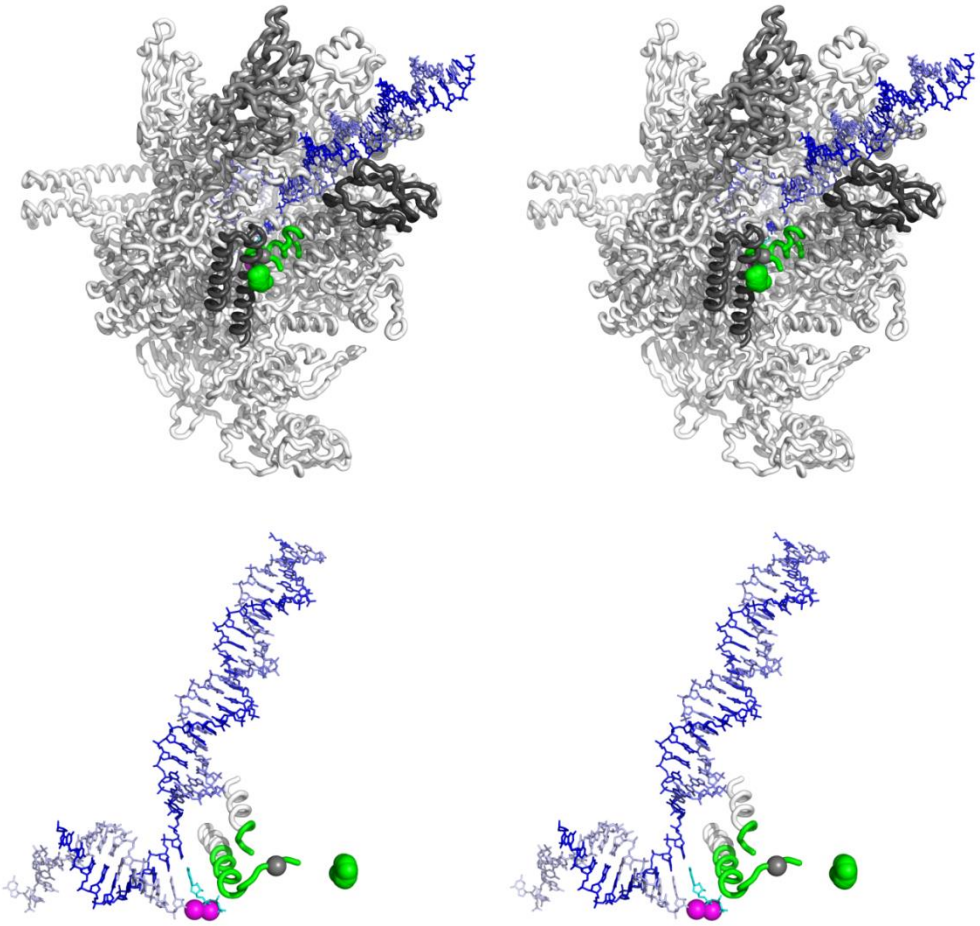


Figure 17. Positions of residue β'942 in open and closed TL by SI3-absent docking. (continued) FRET-defined and crystallographically-defined positions of β'942 in open TL are indicated as red and black spheres, respectively. FRET-defined and crystallographically-defined positions of β'942 in closed TL are indicated as green and gray spheres, respectively. Open TL is in red (ribbon); closed TL is in green (ribbon); DNA template strand is in blue (lines); DNA non-template strand is in slate (lines); RNA is in light blue (lines).
(B) Positions of β'942 in closed TL (stereo view).). Top: β'942 positions with respect to the TEC. Bottom: β'942 positions in the active center.

