SEQUENCE-SPECIFIC INTERACTIONS BETWEEN RNA POLYMERASE AND THE CORE RECOGNITION ELEMENT

By

HANIF VAHEDIAN-MOVAHED

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 Biochemistry

Written under the direction of

Richard H. Ebright and Bryce E. Nickels

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2017

ABSTRACT OF THE DISSERTATION

Sequence-specific interactions between RNA polymerase and the core recognition element

by HANIF VAHEDIAN-MOVAHED

Dissertation Directors:

Richard H. Ebright and Bryce E. Nickels

In bacteria, the flow of biological information from DNA to RNA is carried out by a single enzyme called RNA polymerase (RNAP). Bacterial RNAP is composed of a multi-subunit catalytic core and a dissociable subunit called sigma factor. Since the discovery of sigma factor in 1969, the prevailing view has been that the RNAP core enzyme requires binding to sigma for sequence-specific transcription, because the RNAP core does not contain the determinants for sequence-specific core promoter recognition and DNA unwinding. It has also been assumed that sequence-specific RNAP-DNA interactions are mainly limited to transcription initiation, as sigma could dissociate from the RNAP core after the initiation stage. These two paradigmatic assumptions have been challenged by recent structural evidence from our lab, which indicates in the initiation complex, the RNAP core directly interacts with the non-template-strand segment of the transcription bubble corresponding to positions -4 to +2, and that the interaction with this element is sequence-specific at least at one of its positions. This element has been termed the “core recognition element,” CRE. This thesis addresses three major topics regarding CRE: sequence-specificity, recognition mechanism, and the functional roles.

In chapter 1, using equilibrium binding and dissociation kinetics studies, I demonstrate that the RNAP core shows sequence-specificity at 3-out-of-6 CRE positions (the consensus sequence is T_{-4} n_{-3} n_{-2} n_{-1} T_{+1} G_{+2}). I also determine that RNAP amino acid βR371 mediates specificity at CRE position -4, βW183 mediates specificity at CRE position +1, and βR151, βD446, or βR451 mediates specificity at CRE position +2. In subsequent chapters, I use the RNAP derivative containing the βD446A substitution as a reagent to assess the functional significance of RNAP-CRE^{+2G} interactions on transcription initiation and elongation.

In chapters 2, 3 and 4, using a combination of next-generation sequencing approaches and biophysical and biochemical assays, I show that sequence-specific RNAP-G_{CRE} interactions play functional roles in three key stages of transcription initiation: promoting DNA unwinding at a consensus G_{CRE} sequence, favoring start-site selection at positions upstream of a consensus G_{CRE} sequence, and reducing the probability of abortive transcript release at positions upstream of a consensus G_{CRE} sequence.

In chapter 5, using biochemical assays and mNET-seq, I show that sequence-specific RNAP-G_{CRE} interaction occurs in and plays functional roles in key stages of transcription elongation through the E. coli genome: favoring pause-read-through at positions upstream of consensus G_{CRE} sequence and favoring post-translocated states at positions upstream of consensus G_{CRE} sequence.

In chapter 6, using a promoter-independent transcription assay, I show that RNAP-G_{CRE} interaction occurs in, and plays functional roles in all three domains of life: bacteria, archaea and eukaryote.

In chapter 7, using genome-wide next-generation sequencing approaches, I show that the RNAP core can perform sequence-specific transcription in the absence of sigma factor in a manner that correlates with the presence of an AT-rich region followed by a TG-motif.

In the final chapter, I summarize how my work revealed previously undocumented regulatory events in transcription initiation and elongation. Based on my findings, I describe two implications of this thesis for future consideration: a scenario describing what the architecture of primordial promoter sequences might have looked like and a mechanism for antibiotic-tolerant persistence state.

Acknowledgements

I cannot imagine being anywhere but the Ebright and Nickels labs for my PhD training. My graduate training with them created some of the most enjoyable moments of discoveries, enlightenments and enthusiasm while personally experiencing the objective truth of science. It was a pleasure to spend 5.68 years in the labs of Professors Ebright and Nickels, and I am greatly indebted to their teachings.

I am sincerely grateful to my PhD advisor, Professor Richard Ebright, for his continuous support, enlightening advice, incisive guidance and criticism, while giving me research freedom and providing a rich research environment, which have improved my abilities as a scientist, particularly in the judgment of scientific hypotheses and in experimental design.

I would like to extend my gratitude to my PhD co-advisor, Professor Bryce Nickels, for his effective advice, for his invaluable contributions of making the new next-generation sequencing-based technologies available to my research, and for being a patient mentor in teaching me many things, in particular my very first molecular genetic experiments, while I was a rotation student in his lab.

I would like to thank the members of my thesis committee: Profs. Wilma Olson, Eddy Arnold and Smita Patel, for their insightful feedback, discussions and encouragement.

I would also like to thank the members of the Ebright lab, in particular Yu Zhang for collaboration on chapter one of my thesis and his published crystal structures, which have been very insightful to my work; Yu Zhang and Yu Feng for teaching me the arts of protein purifications; Anirban Chakraborty for teaching me site-specific labeling of proteins and single-molecule FRET during my rotation, which, although not related to this thesis, were

nonetheless insightful. Also, I am thankful to him for a few shopping rides while I was new to the US. I would like to thank David Degen for continuous help with anything in the lab, from seeking an ancient machine to answering any other questions, as well as Yon Ebright for patiently answering my chemistry questions and helping me in interpreting mass spectroscopy data, as well as for teaching me how to cook rice.

I would like to thank all the members of the Nickels lab in particular Seth Goldman for teaching me a lot of experimental tricks, in form of running perfect sequencing gels to library preparation, and for many insightful discussion that we had; Irina Vvdenskaya for teaching me the arts of RNA library preparation and collaboration on chapters 3 and 5; Jeremy Bird, for his collaboration with me on chapter 5 by helping me with the pause assays to be done in a timely manner for publication.

I would also like to extend my gratitude to my collaborators, in particular to the lab of Prof. Carlos Bustamante (University of California, Berkeley) for providing me the opportunity to collaborate with them on single-molecule optical tweezer assays. Although I have not included the data in this thesis, it has been an enlightening collaboration. I would like to thank the lab of Prof. Finn Werner (University College London) for providing me with archaeal (*M. jannaschii*) RNAP; the lab of Prof. Craig Kaplan (Texas A&M University) for providing me with eukaryotic (*S. cerevisiae*) Pol II, as well as Yuanchao Zhang from the Deane Taylor lab (University of Pennsylvania) for helping me with the analysis of the next-generation-sequencing data.

I would also like to thank all the friends and peers of the bacterial transcription community, which I have had the pleasure to have discussions with during my attendance of FASEB meetings (three times), a transcription termination meeting (once), a phage

meeting (once) and the Biophysical Society meeting (four times). They have all been a source of excitement and have been insightful.

Finally, I am thankful to my parents for their sacrifice, never ending support, and for teaching me strength and independence in life.

Dedication

To my parents

Preface

A portion of the work presented in this dissertation has been published and additional sections are expected to appear in scientific journals in the future:

Parts of chapter 3 appear as published in PNAS 113, E2899-90 (2016).

Parts of chapter 5 appear as published in Science, 344, 1285-1289 (2014).

Parts of chapter 1 are in preparation for submission to PNAS.

Parts of chapter 6 and 5 will be combined with data from our collaborator (Carlos Bustamante lab, at UC Berkeley) and is expected to be published in the future.

Chapter 2 (role of CRE in DNA unwinding), chapter 4 (role of CRE in abortive transcription), and chapter 7 (sigma-independent transcription) are also expected to appear in scientific journals in the future.

Table of Contents

List of Tables

Section **Page**

List of Illustrations

Section **Page**

General Introduction

Chapter 1: determination of CRE consensus sequence and recognition mechanism

Section **Page**

Section **Page**

Section **Page**

General Introduction

1. Transcription reactions

Transcription is a process by which biological information encoded in a DNA sequence is copied into an RNA sequence by means of a molecular machine called RNA polymerase (RNAP) (Buc & Strick, 2009). Transcription has three stages: initiation, elongation and termination, which are comprised of a series of reactions carried out by RNAP (Fig 1). These reactions can be described as a series of kinetically and structurally distinct intermediate complexes (Fig. 1) which are listed below (Abbondanzieri, Greenleaf et al., 2005, Buc & Strick, 2009, Dangkulwanich, Ishibashi et al., 2014a, Haugen, Ross et al., 2008a, Losick, Chamberlin et al., 1976, Mustaev, Roberts et al., 2017, Revyakin, Liu et al., 2006, Ruff, Record et al., 2015, Saecker, Record et al., 2011):

(1) RNAP-promoter closed complex (RPc): RNAP binds to promoter DNA but it is not entered to the main channel and is still double stranded DNA.

(2) RNAP-promoter intermediate complex (RPi): The RNAP-DNA complex undergoes large conformational changes (referred to as isomerization) and DNA is loaded into the active center cleft.

(3) RNAP-promoter open complex (RPo): RNAP unwinds ~13 base pairs of promoter DNA surrounding the transcription start site (-11 t0 +2), forming a single-stranded region ("transcription bubble"). Clamp closes at this step.

(4) RNAP-promoter initial transcribing complex (RPitc): RNAP begins synthesis of an RNA product. During initial transcription, RNAP uses a "scrunching" mechanism, in

which RNAP remains stationary on promoter DNA, unwinds, and reels in downstream DNA in each nucleotide-addition cycle.

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

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

In addition to the on-pathway reactions during transcription as described above, kinetically significant off-pathway reactions occur. During initial transcription, RNAP can engage in off-pathway abortive cycles of synthesis and release of short RNA products (Hsu, 2009, Skancke, Bar et al., 2015). During elongation, RNAP can enter into, and return from, off-pathway states involving pausing (Artsimovitch & Landick, 2000a, Dalal, Larson et al., 2006, Herbert, Porta et al., 2006a, Neuman, Abbondanzieri et al., 2003a).

2. Bacterial RNA polymerase

RNA polymerase (RNAP) is the molecular machines that synthesis the RNA from the DNA. Several independent groups isolated this enzyme by ~1960 (Chamberlin & Berg, 1962, Hurwitz, 2005, Hurwitz, Furth et al., 1961). Around 40 years after RNAP discovery (around 2000) different structures were began to be available for different forms of this enzyme and over years of structural studies on them, it have been revealed that all multi-subunit RNAPs have a common structural framework and function by closely related molecular mechanism, see Figure 2, (Ebright, 2000, Werner, 2008, Werner & Grohmann,

2011a, Zhang, Campbell et al., 1999). RNAPs are either multi-subunit (bacteria, eukaryotes, archaea) or single subunit (some bacteriophage, mitochondria). Among multi-subunit RNAPs, bacteria have the simplest form (Fig.2). The bacteria RNAP core comprises of five subunit ($\alpha_1\alpha_2\beta\beta'\omega$) with around 400 kDa molecular mass and it recruit an additional subunit (σ) to form a holoenzyme. σ factor enables the enzyme to initiate transcription at specific promoter regions. This additional subunit has been believed that can dissociate from the enzyme after initiation step. RNAP structure reminiscent a crab's claw shape and has dimensions of ~150 Å x ~100 Å x ~100 Å. A channel is formed between the pincer of the claw which is called "primary channel" with ~25 Å in diameter, large enough to accommodate double-stranded DNA and it is in fact where the DNA is bound, another channel is the secondary channel "NTP entry channel" in which the NTPs entry to the active center through that, also there is another channel "RNA exit channel" which is for the exits of growing RNA chain (Ebright, 2000, Werner, 2008, Werner & Grohmann, 2011a, Zhang et al., 1999). For catalytic reaction RNAP need two Mg2+ ions. One is already present in the enzyme active center and the second one is brought in by the NTPs (Steitz & Steitz, 1993).

3. Bacterial promoter architecture.

Promoter sequence is a region of DNA that define where RNA polymerase bind and initiate transcription, see Fig 3, (Busby & Ebright, 1994a).

3.1 Core promoter elements. Part of the promoter that is covered by the core subunit is called core promoter region (Fig 3). In bacteria, the primary determinants of core promoter recognition were believed to be mainly determined by two conserved hexametric element centered at -35 and -10 elements from the start site and separated by 17 base pair

linker (Haugen et al., 2008a, McClure, Hawley et al., 1983, Pribnow, 1975b). However, after 4 decades of research we now know about additional promoter elements including extended -10 element and discriminator element which although are not highly conserved in promoters but can still play significant roles in promoter recognition. These core promoter elements are recognized by sigma factor (Haugen et al., 2008a).

3.2 Upstream promoter (UP) elements. In addition to above mentioned core promoter elements, other elements have been discovered which are outside the core promoter region, upstream of -35 element that are recognized by α-subunits (Fig. 3). These promoter regions are called upstream promoter (UP) elements (Busby & Ebright, 1994b, Haugen et al., 2008a, Ross, Gosink et al., 1993). UP element like extended -10 and discriminator element are also not very conserved across promoters but can still play a significant role in promoter recognition, regardless of the presence of consensus sequences (Fig. 3 and 4) (Haugen, Ross et al., 2008b).

4. Interactions between RNA polymerase and promoter elements

4.1. Interactions between sigma and core promoter elements. Sigma factor binding to RNAP core (RNAP holo enzyme formation) exert promoter-specific transcription initiation, because Sigma factors have several DNA-binding motifs for sequence-specific promoter recognition (Fig 3). For example, *E. coli* σ70, a primary sigma factor, includes four domains. Domains 2, 3 and 4 are responsible for the recognition of the -10 element, the extended -10 element and the -35 element, respectively. These elements are part of the core promoter region (i.e., nucleotide sequence -40 through +20 relative to the transcription start-site +1; Fig. 3) which is covered by the bulk body of the RNAP core (Busby & Ebright, 1994b, Haugen et al., 2008a). -35 element is recognized by a helix-turn-helix motif

of σ-region 4, which contacts backbone and bases on both template and non-template strands. These interactions are considered among the first sequence specific interactions made between RNAP and the promoter reviewed in (Haugen et al., 2008a). The extended -10 element major groove is recognized by surface-exposed residues in a helix from σ-region 3. These interactions are especially important in promoters, which do not have a good homology score relative to the consensus for their -35 or -10 element. For example, mutating the extended -10 element in a promoter that lack the consensus -35 element can abolish the transcription. In addition, the prevalence of this element is different in different bacteria and is mainly observed in gram-positive promoters. It looks like the extended -10 element can stimulate transcription through increasing the rate of RPc and RPo formation and can improve the maintenance of the DNA strand separation (Camacho & Salas, 1999, Voskuil & Chambliss, 2002).

The -10 element is recognized by σ-region 2 both as double-stranded DNA (before melting) and as single-stranded DNA (after melting). Therefore, mutation in -10 element can affect both KB and ki. Melting may involve base flipping at -11 and staking of an aromatic residue of sigma 2.3 on the base at -12 is proposed to prevent DNA unwinding further upstream (Feklistov & Darst, 2011, Zhang, Feng et al., 2012).

Discriminator element is recognized by σ-region 1.2. These interactions affect the rate of dissociation of RPo to RPc, which is responsible for their response to negative regulator of ppGpp, and DksA Discriminator element is mainly observed for rRNA and tRNA promoters (Haugen et al., 2008a, Travers, 1980, Zhang et al., 2012)

4.2. Interactions between RNAP core and UP elements. The c-terminal domains (α-CTDs) of α subunits of RNAP core contain DNA binding determinants (Fig 3). α-CTDs

recognize the upstream promoter elements but not any of the core promoter elements (Estrem, Ross et al., 1999, Haugen et al., 2008a, Ross et al., 1993). α-CTDs uses a helix–hairpin–helix motif to interacts with DNA minor groove. The two αCTDs could bind to two DNA site upstream of the –35 elements, however, they can also interact with DNA site is 5–6 minor grooves upstream of the –35 hexamer. This is due to the flexible linker that connects the two globular domains of each α subunits and ability of the DNA to bend.

4.3. Interactions between RNAP core and core promoter elements. Until recently, it ha been assumed that the RNAP core does not contain the determinants for sequence-specific core promoter DNA recognition and that, the RNAP core requires binding to an accessory specificity factor (σ-factor) for core promoter DNA recognition. Although, a fragment of the RNAP core β' subunit (i.e., jaw-domain) interacts with downstream dsDNA, but these interactions are assumed to be non-sequence-specific (Drennan, Kraemer et al., 2012, Ederth, Artsimovitch et al., 2002, James, Liu et al., 2012). Additionally, another fragment of RNAP core β' subunit (e.g., β'-zipper-domain) has been proposed to interact with the promoter DNA spacer region but only with a weak sequence-specificity (Bae, Feklistov et al., 2015, Yuzenkova, Tadigotla et al., 2011). In both cases (i.e., jaw-domain and β'-zipper-domain), the sequence-specificity has not been definitively determined and remains to be validated.

The above assumptions have been challenged by several lines of study, which suggest that the RNAP core makes direct sequence-specific interactions with DNA. Initial studies have suggested that the β subunit residues 84 to 642 can be crosslinked in solution to the non-template DNA strand positions -4 through +2 (Naryshkin, Revyakin et al., 2000). This has also been evidenced by the crystal structure of RPo that indicates the RNAP core β

subunits interact with the transcription-bubble-nontemplate strand segment corresponding to positions -4 to +2 (which has been designated as core recognition element, CRE) (Fig. 3 and Fig. 4) (Zhang et al., 2012). The interaction involves the unstacking and insertion of the +2G base into a pocket formed by βR151, βI445, βD446, βR451, βT539, and βV547 (β pocket); the stacking of the +1T base on aromatic residue βW183; H-bond formation with the -2G base by βD199 and βR394; and H-bond formation with the -3A and -4G bases by βR371(Fig. 3 and 4). Moreover, for each RNAP residue that participates in RNAP-CRE interactions with CRE position +2, a single-Ala substitution, (and for two residues, "pocket-filling" single-Phe and single-Trp substitutions) resulted in defects in transcription, indicating that these residues are functionally important for transcription. Moreover, the biochemical data suggest that the interaction at one of these positions (i.e., CRE position +2) is sequence-specific, providing the first evidence that the RNAP core is capable of making direct sequence-specific interactions with core promoter DNA (Zhang et al., 2012).

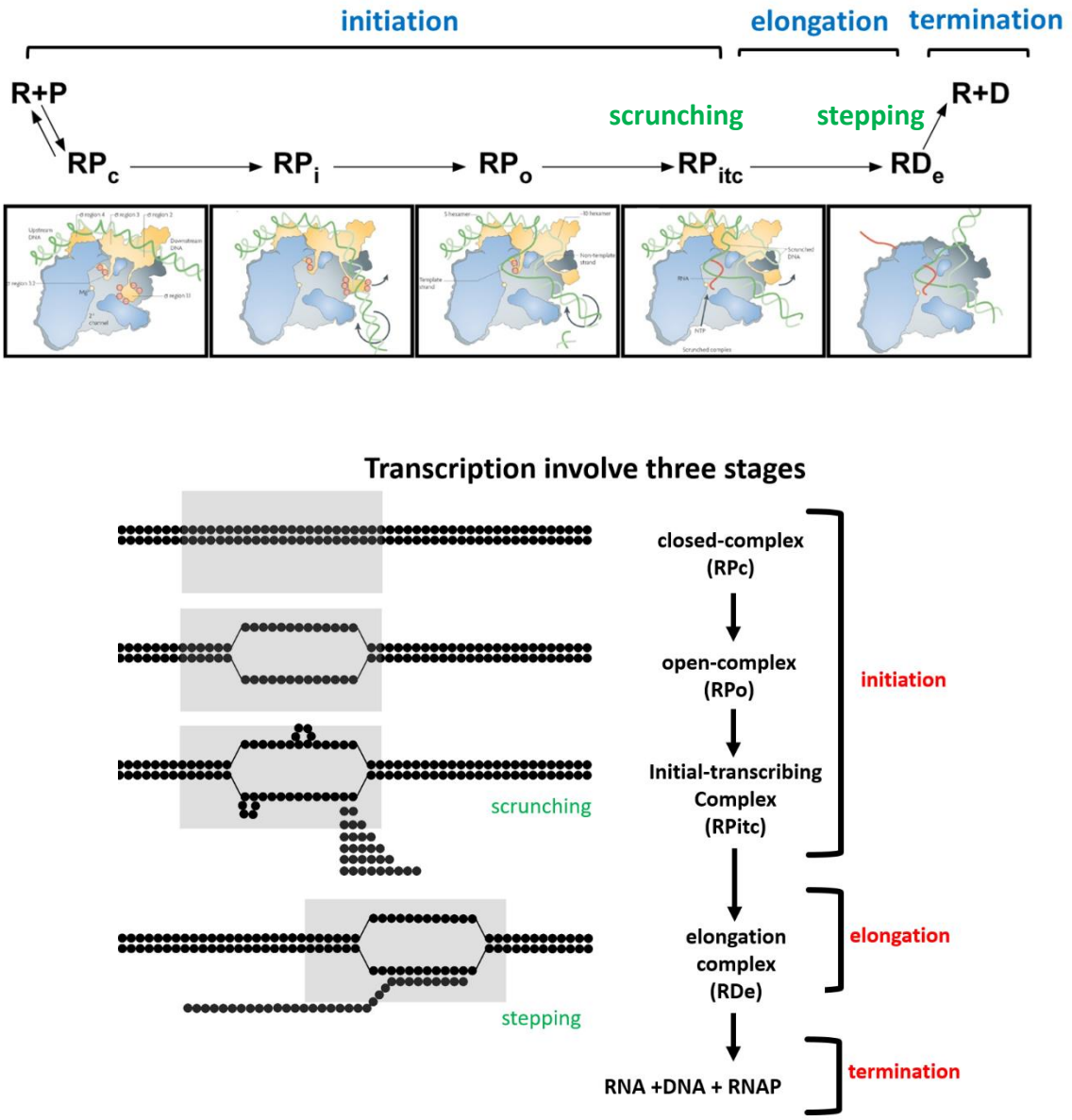
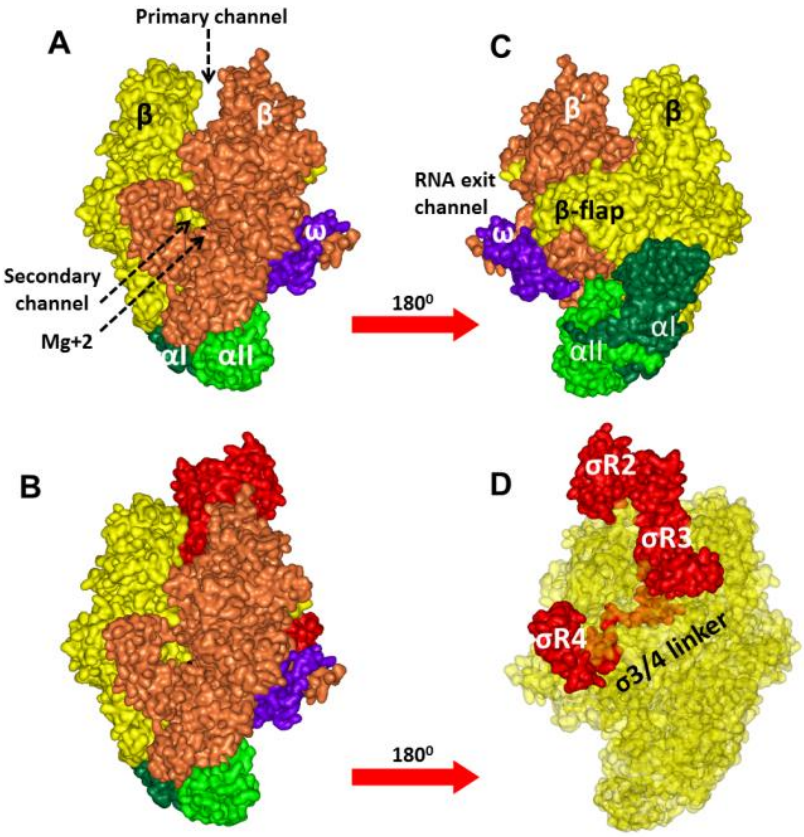


Figure 1. Transcription involve three stages: initiation, elongation and termination. R, RNAP; P, promoter; RPc, closed complex; RPitc, initial transcribing complex; RDe, transcription elongation complex R, RNAP; P, promoter; RPc, closed complex; RPitc, initial transcribing complex; RDe, transcription elongation complex. Schematic structures obtained from (Murakami & Darst, 2003).