C

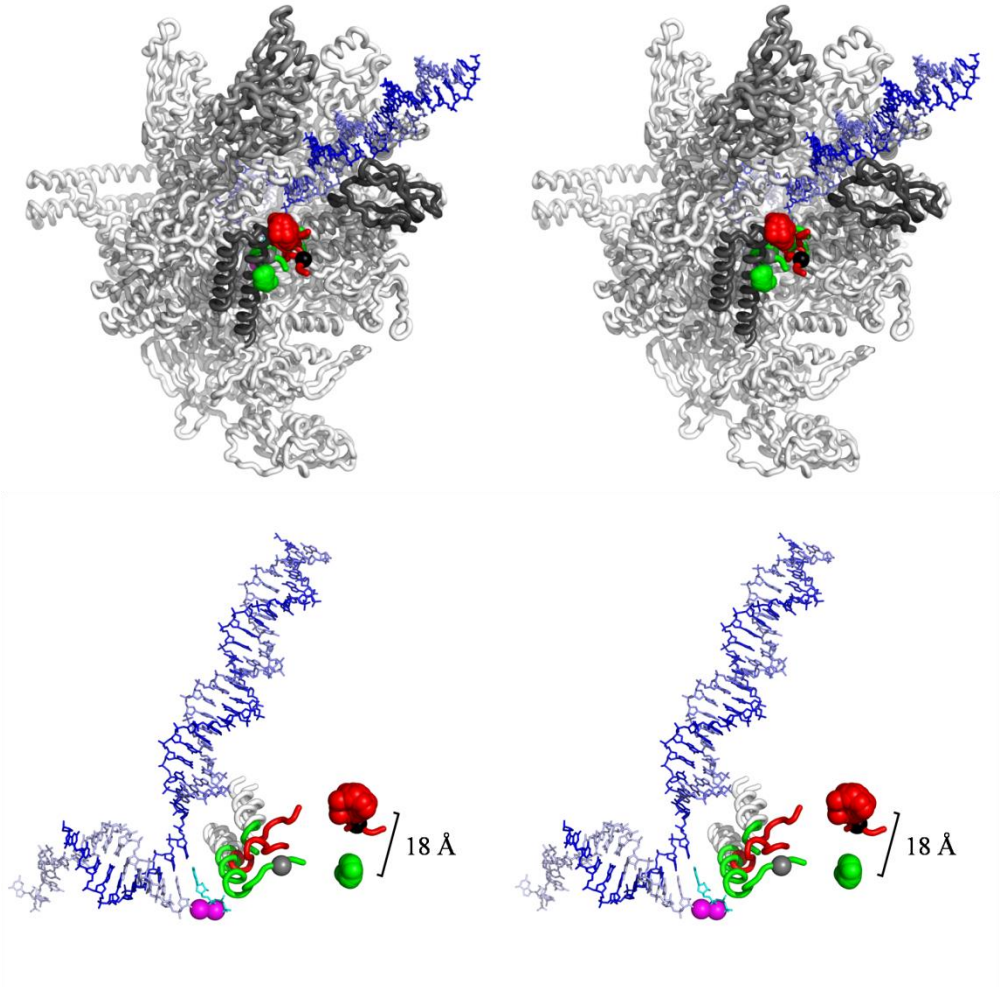


Figure 17. Positions of residue β'942 in open and closed TL by SI3-absent docking. (continued) FRET-defined and crystallographically-defined positions of β'942 in open TL are indicated as red and black spheres, respectively. FRET-defined and crystallographically-defined positions of β'942 in closed TL are indicated as green and gray spheres, respectively. Open TL is in red (ribbon); closed TL is in green (ribbon); DNA template strand is in blue (lines); DNA non-template strand is in slate (lines); RNA is in light blue (lines).
(C) Superimposed positions of β'942 in open and closed TL (stereo view).). Top: β'942 positions with respect to the TEC. Bottom: β'942 positions in the active center.

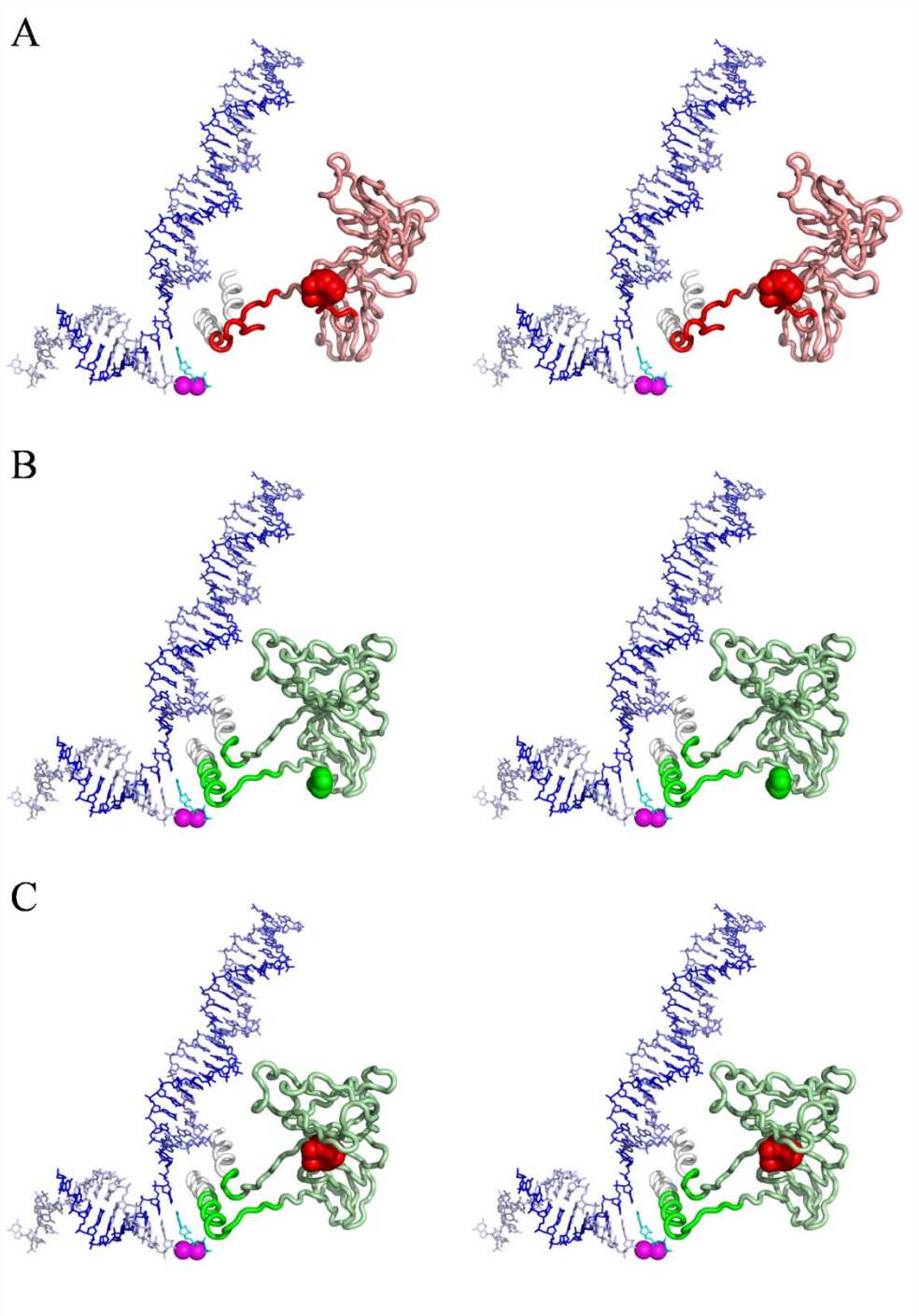


Figure 18. Compatibility of SI3 states and β'942 positions in open and closed TL by SI3-absent docking.
(A) FRET-defined position of β'942 (red sphere) in open TL is compatible with SI3(B) (pink domain).
(B) FRET-defined position of β'942 (green sphere) in closed TL is compatible with SI3(A) (light green domain).
(C) FRET-defined position of β'942 (red sphere) in open TL is incompatible with SI3(A) (light green domain).

4.7. Analysis of effect of TL conformation on SI3 orientation

To further analyze the effect of TL conformation on SI3 orientation, SI3-present docking simulations were performed in a procedure analogous to the SI3-absent docking but using SI3-containing TEC structures as the docking body. A ΔTL 5BYH-derived TEC structure with SI3(B) (compatible with open TL) was used to map the position of β'942 in the open TL, and a ΔTL 4YLN-derived TEC structure with SI3(A) (compatible with closed TL) was used to map the position of β'942 in the closed TL.

The results of SI3-present docking were very similar with those of SI3-absent docking in terms of number of identified solution groups, population of each group, and modeled positions by the best solutions (Table 7 and Figure 19). In results of open TL, 99.3% of the 1,000 obtained solutions belong to a predominant group. Fifty refined solution from this group placed the TL residue β'942 between the rim helices and SI3 domain (Figure 19A, top). The cluster of solutions overlapped with the cluster by SI3-absent docking (Figure 19A, top and bottom). In results of closed TL, clustering of the solutions obtained for closed TL defined three distinct groups, accounting for 47.6%, 24.2% and 27.6% of population, followed by six ungrouped solutions with large χ_r^2. These three groups placed the residues in positions similar with the corresponding positions by SI3-absent docking. Top fifty refined solutions from the first group placed the residue close to the RH (Figure 19B, top). The cluster of solutions highly agreed with the cluster by SI3-absent docking. The modeled positions in the open and closed TL were ~20 Å apart, consistent with the results of SI3-absent docking and crystal structures.

Consistent with the observation in SI3-absent docking, results of the SI3-present docking showed that: (i) FRET-defined position of β'942 in open TL is compatible with SI3(B) (Figure 20A); (ii) FRET-defined position of β'942 in closed TL is compatible with SI3(A) (Figure 20B); (iii) but FRET-defined position of β'942 in open TL is incompatible with SI3(A) (Figure 20C).

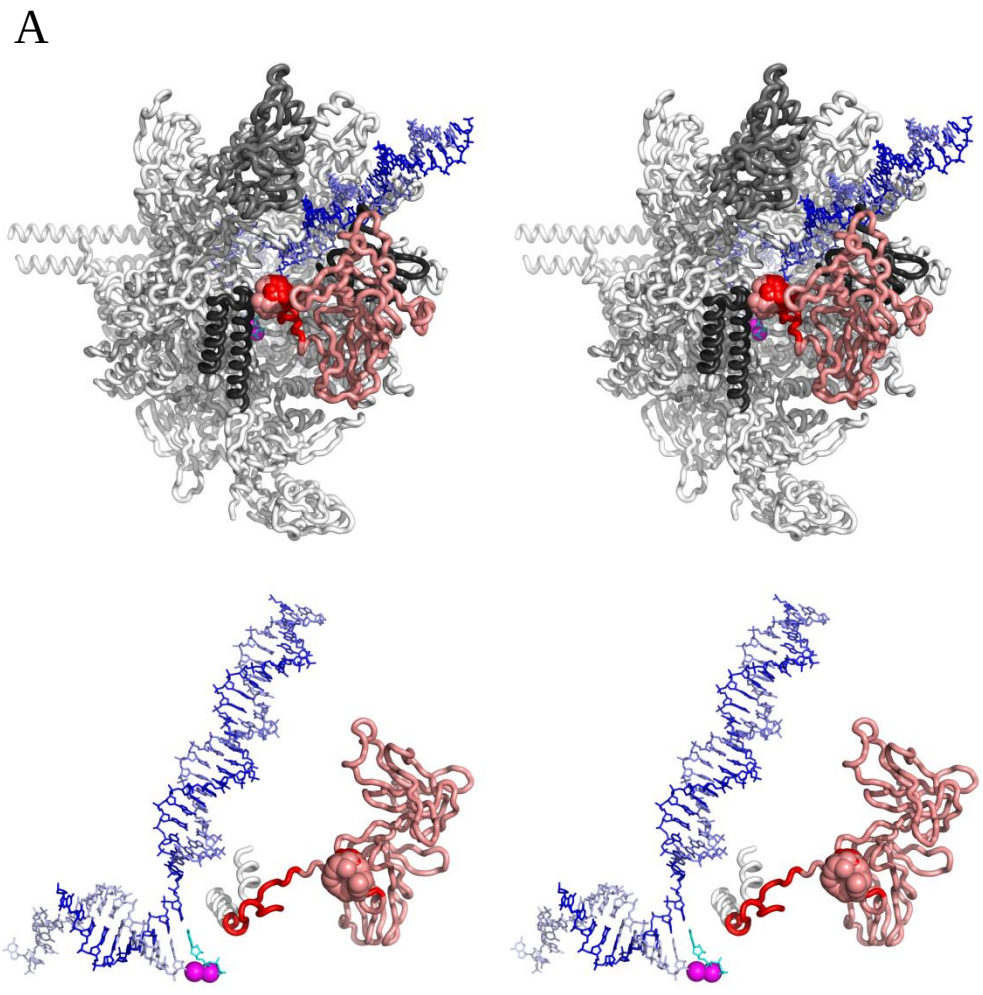


Figure 19. Comparisons of β'942 positions by SI3-present and SI3-absent docking. (A) Positions of β'942 in open TL by SI3-present (pink) and SI3-absent (red) docking. Top: overview. Bottom: active center.