E

Bacteria	Archaea	Eukaryotes				
		RNAPII	RNAPIII	RNAPI	Plant RNAPIV‡	Plant RNAPV‡
RNAP subunits						
β-subunit	Rpo1 (RpoA)	RPB1	C160	A190	NRPD1	NRPE1
β-subunit	Rpo2 (RpoB)	RPB2	C128	A135	NRPD/E2	NRPD/E2
α-subunit	Rpo3 (RpoD)	RPB3	AC40	AC40	RPB3 [1]	RPB3 [1]
α-subunit	Rpo11 (RpoL)	RPB11	AC19	AC19	RPB11	RPB11
ω-subunit	Rpo6 (RpoK)	RPB6	RPB6	RPB6	RPB6 [1]	RPB6 [1]
	Rpo5 (RpoH)	RPB5	RPB5	RPB5	RPB5 [3]	NRPE5
	Rpo8§ (RpoG)	RPB8	RPB8	RPB8	RPB8 [1]	RPB8 [1]
	Rpo10 (RpoN)	RPB10	RPB10	RPB10	RPB10	RPB10
	Rpo12 (RpoP)	RPB12	RPB12	RPB12	RPB12	RPB12
	Rpo4 (RpoF)	RPB4	C17	A14	NRPD/E4	NRPD/E4
	Rpo7 (RpoE)	RPB7	C25	A43	NRPD7 [1]	NRPE7
		RPB9	C11	A12	NRPD9b	RPB9

Figure 2. Bacterial RNAP. (A and C) RNAP core subunits "core enzyme". (B and D) RNAP core subunits associated with sigma factor "holo-enzyme". (E) conservation of RNAP subunits across three domain of life bacteria, archaea and eukaryotes (table obtained from (Werner & Grohmann, 2011b)).

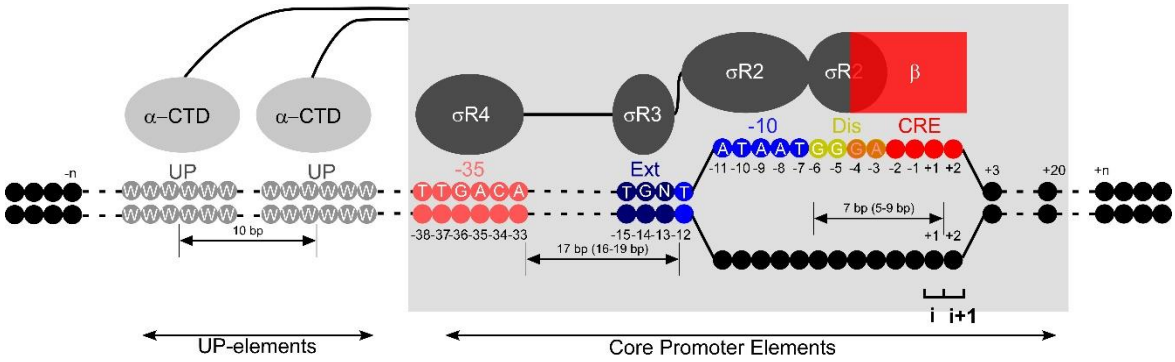


Figure 3. Bacterial promoter architecture. Promoter DNA elements and their sequence-specific interacting protein partner from bacterial RNAP.

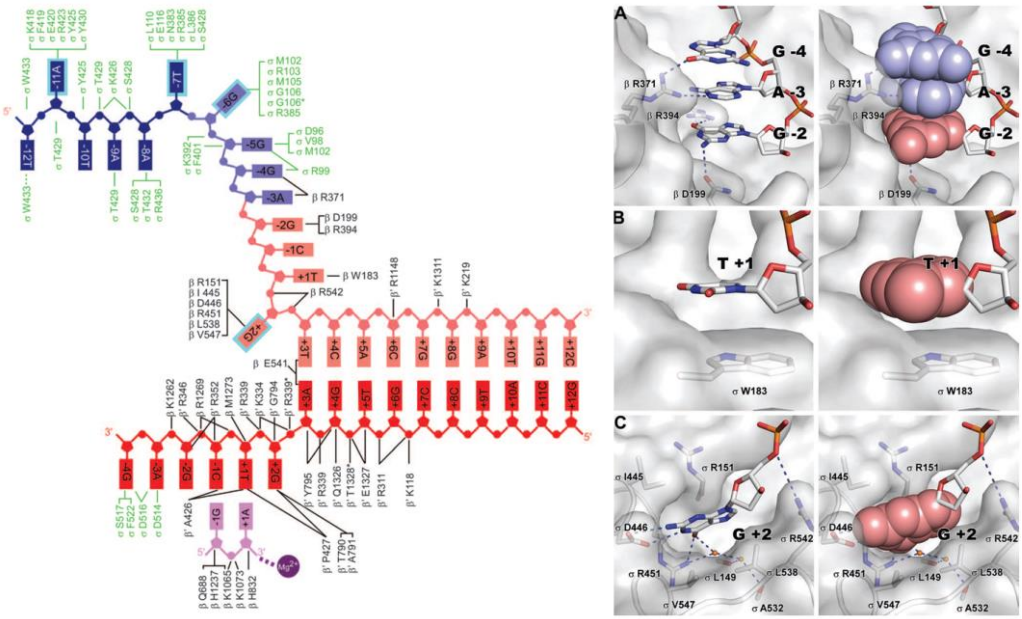


Figure 4. RNAP-DNA open complex structure. Left panel, two-dimensional schematic of interactions of -10 elements, discriminator elements and core recognition element (non-template ssDNA region of -4 to +2). Right panel, the details of CRE interactions with RNAP residues (images obtained from (Zhang et al., 2012)).

Chapter 1

Determination of CRE consensus sequence and recognition mechanism

Hanif Vahedian-Movahed[1], Yu Zhang and Richard H. Ebright

1.1. Background

Until recently, it had been assumed that the RNAP core does not contain the determinants for sequence-specific core promoter DNA recognition and that the RNAP core requires binding to an accessory specificity factor (σ-factor) for core promoter DNA recognition. Although, the c-terminal domains (α-CTDs) of α subunits of RNAP core do contain DNA binding determinants, α-CTDs recognize the UP elements and not any of the DNA elements of the core promoter regions (Estrem et al., 1999, Haugen et al., 2008a, Ross et al., 1993). Also, a fragment of the RNAP core β' subunit (i.e., jaw-domain) interacts with downstream dsDNA, but these interactions are assumed to be non-sequence-specific (Drennan et al., 2012, Ederth et al., 2002). Additionally, another fragment of RNAP core β' subunit (e.g., β'-zipper-domain) has been proposed to interact with the promoter DNA spacer region but only with a weak sequence-specificity (Bae et al., 2015, Yuzenkova et al., 2011). In both cases (i.e., jaw-domain and β'-zipper-domain), the sequence-specificity has not been definitively determined and remains to be validated.

[1] I performed all the experiments except for the crystal structures presented in the last section were YZ performed the experiments and provided the data.

Until relatively recently, it had also been assumed that sequence-specific RNAP-DNA interactions are mainly limited to transcription initiation, except with respect to sigma-factor retention during the initiation-elongation transition stage (Mukhopadhyay, Kapanidis et al., 2001, Nickels, Mukhopadhyay et al., 2004) or when sigma-factor (as a trans regulatory element) associates with an elongating RNAP core . In such cases, the sequence-specific interactions during the elongation stage (which is mediated by a sigma-factor and not the RNAP core) can cause σ-dependent pausing, which would have important functional significance (e.g., in recruiting a transcription factor to the elongation complex) (Goldman, Nair et al., 2015).

The above assumptions have been challenged by several lines of study which suggest that the RNAP core makes direct sequence-specific interactions with DNA, not only in transcription initiation but also in transcription elongation (chapter 4). Initial studies have suggested that the β subunit residues 84 to 642 can be crosslinked in solution to the non-template DNA strand positions -4 through +2 (Naryshkin et al., 2000). This has also been evidenced by the crystal structure of RPo that indicates the RNAP core β subunits interact with the transcription-bubble nontemplate strand segment corresponding to positions -4 to +2 (which has been designated as core recognition element, CRE) (Zhang et al., 2012). Moreover, the biochemical data suggest that the interaction at one of these positions (i.e., CRE position +2) is sequence-specific, providing the first evidence that the RNAP core is capable of making direct sequence-specific interactions with DNA (Zhang et al., 2012).

Although, it has already been determined that the interaction at CRE position +2 is sequence-specific, it has not yet been determined whether RNAP-CRE interactions involving positions -4, -3, -2, -1 and +1 are sequence-specific; and it is not clear which of

the residues of β subunits mediate sequence-specificity. Detailed analysis of sequence specificity and the recognition mechanism of CRE could reveal its involvement in regulatory events that occurs in initiation or post-initiation stage of transcription. Therefore, it is crucial to determine the full extent of RNAP-CRE interactions.

In this study, we determined the full extent of RNAP-CRE interactions for all CRE positions (positions -4 through +1) in that we established that RNAP-CRE interactions are sequence-specific at least at two other positions in addition to the previously characterized position +2 (sequence-specificity at 3-out-of-6 positions). We also determined the individual residues of the RNAP core beta subunit that mediate sequence-specificity for all of the three sequence-specific positions (i.e., positions -4, +1 and +2). Furthermore, we have determined the structural basis of sequence-specificity for the two strongest positions (position +1 and +2). Our results show that the RNAP core makes direct sequence-specific interactions with DNA, a phenomenon which has previously been attributed to sigma factors and to the transcription initiation stage. We propose that this interaction could happen at each stage of transcription and so potentially play functional roles at each stage of transcription. The RNAP-mutants generated in this work provides a tool for future functional analysis of any individual sequence-specific CRE position (see chapter 2).

1.2. Results and discussions

1.2.1. Sequence specificity of RNAP-CRE interactions

The crystal structure of RPo indicates that the RNAP core interact with 5-out-of-6 bases of CRE nucleotides (Fig. 1). Previous biophysical results indicate that RNAP exhibits strong specificity for G at CRE position +2 (in that G >> T ~ A ~ C). To determine whether RNAP-CRE interactions are sequence-specific at other CRE positions +1, -1, -2, -3 and -

4, I constructed all possible nucleotide substitutions at each CRE position, and assessed effects on RNAP-DNA interaction in a fluorescent-based equilibrium binding experiments and high-salt-induced-dissociation off-rate experiments. The assay used here take advantage of high-specificity [TMR211-σ70]-RNAP fluorescent protein-based beacon assay (see methods for the strategy). My results indicate that in addition to the CRE position +2, RNAP exhibits sequence specificity toward positions +1 and -4 but no specificity toward positions -1, -2 and -3.