B

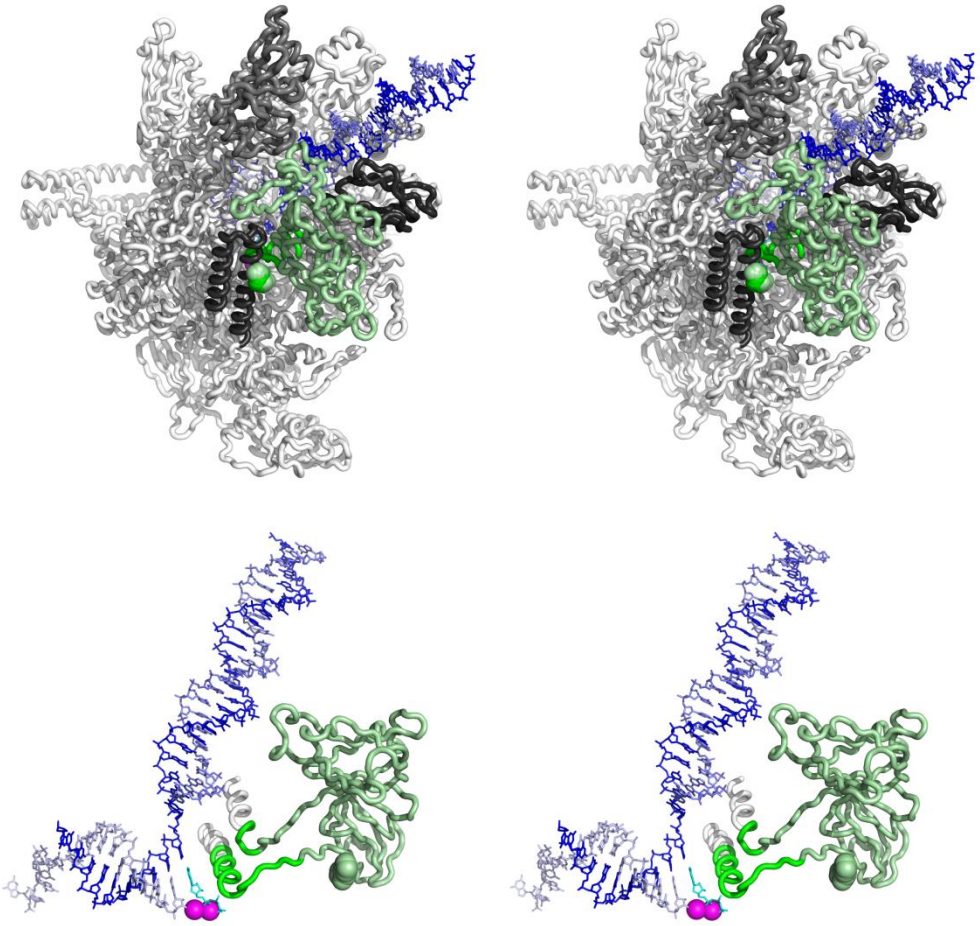


Figure 19. Comparisons of β'942 positions by SI3-present and SI3-absent docking. (continued)
(B) Positions of β'942 in closed TL by SI3-present (light green) and SI3-absent (green) docking. Top: overview. Bottom: active center.

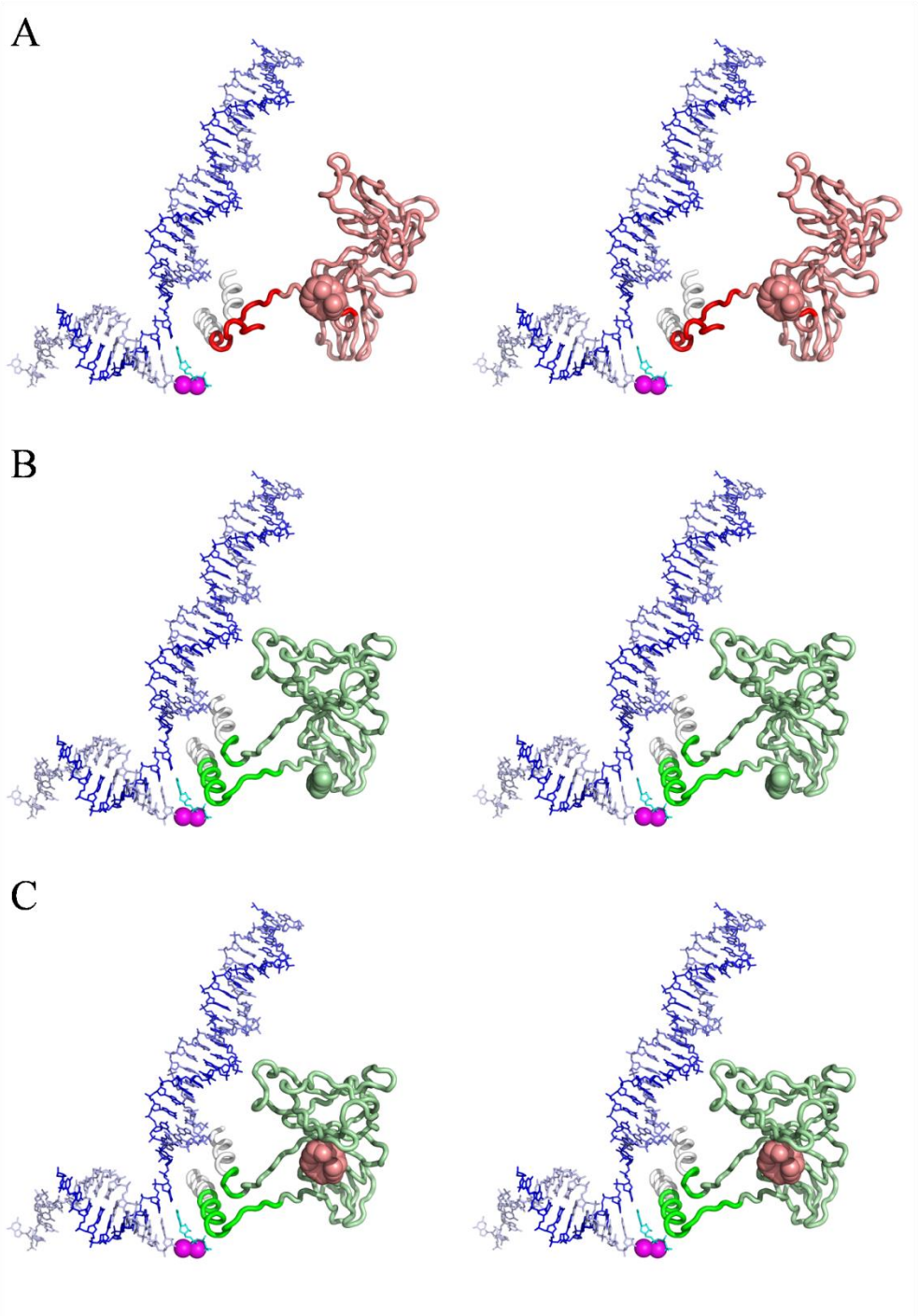


Figure 20. Compatibility of SI3 states and β'942 positions in open and closed TL by SI3-present docking.
(A) FRET-defined position of β'942 (pink sphere) in open TL is compatible with SI3(B) (pink domain).
(B) FRET-defined position of β'942 (light green sphere) in closed TL is compatible with SI3(A) (light green domain).
(C) FRET-defined position of β'942 (pink sphere) in open TL is incompatible with SI3(A) (light green domain).

Table 7. Summary of model solutions in SI3-absent and SI3-present docking

Docking	TL conformation	Group of solutions	Population (%)	Lowest χ_r^2
SI3-absent	open	1	99.9	1.58
	closed	1	53.2	1.50
		2	22.5	1.57
		3	24.0	2.70
SI3-present	open	1	99.3	1.65
	closed	1	47.6	1.52
		2	24.2	1.59
		3	27.6	2.75

5. Discussion

5.1. Conclusions and implications

5.1.1. Occurrence of TL closing/opening in solution

In this work, it has been directly determined that TL closing and opening occurs in solution by monitoring FRET efficiencies between a fluorescent probe Cou incorporated in TL using the method of unnatural amino acid mutagenesis, and a complementary probe fluorescein chemically attached to DNA in NTP-free TECs or NTP-bound TECs. Measured significant changes in FRET efficiencies correspond to significant changes in distances between TL and stationary reference sites in TEC, providing a direct evidence of TL closing unaffected by crystal packing interactions or the stabilizing influence of crystallization co-solvents. The results indicated that TL closing is a result of NTP binding, which agrees with the crystal-structure-based proposal that each nucleotide addition cycle is coupled to the active-center conformational cycling involving TL closing upon NTP binding followed by TL re-opening upon PPi release.

5.1.2. Central role of TL in substrate selection

In addition, it has been demonstrated that substrate identities, including base complementarity, ribose/deoxyribose identity, and phosphate intactness, all affect TL closing to different extents. Unmatched NTP do not induce TL closing. The absence of γ-phosphate as in ADP merely influences TL in a modest degree, while the absence of both β- and γ-phosphates as in AMP almost completely impairs TL closing. The 2'-

deoxyribose identity significantly impairs TL closing, while the 3'-deoxyribose identity only has a mild effect. Consistent with the results, recent genetic, biochemical and structure studies supported a view that TL plays a central role in NTP selection thereby contributing to the high fidelity of transcription (Nudler, 2009). On the other hand, the small effect of 2'-deoxyribose identity on TL closing supported the proposal that TL is probably not the sole checkpoint for selecting NTP over dNTP (Zhang, et al., 2010). Indeed, RNAP residues β'Asn458 and β'Arg425 have been shown important in discrimination against dNTP (Svetlov, et al., 2004). The low sensitivity of TL against the 3'-deoxyribose identity suggested that TL probably participates in discrimination against 3'-dNTP merely through indirect interactions rather than direct interactions. Consistent with this view, direct contacts between TL and the NTP 3'-OH were not observed in crystal structures, however, the β'Gln929 counterpart in RNAP II interacted with the β'Asn458 counterpart, which, in turn, formed a hydrogen bond with the NTP 3'-OH group (Wang, et al., 2006).