My results showed that at position +1 RNAP exhibits strong specificity for T (T > C >> A ~ G). RNAP has an equilibrium dissociation constant that is lower by a factor of 20, and an off-rate value that is lower by a factor of 32 for a promoter derivative having a T at position +1. Moreover, RNAP-DNA complexes formed with DNA derivatives that have a non-consensus base "A" at position +2 and a consensus base "T" at position +1, reduced the complex stability by a factor of 10 but does not eliminate binding (Fig. 3). This indicates that the interaction at position +1 can be formed independent of the sequence-specific interaction at position +2. Furthermore, RNAP-DNA complexes formed with DNA derivative that have a deoxyuridine (dU) at position +1 compared with sequences that have a +1 T shows no change in the stability complex as inferred from off-rate values. This indicates that the methyl group of the T is not important for the high-specificity interactions. Possibly, the presence of two carbonyl groups make the pyrimidine ring of T uniform and preferable for proper stacking interaction with Trp 183 as compared to C in which the pyrimidine ring possess an amine group and a carbonyl group.

My results also show that at position -4, RNAP exhibits moderate specificity for T or G (T ~ G > A ~ C). RNAP has an equilibrium dissociation constant that is lower by a factor

of ~13 or ~4 for a promoter derivative having a T or a G at position -4, respectively. Additionally, RNAP has an off-rate value that is lower by a factor of ~5 or ~4 for a promoter derivative having a T or a G at position -4, respectively (Fig 2).

I also performed analogous equilibrium binding and dissociation kinetic experiments for position -1, -2 and -3, and the results suggest that the interactions at these positions are not strongly sequence-specific (Fig. 4). We conclude that RNAP-CRE interactions shows sequence-specificity at 3-out-of-6 positions and that the consensus sequence for CRE is $T_{-4}\, N_{-3}\, N_{-2}\, N_{-1}\, T_{+1}\, G_{+2}$ (Fig. 2).

1.2.2. RNAP residues that mediate sequence specificity of RNAP-CRE interactions

To identify individual RNAP amino acids that mediate specificity at CRE positions +2, +1 and -4, I constructed single Ala substitutions of RNAP residues that contact CRE positions +2, +1 and -4 and assessed effects on RNAP-DNA interactions with promoter derivatives containing all possible nucleotide substitutions at CRE positions +2, +1 and -4. To enable the use of the molecular beacon assay (described in previous section), I generated a fluorescently labelled holo enzyme reconstituted from a mutant core enzyme and a fluorescently labelled sigma factor [TMR211-σ70] for each single Ala substituted RNAP derivative to be tested.

My equilibrium binding and dissociation kinetics results show that Ala substitution of βR151, βD446, or βR451 eliminates specificity at CRE position +2 (Fig. 5. and 6). Additionally, analogous experiments using "pocket-filling" Phe and Trp substitution of one of this residue “βD446” also eliminate specificity at CRE position +2 (Fig. 7 and 8). These

bulky amino acids could prevent the access of the base into the pocket. Therefore, we conclude that βR151, βD446, or βR451 residues determine specificity at position +2.

Next, we tested the effect of Ala substitution of βW183 which contact CRE position +1. The results show Ala substitution of βW183 eliminates specificity at CRE position +1 (Fig 9). We conclude that βW183 residue determine specificity at position +1.

Next, we tested the effect of Ala substitution of βR371 which contact CRE position -4. The results show Ala substitution of βR371 eliminates specificity at CRE position -4 (Fig. 10). Therfore, βR371 residue determine specificity at position -4.

All in all, we conclude that βR151, βD446, or βR451 mediate specificity at CRE position +2, βW183 mediates specificity at CRE position +1, and βR371 mediates specificity at CRE position -4.

1.2.3. Structural basis of sequence specificity of RNAP-CRE interactions

To define the structural basis of specificity at CRE positions +1 and +2, my lab mate Yu Zhang, determined crystal structures of RPo derivatives containing all possible nucleotide substitutions at CRE positions +2 and +1 (Fig. 11.) ; positions which showed the highest specificity. The structures of RPo derivatives containing substitutions at CRE position +2 indicate that, in contrast to the consensus base G which inserts into the β pocket, the non-consensus bases A, T, and C do not insert into the β pocket. The structures of RPo derivatives containing substitutions at CRE position +1 indicate that, in contrast to the consensus base T which stacks squarely on βW183, the non-consensus bases G, A, and C do not stack on, or only partially stack on, βW183. We conclude that specificity at CRE

positions +2 and +1 manifests itself not only in quantitative differences in binding thermodynamics and kinetics, but also in qualitative differences in structure.

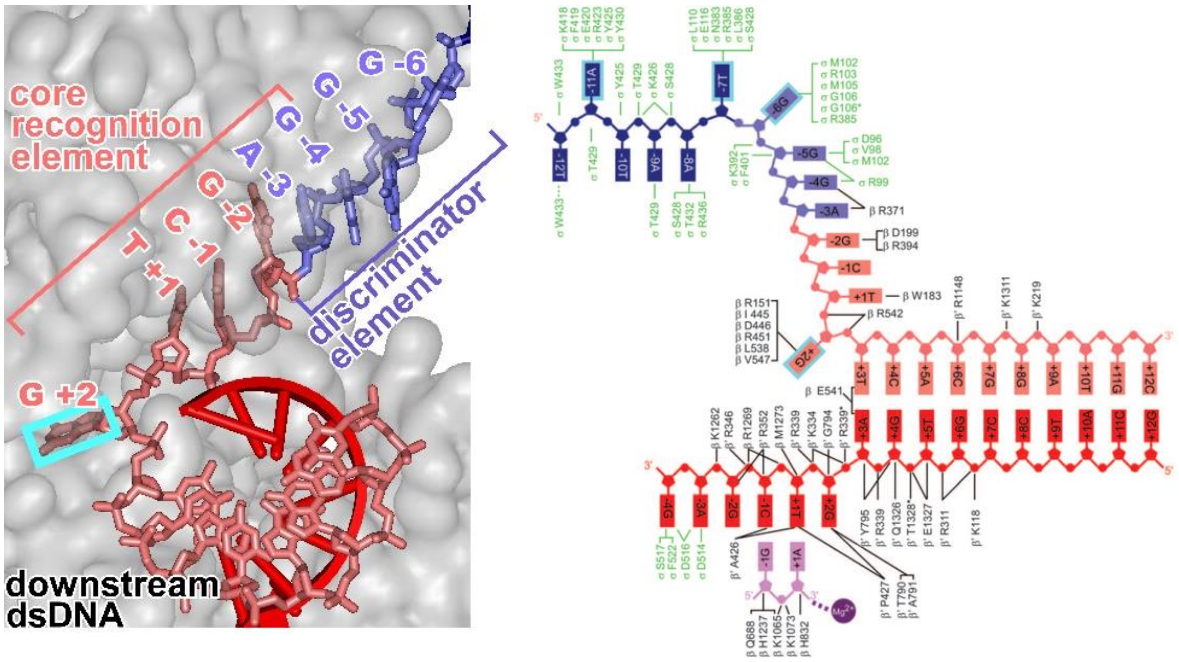


Figure 1. Overview of the "Core Recognition Element", (CRE). (Left panel) View of the downstream edge of the transcription bubbles. Grey, β-subunit of RNAP; light blue, discriminator element of DNA nontemplate strand; pink, DNA nontemplate strand downstream of discriminator element; red, DNA template strand; cyan box, base +2 unstacked and inserted into β-pocket. (Right panel), two-dimensional schematic of interactions of -10 elements, discriminator elements and core recognition element (non-template ssDNA region of -4 to +2 (images obtained from (Zhang et al., 2012)).

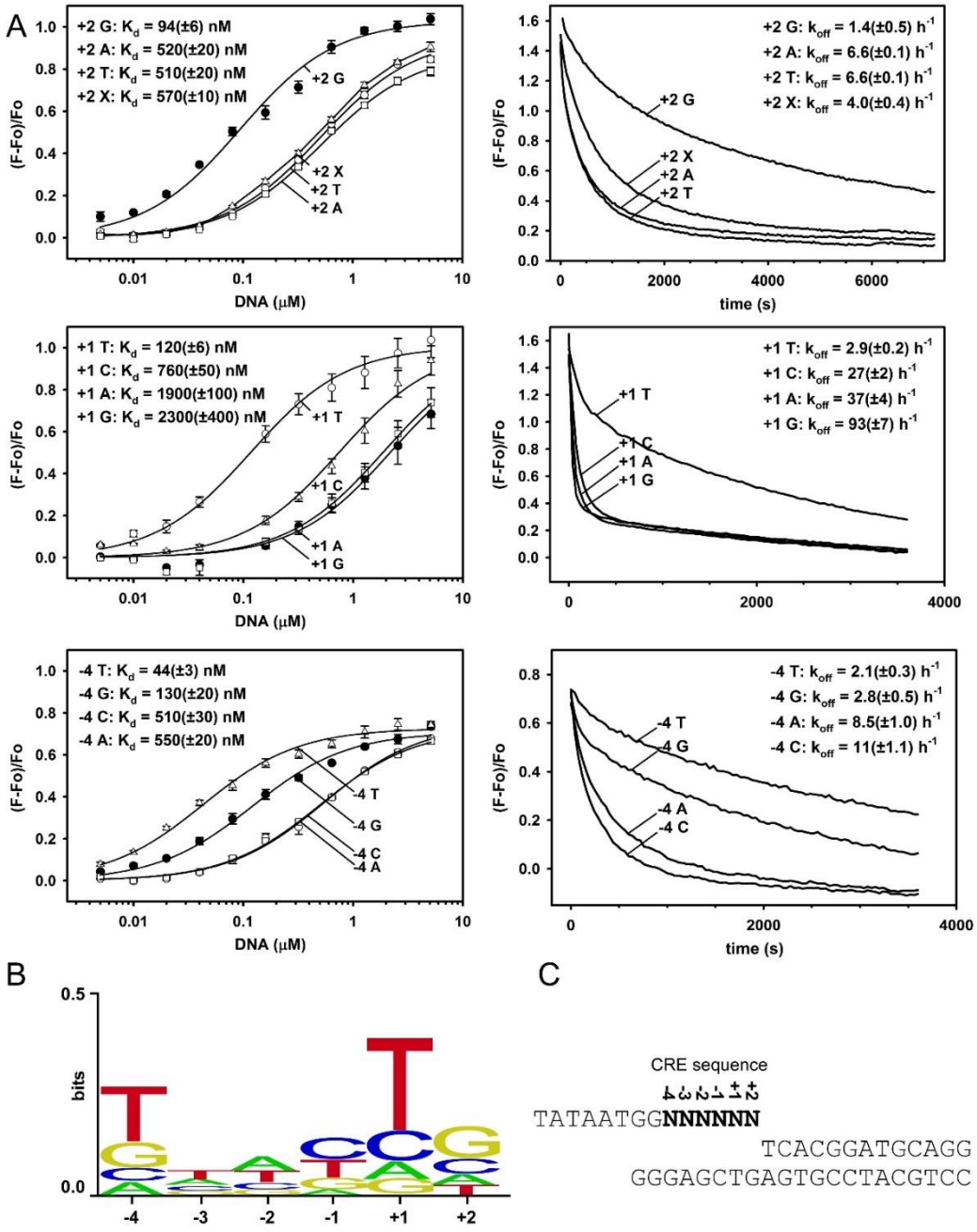


Figure 2. Specificity of RNAP-CRE interactions. A. equilibrium binding (left) and dissociation kinetics (right) for CRE position +2, +1, and -4. B. Consensus sequence for CRE based on energy matrix of the interactions derived from equilibrium binding data of position +2, +1, -1, -2, -3 and -4. C. Sequences of the scaffolds used in fluorescent beacon assays of [TMR211-σ70]-RNAP. N indicates the single-base-substitution positions.