5.1.3. Positions of β'942 in open and closed TL in solution

Moreover, the positions of residue β'942 in the open and closed TL have been mapped by distance-restrained docking using the systematic FRET measurement data as distance restrains. The best-fit FRET-derived models placed the open and closed TL in the secondary channel with the two positions apart by a significant distance. Compared with the open TL, the closed TL was placed in a position at which the segment can more effectively block the secondary channel thereby hindering substrate dissociation and

competition. From this point of view, the results were consistent with the observations in biochemical and structural studies. However, the FRET-defined positions were significantly different from the crystallographically defined positions. This may be explained by the difference in how crystallography and ensemble FRET determine loop conformations, and by the relatively low accuracy of FRET-derived positioning method. First, NMR-based structural biology studies indicated that solvent-exposed loops are "floppy" and adopt many conformations. The cystallographically defined loop conformations may have been affected by crystal packing interactions, or the stabilizing influence of crystallization co-solvents, which means they might be neither the lowest-energy conformation nor the real conformation in solution. The open TL exhibiting a disordered (as seen in many RNAP structures, e. g. PDB 2O5I) or an extended (as seen in a few RNAP structures, e. g. PDB 5BYH, Figure 16C) conformation is likely to be an example. It can be also true for the closed TL because the segment centered on β'942 (β'939-945) adopted an extended conformation in *E. coli* RNAP complexes (e. g. PDB 4YLN, PBD 5IPM) due to the existence of SI3 (Figure 16C). On the other hand, ensemble FRET measurements provide averaged FRET efficiency within the molecule population, by which averaged probe-probe distances are calculated. As a result, the FRET-derived models merely represent a mixed conformation. A few conformations far away in position from to the major population can largely alter the mean conformation. Second, the accuracy of FRET-derived distance-restrained rigid body docking mainly depends on the uncertainties of distance determination, the number of distance restrains, and the geometry of the problem (Beckers, et al., 2015). Typical values of error are on the order of 10-20 Å (Nagy, et al., 2015). The uncertainties of distance determination usually

arise due to the intrinsic uncertainty of the relative orientation of dye molecules and the experimental uncertainty of quantum yield determination. The accuracy of determining an unknown position depends on the number of restrains used and the distribution of the reference sites. According to a benchmark study of distance-restrained docking using systematic FRET measurements, at least 20 restrains are needed to obtain accuracy of 10 Å or better (Knight, et al., 2005). The reference sites should form a triangulational, or more correctly, trilaterational distribution. In this work, only eleven reference sites were used, and unfortunately, they were all on the DNA due to extreme difficulties in placing a second fluorescent probe on RNAP without non-specific labelling and loss of transcriptional activity.

5.1.4. Effect of TL conformation on SI3 orientation

The modeled position of β'942 in open TL was found incompatible with the SI3(A) that accommodates the closed TL conformation. This observation implied that the SI3 module probably varies in at least two orientational states in response to the conformational change of TL. It was consistent with the conclusions based on EM structure and hybrid-approach structure of *E. coli* RNAP (Hudson, et al., 2009; Opalka, et al., 2010). SI3 orientation cycling may have the following potential roles. Movement of the large module (~ 120 nm^3) coupled to conformational cycling of TL in each nucleotide addition step is likely to hinder the inter-conversion between the open and closed TL states thereby altering the catalytic rate. In addition, the domain is near the RNAP jaw and β DR1, potentially making direct contacts with these domains and serving for

communication between the TL in interior and the protein exterior. Also, SBHMb in the domain is ~ 15-20 Å from the DNA duplex downstream of transcription bubble, allowing the module to act as a target of transcription factors that bind on the downstream DNA and regulate the catalysis by trapping TL in a certain conformational state.

5.2. Significance in methods development

Roadblocks encountered in this work have fully demonstrated the challenges of labelling TL and preparing labelled TEC for FRET measurements. For instance, although the method of unnatural amino acid mutagenesis can specifically introduce the probe Cou to the target position in TL, the yield of labelled full-length β' was unacceptably low due to the rather low amber-suppression efficiency by the orthogonal tRNA/synthetase pair. The labelled β' in a low amount was difficult to isolate and difficult to meet the high-amount requirement for *in vitro* reconstitution of RNAP. Moreover, the presence of Cou in TL largely impaired the catalytic activity of labelled RNAPs. In addition, the RNAPs prepared by *in vitro* reconstitution exhibited relatively low monodispersity, which depressed their transcriptional activity and ability of forming TEC *in vitro*. Besides, the presence of excess labelled DNA and RNAP interfered fluorescence measurements of the formed TECs, thus an effective method was needed to purify labelled TEC. Four procedures were developed and successfully solved the specific problems.

First, a labelling procedure using split β' was developed and substantially improved the yield of β' having Cou incorporated in it. Better incorporation at amber codon closer to translation start may be resulted from less time and material spent in synthesizing truncated protein of a smaller size. This procedure can be theoretically applied to introduce other unnatural amino acids to β' or β (by using split β fragments, Artsimovitch, et al., 2003), particularly for those unnatural amino acids whose incorporation efficiencies have been shown poor in full-length β' or β.

Second, a labelling procedure involving Ala substitution of amino acid residues following Cou was developed and successfully improved the transcriptional activity of Cou-labelled RNAPs. The dramatic increases in activity were likely resulted from the removal of interactions (or steric hindrance) by the side chain of amino acid residue(s) immediately next to Cou thereby providing more flexibility in transition between TL conformational states. This procedure serves a reference for introducing an unnatural amino acid in protein interior with minimized loss of protein functions.

Third, a one-pot *in vivo* preparation procedure of RNAP incorporated with Cou was developed and dramatically improved the monodispersity of the labelled RNAPs thereby improving their transcriptional activity and TEC-formation ability. These improvements might be attributed to correct protein folding in physiological environment during *in vivo* assembly of RNAP, rather than artificially renaturating conditions during *in vitro* reconstitution. This procedure was faster than *in vitro* reconstitution due to the use of subunit co-overexpression and *in vivo* assembly. In addition, the yield of the labelled RNAP was higher than that by reconstitution. This procedure can be applied for

incorporation of other amino acids into RNAP with minimized loss of its transcriptional activity.

Fourth, an electroelution-based purification procedure for RNAP complexes was developed and eliminated most fluorescence noise from excess components. This procedure can be readily applied for purification of any other fluorescently labelled RNAP complexes, as well as fluorescently unlabelled RNAP complexes simply by using a labelled RNAP complex as fluorescent marker.