A

TATAATGGGAGC**TG**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**CG**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**AG**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**GG**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

B

TATAATGGGAGC**TA**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**CA**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**AA**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**GA**
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

DNA site	RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{off,co}$ ns
+1T	2.9	[1]
+1C	27	9
+1A	37	13
+1G	93	32
+1dU	3.1	1

DNA site	RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{off,co}$ ns
+1T;+2A	23	[1]
+1C;+2A	30	1
+1A;+2A	72	3
+1G;+2A	110	5

Figure 3. Specificity of RNAP-CRE interactions of position +1. Dissociation kinetics data for Specificity of CRE positions +1 in the presence (A) or absence (B) of sequence-specific interactions at position +2.

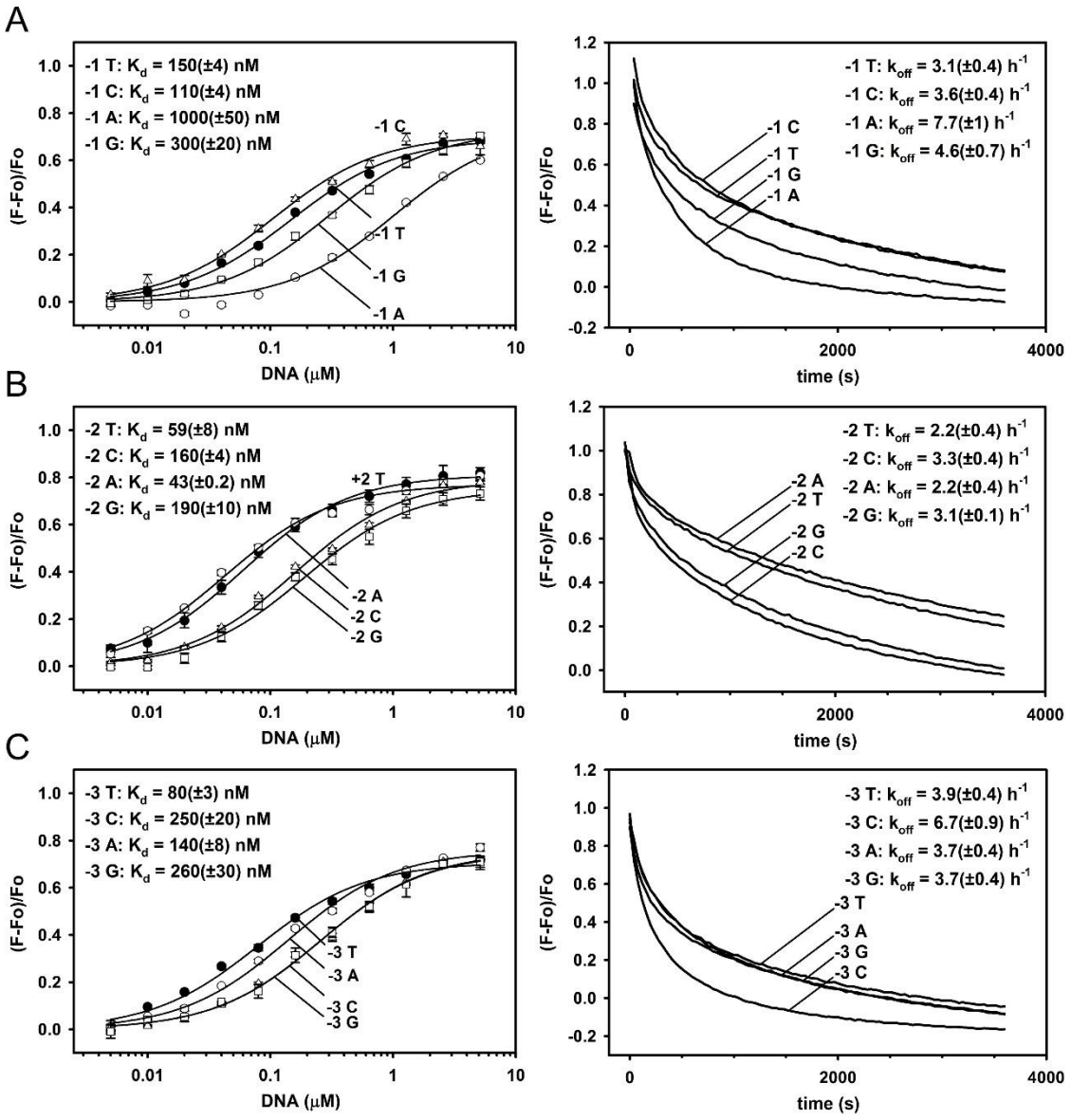


Figure 4. Specificity of RNAP-CRE interactions of positions -1, -2, -3. Equilibrium binding (left) and dissociation kinetics (right) for Specificity of CRE positions -1, -2, and -3.

A

```
TATAATGGGAGCTG                  TATAATGGGAGCTT
             TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTGAGTGCCTACGTCC           GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGCTA                  TATAATGGGAGCTX
             TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTGAGTGCCTACGTCC           GGGAGCTGAGTGCCTACGTCC
```

B

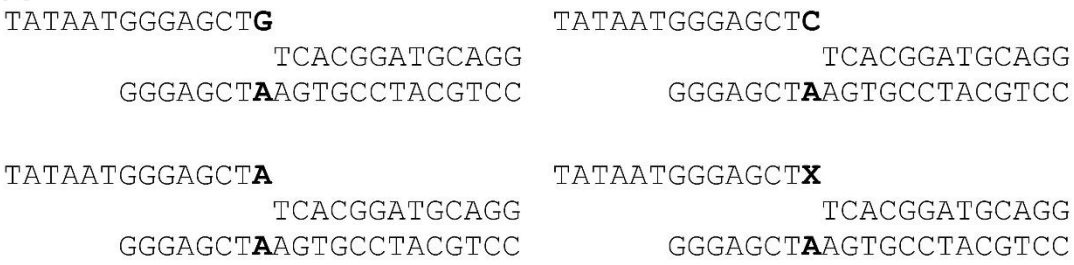
DNA site	RNAP		[Ala151]β-RNAP		[Ala446]β-RNAP		[Ala451]β-RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$
+2G	1.4	[1]	10	[1]	11	[1]	100	[1]
+2X	4.8	3	8.7	0.8	5.9	0.5	140	1
+2A	6.5	5	13	1	16	1	210	2
+2T	7.4	5	17	2	14	1	170	2

C

DNA site	RNAP		[Ala151]β-RNAP		[Ala446]β-RNAP		[Ala451]β-RNAP	
	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$
+2G	0.13	[1]	1.1	[1]	0.93	[1]	6.3	[1]
+2X	0.57	4	1.0	1	0.56	1	8.4	1
+2A	0.85	6	1.2	1	0.79	1	7.8	1
+2T	0.81	6	1.8	2	0.90	1	11	2

Figure 5. Biochemical basis of specificity of CRE position +2. (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. Note the "G" base at the template stand position +2 (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]-β-RNAP, [TMR211-σ70][Ala151]-β-RNAP, [TMR211-σ70][Ala446]–β-RNAP and [TMR211-σ70][Ala451]-β-RNAP) derivatives in beacon assays.

A



B

DNA site	RNAP		[Ala151]β-RNAP		[Ala446]β-RNAP		[Ala451]β-RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{d,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,c}$ ns
+2G	0.8	[1]	4.5	[1]	11	[1]	76	[1]
+2X	5.1	6	13	3	18	2	190	2
+2A	5.2	6	13	3	15	1	190	2
+2C	3.6	4	9.3	2	9.4	1	160	2

C

DNA site	RNAP		[Ala151]β-RNAP		[Ala446]β-RNAP		[Ala451]β-RNAP	
	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,co}$ s
+2G	0.11	[1]	0.62	[1]	1.0	[1]	5.9	[1]
+2X	0.77	7	1.1	2	1.2	1	9.5	2
+2A	0.65	6	1.1	2	1.2	1	6.2	1
+2C	0.43	4	0.6	1	0.64	1	5.3	1

Figure 6. Biochemical basis of specificity of CRE position +2. (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. Note the "A" base at the template stand position +2 (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]-β-RNAP, [TMR211-σ70][Ala151]-β-RNAP, [TMR211-σ70][Ala446]–β-RNAP and [TMR211-σ70][Ala451]-β-RNAP) derivatives in beacon assays.

A

```
TATAATGGGAGCTG                  TATAATGGGAGCTT
             TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTGAGTGCCTACGTCC           GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGCTA                  TATAATGGGAGCTX
             TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTGAGTGCCTACGTCC           GGGAGCTGAGTGCCTACGTCC
```

B

DNA site	RNAP		[Phe446]β-RNAP		[Trp446]β-RNAP	
	k_{off} (h^{-1})	$k_{off,X}/k_{off,con}$ s	k_{off} (h^{-1})	$k_{off,X}/k_{off,con}$ s	k_{off} (h^{-1})	$k_{off,X}/k_{off,con}$ s
+2G	1.4	[1]	6.6	[1]	5.0	[1]
+2X	4.8	3	4.0	0.6	3.5	0.7
+2A	6.5	5	7.5	1	7.9	2
+2T	7.4	5	5.7	0.9	5.1	1

C

DNA site	RNAP		[Phe446]β-RNAP		[Trp446]β-RNAP	
	K_d (uM)	$K_{d,X}/K_{d,con}$ s	K_d (uM)	$K_{d,X}/K_{d,con}$ s	K_d (uM)	$K_{d,X}/K_{d,con}$ s
+2G	0.13	[1]	1.1	[1]	0.56	[1]
+2X	0.57	4	0.70	0.7	0.51	0.9
+2A	0.85	6	1.3	1	0.77	1
+2T	0.81	6	1.1	1	0.56	1

Figure 7. Biochemical basis of specificity of CRE position +2 (pocket-filling mutants). (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. Note the "G" base at the template stand position +2 (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]b-RNAP, [TMR211-σ70][Phe446]b-RNAP and [TMR211-σ70][Trp446]b-RNAP) derivatives in beacon assays.

A

```
TATAATGGGAGCTG                  TATAATGGGAGCTC
              TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTAAGTGCCTACGTCC           GGGAGCTAAGTGCCTACGTCC

TATAATGGGAGCTA                  TATAATGGGAGCTX
              TCACGGATGCAGG                   TCACGGATGCAGG
      GGGAGCTAAGTGCCTACGTCC           GGGAGCTAAGTGCCTACGTCC
```

B

DNA site	RNAP		[Phe446]β-RNAP		[Trp446]β-RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{d,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$
+2G	0.8	[1]	6.8	[1]	3.1	[1]
+2X	5.1	6	11	2	5.1	2
+2A	5.2	6	13	2	6.8	2
+2C	3.6	4	7.7	1	3.7	1

C

DNA site	RNAP		[Phe446]β-RNAP		[Trp446]β-RNAP	
	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$
+2G	0.11	[1]	0.75	[1]	0.57	[1]
+2X	0.77	7	0.92	1	0.66	1
+2A	0.65	6	0.95	1	0.81	1
+2C	0.43	4	0.48	1	0.37	1

Figure 8. Biochemical basis of specificity of CRE position +2 (pocket-filling mutants). (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. Note the "A" base at the template stand position +2 (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]b-RNAP, [TMR211-σ70][Phe446]b-RNAP and [TMR211-σ70][Trp446]b-RNAP) derivatives in beacon assays.

A

TATAATGGGAGC**T**G
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**C**G
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**A**G
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGGGAGC**G**G
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

B

DNA site	RNAP		[Ala183]β-RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$
+1T	2.9	[1]	70	[1]
+1C	27	9	90	1
+1A	37	13	140	2
+1G	93	32	140	2

C

DNA site	RNAP		[Ala183]β-RNAP	
	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$
+1T	0.12	[1]	3.8	[1]
+1C	0.76	6	4.8	1
+1A	1.9	20	16	4
+1G	2.3	20	11	3

Figure 9. Biochemical basis of specificity of CRE position +1. (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]b-RNAP and [TMR211-σ70][Ala183]b-RNAP) derivatives in beacon assays.