References

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

Artsimovitch, I., Svetlov, V., Murakami, K. & Landick, R. (2003) Co-overexpression of Escherichia coli RNA Polymerase Subunits Allows Isolation and Analysis of Mutant Enzymes Lacking Lineage-specific Sequence Insertions. *J. Biol. Chem.* **278**, 12344-12355.

Bae, B., Davis, E., Brown, D., Campbell, E., Wigneshweraraj, S. & Darst, S. (2013) Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ^{70} domain 1.1. *Proc. Natl. Acad. Sci. USA* **110**, 19772-19777.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A., Rafickov, R., Mustaev, A. & Nudler, E. (2005) A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183-193.

Berkers, M., Drechsler, F., Eilert, T., Nagy, J. & Michaelis, J. (2015) Quantitative structural information from single-molecule FRET. *Faraday Discuss*., **184**, 117-129.

Cai, Q. et al. (2007) Nanometer distance measurements in RNA using site-directed spin Labelling. *Biophys. J.* **93**, 2110-2117.

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

Chakraborty, A., Mazumder, A., Lin, M., Hasemeyer, A., Xu, Q., Wang, D., Ebright, Y. W. & Ebright, R. H. (2015) Site-specific Incorporation of Probes into RNA Polymerase by Unnatural-amino-acid Mutagenesis and Staudinger-Bertozzi Ligation. Bacterial Transcriptional Control. *Methods in Molecular Biology*, **1276**, 101-131.

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., and Bowman, R. (1965) Fluorescence polarization: mea- surement with ultraviolet-polarizing filters in a spectrophotofluorometer. *Science* **147**, 729–732.

Chlenov, M., Masuda, S., Murakami, K., Nikiforov, V., Darst, S. & Mustaev, A. (2005) Structure and Function of Lineage-specific Sequence Insertions in the Bacterial RNA Polymerase β' Subunit. *J. Mol. Biol*. **353**, 138–154.

Choi, U. B., Strop, P., Vrljic, M., Chu, S., Brunger, A. T. and Weninger, K. R. (2010) Single-molecule FRET–derived model of the synaptotagmin 1–SNARE fusion complex. *Nat. Struct. Mol. Biol.*, **17**, 318–324.

Clegg, R. (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.

Da, L., Wang, D. & Huang, X. (2012) Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II. *J. Am. Chem. Soc.*, **134**, 2399–2406.

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

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

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

Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., NIkiforov, V. & Goldfarb, A. (2002) Swing-Gate Model of Nucleotide Entry into the RNA Polymerase Active Center. *Mol. Cell* **10**, 623–634.

Erie, D., Yager, T. & von Hippel, P. (1992) The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu. Rev. Biophys. Biomol. Struct*. **21**, 379-415.

Eastman, J. W. (1967) Quantitative spectrofluorimetr--the fluorescence quantum yield of quinine sulfate. *Photochem Photobiol*. **6**, 55–72.

Feng, Y., Degen, D., Wang, X., Gigliotti, M., Liu, S. Zhang, Y., Das, D., Michalchuk, T., Ebright, Y., Talaue, M., Connell, N. & Ebright, R. (2015) Structural Basis of Transcription Inhibition by CBR Hydroxamidines and CBR Pyrazoles. *Structure* **23**, 1470-1481.

Goerke, A. & Swartz, J. (2009) High-Level Cell-Free Synthesis Yields of Proteins Containing Site-Specific Non-Natural Amino Acids. *Biotechnol. Bioeng*. **102**, 400-416.

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

Haugen, S., Ross, W., and Gourse, R. (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nature Reviews Microbiology* **6**, 507-519.

Hudson, B.P., Quispe, J., Lara-González, S., Kim, Y., Berman, H.M., Arnold, E., Ebright, R.H., and Lawson, C.L. (2009). Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. *Proc. Natl. Acad. Sci. USA* **106**, 19830-19835.

Huiwitz, J. (2005) The Discovery of RNA Polymerase. *J. Biol. Chem.* **280**, 42477-42485.

Imashimizu, M., Takahashi, H., Oshima, T., Mclntosh, C., Bubunenko, M., Court, D.& Kashlev, M. (2015) Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biology*. 16:98.

Iyer, L., Koonin, E. & Aravind, L. (2004) Evolution of bacterial RNA polymerase: implications for large-scalebacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* **335**, 73-88.

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

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

Kaplan, C.D., Larsson, K.M. & Kornberg, R.D. (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell* **30**, 547–556.

Kashkina, E., Anikin, M., Tahirov, T., Kochetkov, S., Vassylyev, D. & Temiakov, D. (2006) Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Research* **34**, 4036–4045.

Kettenberger, H., Armache, K. J. & Cramer, P. (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, **16**, 955-965.

Knight, J., Mekler, V., Mukhopadhyay, J., Ebright, R. & Levy, R. (2005) Distance-Restrained Docking of Rifampicin and Rifamycin SV to RNA Polymerase Using

Systematic FRET Measurements: Developing Benchmarks of Model Quality and Reliability. *Biophys. J.* **88**, 925–938.

Kuhlman, P., Duff, H. & Galant, A. (2004) A fluorescence-based assay for multisubunit DNA-dependent RNA polymerases. *Analytical Biochemistry* **324**, 183–190.

Kulbachinskiy, A. & Mustaev, A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem*. **281**,18273-18276.

Lakowicz, J. Principles of Fluorescence Spectroscopy. New York: Springer US. 2006. 3rd edition. Print.

Liu, B., Zuo, Y. & Steitz, T. (2016) Structures of *E. coli* σS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc. Natl. Acad. Sci. USA* **113**, 4051–4056.

Liu, C. (2007). The use of single-molecule DNA nanomanipulation to study transcription kinetics. Ph.D. Dissertation. (Rutgers University, NJ: UMI Dissertations Publishing).

Liu, S., Lv, P., Li, D., Guo, X., Zhang, B., Yu, M., Li, D., Xiong, Y., Zhang, L. & Tian, C. (2015) K^+ preference at the NaK channel entrance revealed by fluorescence lifetime and anisotropy analysis of site-specifically incorporated (7-hydroxycoumarin-4-yl)ethylglycine. *Chem. Commun*. **51**, 15971-15974.

Malhotra, A., Severinova, E. & Darst, S. (1996) Crystal structure of a σ70 subunit fragment from *E. coli* RNA polymerase. *Cell* **87**, 127-136.

Malinen AM, et al. (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res* **40**, 7442–7451.

Mejia, Y., Nudler, E. & Bustamante, C. (2015) Trigger loop folding determines transcription rate of Escherichia coli's RNA polymerase. *Proc. Natl. Acad. Sci. USA* **112**, 743-748.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A., Niu, W., Ebright, Y., Levy, R. & Ebright, R. (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 ω and eukaryotic RNA

polymerase subunit RPB6 are sequence, structural, and functional homologs and promoter RNA polymerase assembly. *Proc. Natl. Acad. Sci. USA* ***98***, 892-897.

Mukhopadhyay, J., Kapanidis, A., Mekler, V., Kortkhonjia, E., Ebright, Y. & Ebright, R. (2001) Translocation of σ70 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., Ebright, Y. & Ebright, R. (2003) Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. *Meths. Enzymol*. **371**, 144-159.

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

Nudler, E. (2009) RNA Polymerase Active Center: The Molecular Engine of Transcription. *Annu. Rev. Biochem*. **78**, 335–361.

J. Nagy, D. Grohmann, A. C. M. Cheung, S. Schulz, K. Smollett, F. Werner and J. Michaelis (2015) Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun*. **6**, 6161.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V. & Ebright, R. (2000) Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**, 601-611.

Nikolai, N., Kim, Y., Dong, Q. & Ebright, R. H. (2001) Site-Specific Protein–DNA Photocrosslinking: Analysis of Bacterial Transcription Initiation Complexes. Methods in *Molecular Biology*. **148**, 337-361.

Opalka N, Brown J, Lane WJ, Twist K-AF, Landick R, et al. (2010) Complete Structural Model of Escherichia coli RNA Polymerase from a Hybrid Approach. *PLoS Biol* **8(9)**: e1000483. doi:10.1371/journal.pbio.1000483.

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

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

Saecker, R., Record, M. T. Jr. & deHaseth, P. (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.

Seibold, S., Singh, B., Zhang, C., Kireeva, M., Domecq, C., Bouchard, A., Nazione, A., Feig, M., Cukier, R. Coulombe, B., Kashlev, M., Hampsey, M. & Burton, Z. (2010) Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *Biochim. Biophys. Acta*. **1799**, 575–587.

Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S. & Goldfarb, A. (1996) Structural Modules of the Large Subunits of RNA Polymerase. Introducing archaebacterial and chloroplast split sites in the β and β' subunits of Escherichia coli RNA polymerase. *J. Biol. Chem*. **271**, 27969-27974.

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

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V. & Goldfarb, A. (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J*. **22**, 2234–2244.

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

Svetlov, V., Vassylyev, D.G. & Artsimovitch, I. (2004) Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J. Biol. Chem*. **279**, 38087–38090.

Toulokhonov I, Zhang J, Palangat M, Landick R. 2007. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Mol. *Cell* **27**, 406–419.

Tuske, S., Sarafianos, S., Wang, X., Hudson, B., Sineva, E., Mukhopadhyay, J., Birktoft, J, Leroy, O., Ismail, S., Clark, A., Dharia, C., Napoli, A., Laptenko, O., Lee, J., Borukhov, S., Ebright, R., and Arnold, E., (2005) Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122**, 541-552.

Vassylyev, D., Vassylyeva, M., Perederina, A., Tahirov, T. & Artsimovitch, I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157-162. (Vassylyev, et al., 2007a)

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

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D. & Kornberg, R.D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954.

Wang, J., Xie, J. & Schultz, P. (2006) A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* **128**, 8738-8739.

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

Werner, F. (2013) Molecular Mechanisms of Transcription Elongation in Archaea. *Chem. Rev*. **113**, 8331−8349.

Yang, Y., Darbari, V., Zhang, N., Lu, D., Glyde, R., Wang, Y., Winkelman, J., Gourse, R., Murakami, K., Buck, M. & Zhang, X. (2015) Structures of the RNA polymerase- σ^{54} reveal new and conserved regulatory strategies. *Science*, **349**, 882-885.

Yi, L., Sun, H., Itzen, A., Triola, G., Waldmann, H., Goody, R. & Wu, Y. (2011) One-Pot Dual-Labelling of a Protein by Two Chemoselective Reactions. *Angew. Chem. Int. Ed*. **50**, 8287-8290.

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

Young T. S., Ahmad, I., Yin J.A. & 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., Minakhin, L., Richter, C., Severinov, K. & Darst, S. (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811-824.

Zhang, J., Palangat, M. & Landick, R. (2010) Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol*. **17**, 99-104.

Zuo, Y. & Steitz, T. (2015) Crystal Structures of the *E. coli* Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* **58**, 534-540.

Zuo, Y., Wang, Y. & Steitz, T. (2013) The Mechanism of E. coli RNA Polymerase Regulation by ppGpp Is Suggested by the Structure of their Complex. *Mol. Cell* **50**, 430-436.