A

TATAATGG**G**AGCTG
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGG**T**AGCTG
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGG**C**AGCTG
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

TATAATGG**A**AGCTG
TCACGGATGCAGG
GGGAGCTGAGTGCCTACGTCC

B

DNA site	RNAP		[A371]β-RNAP	
	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$	k_{off} (h^{-1})	$k_{off,x}/k_{off,cons}$
-4T	2.1	[1]	11	[1]
-4C	11	5	32	3
-4A	8.5	4	10	1
-4G	2.8	1	15	1

C

DNA site	RNAP		[A371]β-RNAP	
	K_d (μM)	$K_{d,x}/K_{d,cons}$	K_d (μM)	$K_{d,x}/K_{d,cons}$
-4T	0.04	[1]	0.75	[1]
-4C	0.51	12	1.81	2
-4A	0.55	13	1.38	2
-4G	0.13	3	1.30	2

Figure 10. Biochemical basis of specificity of CRE position -4. (A) Sequences of DNA scaffolds used to detect interaction with RNAP derivative. (B) Dissociation kinetics values (C) equilibrium dissociation constants of fluorescently labelled RNAP derivatives ([TMR211-σ70]b-RNAP and [TMR211-σ70][Ala371]b-RNAP) derivatives in beacon assays.

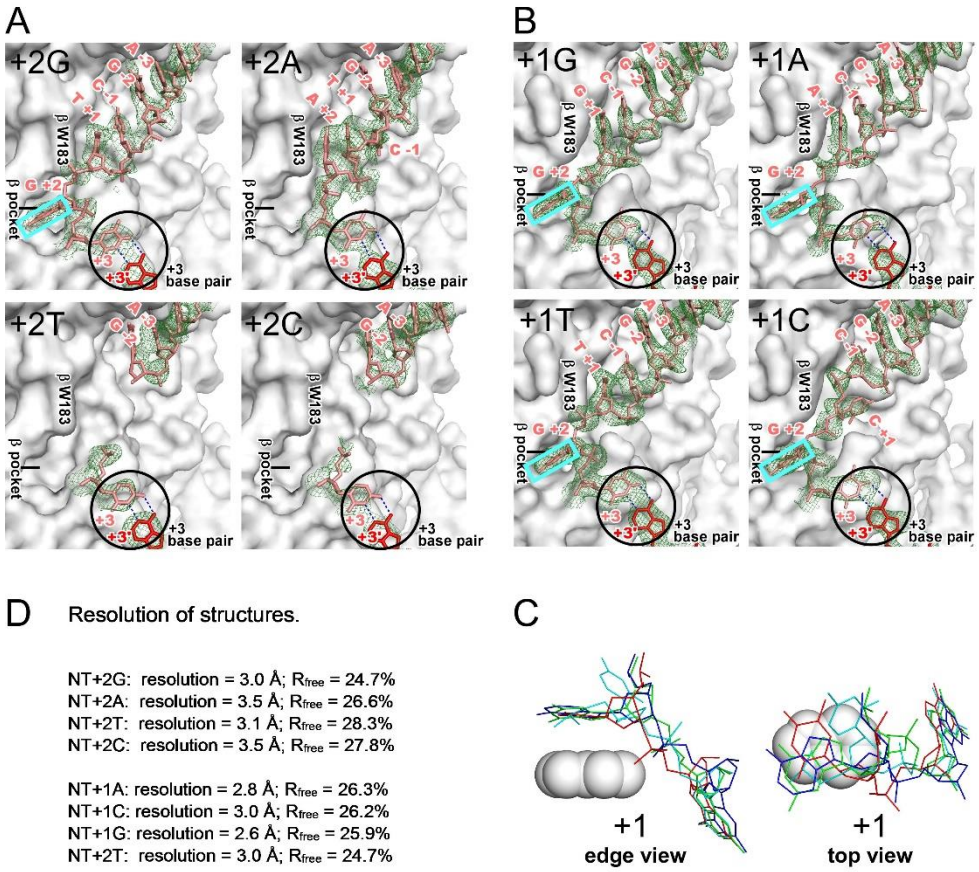


Figure 11. Structural basis of specificity of CRE positions +2 and +1. (A) RPo structures for single base substitutions at position +2 of non-template strand (CRE position +2). Grey, β-subunit of RNAP; pink, DNA nontemplate strand; red, DNA template strand; cyan box, bases unstacked and inserted into β pockets; black circle, the first base pair of downstream duplex. Residues are numbered as in *E. coli* RNAP. (B) RPo structure for single base substitutions at +1 of non-template strand (CRE position +1). Grey, β subunit of RNAP; pink, DNA nontemplate strand; red, DNA template strand; cyan box, bases unstacked and inserted into , β-pocket; black circle, the first base pair of downstream duplex. Residues are numbered as in *E. coli* RNAP. (C) Edge view and top view of the base-stacking at position +1. Red, NT+1T; cyan, NT+1C; blue, NT+1A; Green, NT+1G; white, indole ring of W183 in β-subunit of RNAP. D. structures quality.

1.3. Material and methods

Plasmids. Plasmid used for protein purifications were pRL706 and pRL706 derivatives constructed using site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit; Agilent, Inc.).

***E. coli* RNAP.** For fluorescence-detected RNAP-DNA interaction assays, *E. coli* wild-type RNAP core and mutants-RNAP core were prepared from E. coli strain XE54 (Tang, Severinov et al., 1994) that were transformed with either plasmids pRL706 or pRL706 derivatives. Protein purification procedures were essentially as in (Mukhopadhyay, Mekler et al., 2003b). The purified proteins included RNAP-βWT, RNAP-βD446A, RNAP-βD446F, RNAP-βD446W, RNAP-βD4451A, RNAP-βD151A, RNAP-βW183A and RNAP-βR371A. They were then used to form RNAP holoenzyme containing σ70 labelled at position 211 with 5-tetramethylrhodamine (as described below).

TMR211-σ70. *E. coli* σ70 derivative containing a single Cys residue (Cys211) was prepared as in (Mukhopadhyay et al., 2003b), but using plasmid pGEMD(-Cys)211C (Mekler, Kortkhonjia et al., 2002). Yields were 4-60 mg/l, and purities were > 95%. A σ70 derivative labeled at position 211 with tetramethylrhodamine (TMR211-σ70) was prepared as in (Mekler, Minakhin et al., 2011a, Zhang et al., 2012).

[TMR211-σ70]-RNAP. For fluorescence-detected RNAP-DNA interaction assays [TMR211-σ70]-RNAP-β WT holoenzymes, and any one of the [TMR211-σ70]-RNAP-β derivatives holoenzymes, were prepared as described (Zhang et al., 2012). RNAP core (3μM) and TMR211-σ70 (12 μM) were incubated in 1.5 ml 20 mM Tris-HCl (pH 7.9), 200 mM NaCl, 0.1 mM EDTA, 5 mM 2-mercaptoethanol, and 5% glycerol (running buffer) for

2 h at 4°C. This reaction mixture was applied to a HiLoad 16/60 Superdex S200 column (GE Healthcare) that was equilibrated with running buffer, and the column was eluted with 180 ml of the same buffer. Fractions containing RNAP holoenzyme were pooled, concentrated to ~3 mg/ml using 30 kDa MWCO Amicon Ultra-15 centrifugal ultrafilters (EMD Millipore), and stored at -80°C in a solution of 10 mM Tris-HCl (pH 7.9), 100 mM NaCl, 0.05 mM EDTA, 2.5 mM 2-mercaptoethanol, and 50% glycerol.

Oligonucleotides. Oligodeoxyribonucleotides (purchased from IDT) were dissolved in nuclease free water (Ambion) to 3 mM and stored at -80°C.

Nucleic-acid scaffolds. Nucleic-acid scaffolds for fluorescence-detected RNAP-DNA interaction assays and crystallization were prepared essentially as in (Zhang et al., 2012). Nontemplate-strand oligonucleotide (0.5 mM) and template-strand oligonucleotide (0.55 mM) were combined in a solution of 25 μl 5 mM Tris-HCl (pH 7.7), 200 mM NaCl, and 10 mM MgCl2, heated 5min at 95 °C, then cooled to 25 °C in 2 °C increments of 60 s per step in a thermal cycler (Applied Biosystems) and were then stored at -80 °C, for long storage time, or stored at -20 °C for shorter storage time. Sequences of all scaffolds prepared in this work are presented in Figs 1-10.

RNAP-DNA equilibrium binding assays. Equilibrium binding experiments were performed using a σ70-molecular-beacon assay as described in (Mekler et al., 2011a, Mekler, Pavlova et al., 2011b) in a 96-well microplate format (Zhang et al., 2012). Reaction mixtures of [TMR211-σ70]-RNAP at 1 nM and nucleic-acid scaffold at 0, 5, 10, 20, 40, 160, 320, 640, 1,280, 2,560, or 5,120 nM in 100 ul reaction buffer (20 mM Tris-HCl (pH 8.0), 1 M NaCl, 10 mM MgCl2, 1 mM DTT, 0.02% Tween-20, and 5% glycerol) were incubated 2 h at 25°C . Fluorescence emission intensities were measured using a microplate

reader (GENios Pro; TECAN, Inc; excitation wavelength = 550 nM; emission wavelength = 590 nM). Equilibrium dissociation constants (Kd), were determined by non-linear regression using the equation:

$F/F_o = \{[S]/(Kd + [S])\} + 1$

where F is the emission intensity at a given concentration of the scaffold, F_o is the fluorescence emission intensity in the absence of the scaffold, and [S] is the concentration of the scaffold.

RNAP-DNA dissociation kinetic assays. High-salt-induced dissociation experiments were performed using a σ70-molecular-beacon assay (Mekler et al., 2011a, Mekler et al., 2011b) in a 96-well microplate format (Zhang et al., 2012). Reaction mixtures containing [TMR211-σ70]-RNAP and 16 nM nucleic-acid scaffold in 75 ul 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM MgCl2, 1 mM DTT, 0.02% Tween-20, and 5% glycerol and were incubated 2 h at 25°C to allow formation of RNAP-scaffold complexes. To induce dissociation of RNAP-scaffold complexes, 25 µl of the same buffer yet including 4 M NaCl was injected to the reaction mixture. Fluorescence emission intensities were then immediately measured continuously for 2 h using a microplate reader (GENios Pro; TECAN, Inc; excitation wavelength = 550 nM; emission wavelength = 590 nM). Dissociation rate constants, k_{off}, were determined by nonlinear regression using the equation:

$F/F_o = A + B[\exp(k_{off}t)] + 1$

Where F is the emission intensity at a given time, F_o is the fluorescence emission intensity at time 0, t is time after NaCl injection, and A and B are unconstrained constants.

***T. thermophilus* RNAP holoenzyme.** *T. thermophilus* RNAP holoenzyme was prepared as described previously (Zhang et al., 2012).

Assembly of transcription initiation complexes. Nucleic-scaffolds were prepared as described in (Zhang et al., 2012) with synthetic oligodeoxynucleotides (Figure S??). Complexes for crystallization were prepared by incubating 20 µl 18 µM T. thermophilus holoenzyme (in 20 mM Tris-HCl, pH 7.7, 100 mM NaCl, and 1% glycerol), 1 µl 0.5 mM of the nucleic-acid scaffold (in 5 mM Tris-HCl, pH 7.7, 200 mM NaCl, and 10 mM MgCl2), and 1 µl 25 mM ribodinucleotide GpA for1 h at 25 °C.

Structure determination. Crystals of *T. thermophilus* RPo complexes were grown, optimized, and handled essentially as in (Zhang et al., 2012). By using vapor-diffusion hanging-drop technique, 1 µl RPo (in 20 mM Tris-HCl, pH 7.7, 100 mM NaCl, and 1% glycerol) was mixed with 1 µl reaction buffer (100 mM Tris-HCl, pH 8.4, 200 mM KCl, 50 mM MgCl2, and 9.5% PEG4000), and equilibrated against 400 µl reaction buffer. The rod-like small crystals that appeared in 1-2 days were used as micro-seeding seeds for crystal-size optimization. Crystals larger than 0.2mm X 0.2mm X 0.1 mm were harvested by transferring into RB containing 17.5% (v/v) (2R, 3R)-(-)-2,3-butanediol (Sigma-Aldrich) and were flash-cooled with liquid nitrogen.

Diffraction data were collected at the Cornell High Energy Synchrotron Source (Chessin & Summers, 1970) beamline F1, then processed and scaled using HKL2000 (Otwinowski & Minor, 1997). The phases were determined by molecular replacement using Molrep (Vagin & Teplyakov, 1997) in CCP4 program suite (Winn, Ballard et al., 2011) with one RNAP molecule from the structure of *T. thermophilus* RPo (PDB 4G7H; (Zhang, 2012 #2)) as the search model. The reciprocal refinements were performed with Phenix (Adams,

Afonine et al., 2010) and the iterative model building was performed with Coot (Emsley, Lohkamp et al., 2010). The final crystallographic models will be deposited in the PDB.

Chapter 2

RNAP-CRE interaction promote promoter unwinding

Hanif Vahedian Movahed and Richard H. Ebright

2.1. Background

The interactions between RNAP core and CRE position +2 (+2G) is reminiscent of the mode of interaction of sigma subunit with two bases of the -10 element (A-11 and T-7) and one base of the discriminator element (G-6) Fig 1. Sigma subunit unstacks these three DNA bases, flips them out of base stacks and inserts them into pockets formed by specific residues of sigma. This mode of interaction provides an effective means of sequence recognition, because it enables contact with essentially all atoms of the base. Additionally, this mode of interaction provide an effective means to use binding energy to drive DNA unwinding, because this interaction occurs only when DNA is unwound. This mechanism of flipping out a base and insertion into a pocket is responsible for sigma ability to promote DNA unwinding. As discussed in general introduction and chapter one, RNAP core also flips out a base and inserts it into a β-pocket in analogues manner to sigma. As in -11, -7 and -6 bases interactions with sigma, the CRE +2 interactions with core also require an unwound transcription bubble. Therefore, we propose that RNAP-CRE interactions are expected to play functional roles in promoter unwinding by stabilizing the unwound state.

2.2. Results and discussions

The equilibrium binding data presented in chapter 1 show that having a consensus CRE sequence at position +2, +1 or -4 reduces the equilibrium dissociation constant (Kd) of the RNAP-DNA open complex by a factor of 5, 20 and 5, respectively. Additionally, the dissociation kinetic data presented in chapter 1 show that having a consensus CRE sequence at position +2, +1 or -4 reduces the dissociation rate constant (off-rate) of the RNAP-DNA open complex by a factor of 5, 32 and 13, respectively. This data indicate that RNAP-DNA open complexes where DNA contain consensus CRE sequences have higher stability compared to when the DNA sequences contain non-consensus CRE sequences on a pre-formed-unwound-scaffold DNA. We conclude that RNAP-DNA interaction stabilize the open complex formation.

To further confirm our previous findings, I sought to monitor the open complex formation level of a duplex promoter DNA (rather than a pre-formed-unwound-scaffold DNA), using a previously developed real-time fluorescent promoter melting assay (Ko & Heyduk, 2014). In this method, a Cy3 probe is incorporated at position +2 in dsDNA and have low fluorescence but upon single strand DNA formation the fluorescence increases. I modified this assay by incorporating the Cy3 fluorescence probe to the non-template strand at position -1 instead of the positon +2 to allow base substitution at position +2. Upon mixing of RNAP with DNA, RNAP unwinds DNA surrounding the transcription start-site (from regions -11 toward +2) and therefore the Cy3 fluorescence enhances until it reaches a plateau over time. Upon addition of the NTP, RNAP translocate and the previously unwound single stranded DNA region will rewound to form double stranded DNA decreasing the Cy3 fluorescence intensity. By comparing the fluorescence level at plateau

before and after NTP addition, we would be able to estimate the differences in unwound state between a WT-RNAP and [Ala446]β-RNAP defective in making RNAP-CRE interactions. I prepared N25 promoter derivative containing a consensus base G or a non-consensus base T at position +2 and monitored the Cy3 fluorescence upon addition of WT-RNAP and [Ala446]β-RNAP.

I performed assays with RNAP-βWT or RNAP-βD446A on templates containing a Cy3 linked to the base at position -1 and contain G or T at position +2 (Fig. 1B). I found that when template contain +2T, the RNAP-βD446A and RNAP-βWT exhibited fluorescence enhancement and plateaued at similar levels, whereas when the template contain +2G, the RNAP-βD446A plateaued ~70% of the maximum level reached by RNAP-βWT. The results indicate that, on templates containing +2G, a greater proportion of RPos with RNAP-βWT are in fully unwound states compared to RPos with RNAP-βD446A. While on templates containing +2T, similar proportion of RPos are formed with both RNAP-βWT and -βD446A. I propose two possible scenarios where this 30% difference in fluorescence enhancement could be originated from: First scenario, we can imagine that among the RPo populations, at a time, only ~30% of the complexes are able to form fully unwound RPos with unwound -1 position. Second scenario, entire RPo populations are in unwound states but because of the less stable unwound state at position +2, the conformation of the Cy3 at position -1 is not at optimum to produce maximum quantum yield, therefore all molecules fluorescent but each with only 70% of the maximum signal. This could be inferred from the pattern of the fluorescent reduction upon NTP addition. As can be seen in Fig 1, upon addition of the NTP, when template contain +2T, the RNAP-βD446A exhibited similar fluorescence reduction starting from similar levels and declining to almost similar levels

as with RNAP-βWT. However, when template contain +2G, the RNAP-βD446A, show a fluorescent enhancement-phase where the signal abruptly increase to reach a new plateau at level similar to RNAP-βWT before NTP addition and only then the fluorescent signal start to decline; whereas with RNAP-βWT signal start to decline form where it was plateaued before NTP addition. This data indicate that the second scenario is more possible, where upon NTP addition to position i and i+1 sites the DNA is fully unwound for all the molecules at once, causing an immediate enhancement in fluorescent, followed by until translocation where the DNA at position -1 is re-wound and fluorescent start to decline. If the first scenario was the case, only 30% of the molecules showed required unwinding before translocation while 70% where already starting to translocate and show fluorescent decline. We conclude that sequence specific RNAP-G_{CRE} interactions promote DNA unwinding in transcription initiation and stabilize the open complex, which confirm our previous finding from equilibrium binding and dissociation kinetic experiments.

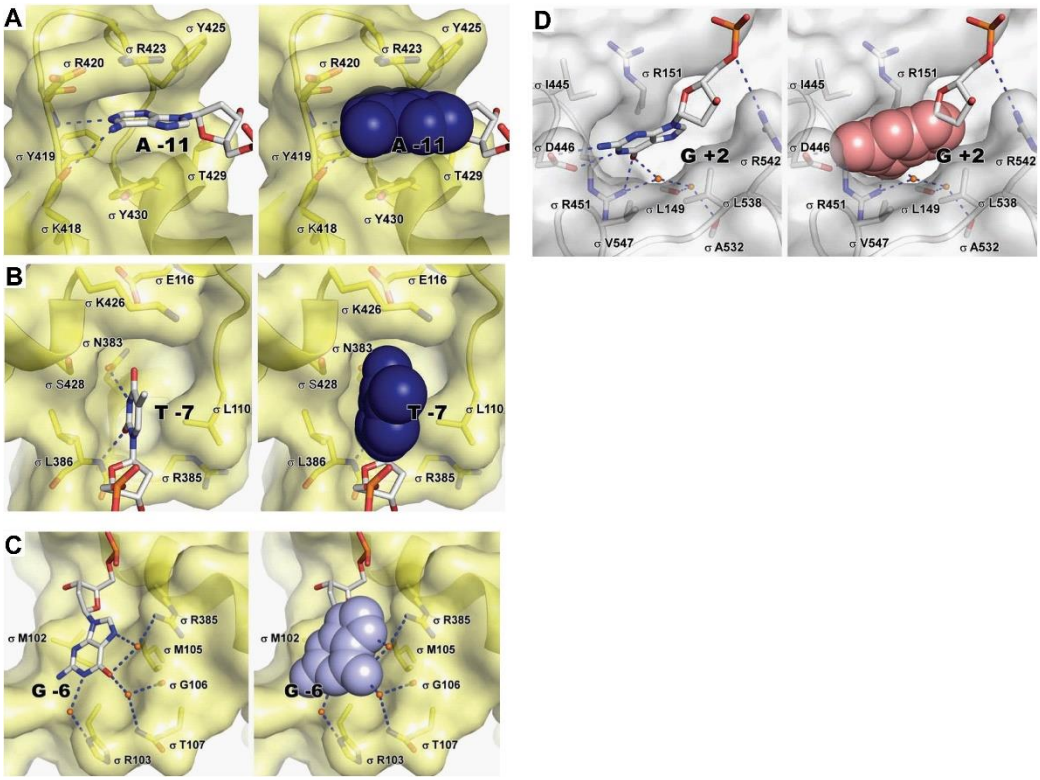


Figure 1. Comparison of the mode of interactions in sigma and core with promoter DNA. (A-C) sigma flips out bases at positions -11, -7 and -6 and inserts into pockets. (D) RNAP core, in analogous manner to sigma, flips out base +2 G into a pocket formed by residues of β-subunit.

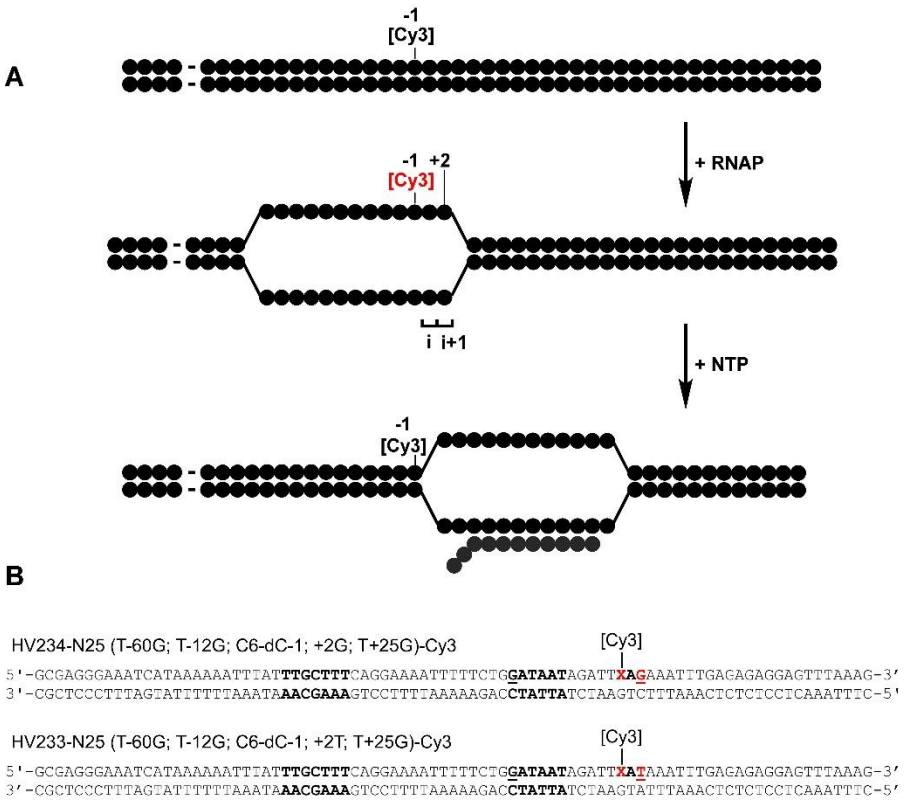


Figure 2. Real-time fluorescent promoter melting assay. (A) Assay design, the cy3 dye is incorporated at position -1 with an internal amino modifier C6-dC. The fluorescent is quenched when DNA is double-stranded and increases upon strand separation during open complex formation. Therefore, the CY3 dye report the closed or open state of bp at position -1, which reflect open complex formation. Open NTP addition and RNAP translocation, the base at position -1 rewind and fluorescent decreases. (B) Fluorescently labelled templates used in this study. Both template include a cy3 attached to base C-1. One template contain a consensus G_{CRE} at position +2 and the other contain a non-consensus T at position +2.

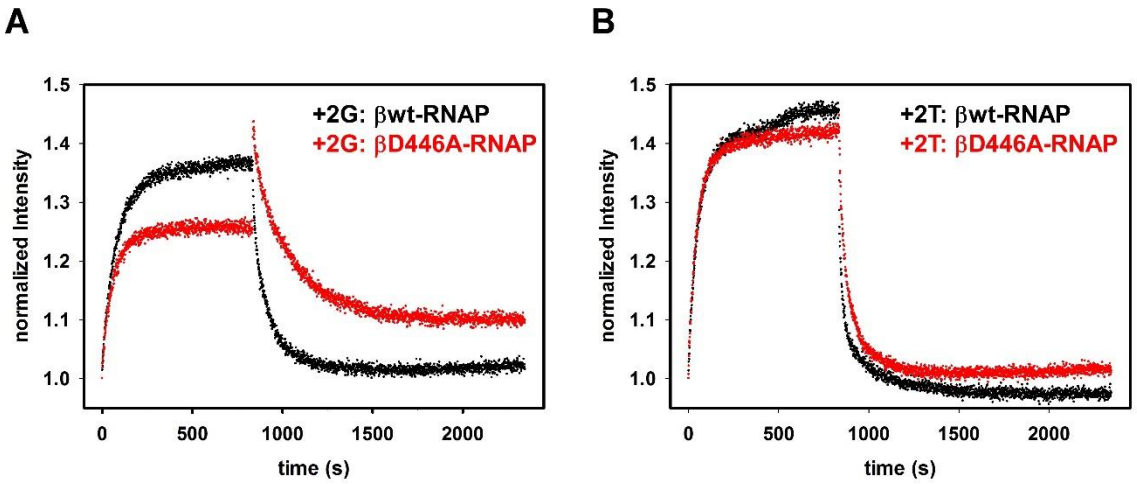


Figure 2. Effect of RNAP-CRE interactions on promoter melting. Kinetic traces for promoter melting and escape for template containing a consensus G_{CRE} at position +2 (A) and the other contain a non-consensus T at position +2 (B).

2.3. Materials and Methods,

Proteins. *E. coli* wild-type RNAP holo and mutants-RNAP holo were prepared from E. coli strain XE54 (Tang et al., 1994) that were transformed with either plasmids pRL706 or pRL706-βD446A derivatives. Protein purification procedures were essentially as in (Mukhopadhyay et al., 2003b). The purified proteins included RNAP-βWT and RNAP-βD446A.

Oligos and reagents. Cy3 NHS ester was purchased from GE Healthcare. The oligonucleotides were purchased from IDT (table X). The non-template strand were ordered to be synthesis with a single internal amino modifier C6 dT at the +2, -2 or -3 positions; for position -1 the internal amiono modifier C6 dC was used. Oligos were solubilized in nuclease free water.

oligo labeling and template preparations. Non-template strands oligo with internal amino-modifier were labeled with Cy3-NHS-esther (per the procedure of the manufacturer), ethanol precipitated to remove excess dye. Then the samples were further purified using HPLC (C18 column) to purify the labeled-oligo from non-labelled-oligos and any remaining free Cy3 dyes. To from duplex DNA template, the Cy3-labelled nontemplate-strand oligonucleotide (1 uM) and template-strand oligonucleotide (1 uM) were combined in 50 µl 5 mM Tris-HCl (pH 7.7), 200 mM NaCl, and 10 mM MgCl2, heated 2min at 95°C, and cooled to 25°C in 2°C steps with 60 s per step in a thermal cycler (Applied Biosystems).

Real-time promoter melting assay. The promoter melting assay were performed as in (Ko & Heyduk, 2014). The assays monitors the change in fluorescence emission of Cy3

attached to a base in DNA region which undergoes DNA unwinding and rewinding, where Cy3 fluorescent enhancement report promoter unwinding (ssDNA formation) and fluorescence decrease report promoter rewinding (dsDNA formation). Fluorescence measurements were performed using a QuantaMaster QM1 spectrofluorometer (PTI, Edison, NJ) (excitation wavelength = 550 nm; emission wavelength = 570 nm; and excitation and emission slit widths = 5 nm). For determination of promoter melting kinetics, 120 μl 5 nM Cy3-labelled DNA template in 50 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 150 mM KCl and 50μg/ml BSA and 5% glycerol was mixed with 75 μl 200 nM RNAP, in the same buffer to a final concetration of 75 nM, at 24°C in a cuvette chamber with a mixing dead time ∼5 s, and fluorescence emission intensities were monitored for 15 min at 24°C. To measure promoter escape kinetics, 5 ul of a pre-mixed solution of NTP and heparin was injected to the reaction above to a final concentration of 100 μM and 200 ug/ml, respectively, in the same cuvette chamber with a mixing dead time ∼5 s, and fluorescence emission intensities were monitored for additonal 30 min at 24°C.

Chapter 3

RNAP-CRE interaction is a determinant of transcription start-site (TSS) selection

[1]Hanif Vahedian Movahed*, Irina O. Vvedenskayaa*, Yuanchao Zhang, Deanne M. Taylor, Bryce E. Nickels and Richard H. Ebright

3.1. Background

In RPo, the second-most downstream nucleotide of the transcription bubble template strand occupies the active center i site and serves as the transcription start site (TSS), and the downstream-most nucleotide of the transcription bubble template strand occupies the active center i+1 site. The position of the TSS relative to the position of the promoter −10 element is variable (Aoyama & Takanami, 1985, Jeong & Kang, 1994, Lewis & Adhya, 2004, Liu & Turnbough, 1994, Sorensen, Baker et al., 1993, Vvedenskaya, 2015, Walker & Osuna, 2002). TSS selection preferentially occurs at the position 7-bp downstream of the promoter −10 element, but can occur over a range of at least five positions, encompassing the positions 6-, 7-, 8-, 9-, or 10-bp downstream of the promoter −10 element. Thus, there must be flexibility in the structure of RPo that enables the position of

[1] I performed all the primer extension experiment, generated reagents (proteins and plasmid constructs) to assist Irina Vvedenskayaa in MASTER and mNET-Seq experiments. Irina Vvedenskayaa prepared the sequencing libraries. MASTER DNA libraries were from previous published work (Vvedenskaya, 2015). Yuanchao Zhang from Deanne Taylor's lab performed data analysis for MASTER and mNET-Seq data. Parts of this chapter appears as published:

*Vvedenskaya IO, *Vahedian-Movahed H, Zhang Y, Taylor DM, Ebright RH, Nickels BE (2016) Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. PNAS, 113: E2899-90.

the TSS to vary relative to the position of the −10 element. We previously have proposed that variability in TSS selection is mediated by variability in the size of the unwound transcription bubble (Fig. S1A) (Robb, Cordes et al., 2013, Vvedenskaya, 2015, Winkelman, Vvedenskaya et al., 2016b). According to this model, RPo generally contains a 13-bp unwound transcription bubble that places the template-strand nucleotide 7-bp downstream of the −10 element in the i site and places the templatestrand nucleotide 8-bp downstream of the −10 element in the i+1 site (Fig. 1A and Fig. S1A) (TSS = 7). For TSS selection to occur unwound, the unwound DNA is pulled into and past the RNAP active center, and the unwound DNA is accommodated as single stranded DNA bulges within the transcription bubble, yielding a "scrunched" complex (Fig. S1A) (TSS = 8 and TSS = 9). For TSS selection to occur at positions further upstream, the opposite occurs: downstream DNA is rewound, downstream DNA is extruded from the RNAP active center, and the extrusion of DNA from the RNAP active center is accommodated by stretching DNA within the transcription bubble, yielding an "antiscrunched" complex (Fig. S1A) (TSS = 6). According to this model, any protein–DNA or protein–protein interaction that affects the energy landscape for transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo potentially could modulate TSS selection (Winkelman, Chandrangsu et al., 2016a, Winkelman et al., 2016b)

As discussed, in the structure of RPo, the RNAP core makes direct protein–DNA interactions with the non–template-strand DNA segment at the downstream part of the transcription bubble (Zhang et al., 2012); "core recognition element" (CRE; Fig. 1A) (Zhang et al., 2012). RNAP–CRE interactions with the non–template-strand nucleotide at the extreme downstream end of the transcription bubble (i.e., TSS+1NT) are sequence

specific, with preference for the base G (G_{CRE}) (Fig. 1, red G) see chapter 1 and (Zhang et al., 2012). As it was shown in previous sections, sequence-specific RNAP–G_{CRE} interactions facilitate promoter unwinding to form the transcription bubble, stabilize the unwound transcription bubble, and define the downstream end of the transcription bubble, see chapter 1 and (Zhang et al., 2012). According to this proposal, sequence-specific RNAP–G_{CRE} interactions should affect the energy landscape for transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo and therefore potentially could affect TSS selection (Fig. S1B). Here we tested the proposal that sequence-specific RNAP–G_{CRE} interactions affect TSS selection. To do this, we used high-throughput sequencing–based approaches to compare TSS selection by WT RNAP to TSS selection by a mutant RNAP defective in sequence-specific RNAP–G_{CRE} interactions. Our results demonstrate that sequence-specific RNAP–CRE interactions are a determinant of TSS selection.

3.2. Results and discussions

3.2.1. Sequence-Specific RNAP–CRE Interactions Are a Determinant of TSS Selection in Vitro. In crystal structures of RNAP–promoter open complexes, residue D446 of the RNAP β subunit makes direct H-bonded interactions with Watson–Crick H-bond–forming atoms of G at G_{CRE} (Zhang et al., 2012). The interactions by βD446 determine specificity at G_{CRE}. Thus, substitution of βD446 by alanine eliminates the ability of RNAP to distinguish A, G, C, and T at the G_{CRE} position see chapter 1. Accordingly, an RNAP derivative carrying the βD446A substitution can serve as a reagent to assess the functional significance of sequence-specific RNAP-G_{CRE} interactions (Fig. 1B, Lower Left). To define the contribution of sequence-specific RNAP–G_{CRE} interactions to TSS selection, we

used a high-throughput sequencing–based methodology termed massively systematic transcript end readout (MASTER) (Vvedenskaya, 2015). MASTER entails the construction of a template library that contains up to 4^{10} (~1,000,000) bar-coded sequences, production of RNA transcripts from the template library in vitro or in vivo, and analysis of transcript ends using highthroughput sequencing (Vvedenskaya, 2015, Winkelman et al., 2016b).

To analyze the effect of disrupting sequence-specific RNAP–GCRE interactions on TSS selection, we used a MASTER template library, lacCONS-N7, that contained 47 (~16,000) sequence variants at positions 4–10 bp downstream of the −10 element of a consensus Escherichia coli σ70-dependent promoter (Fig. 1B, Upper) (Vvedenskaya, 2015). We performed in vitro transcription experiments with the lacCONS-N7 template library, using, in parallel, WT RNAP (RNAP-βWT) or the RNAP derivative containing the βD446A substitution (RNAP-βD446A). RNA products generated in the transcription reactions were isolated and analyzed using high-throughput sequencing of RNA barcodes and 5′ ends (5′ RNAseq) to define, for each RNA product, the template that produced the RNA and the TSS position (Fig. 1B, Lower Right). For each sequence variant, we calculated the percentage of reads starting at each position within the randomized TSS region, %TSSY = 100 × (no. reads starting at position Y/total no. reads starting at positions 4–10).

To determine the effect of disrupting RNAP–G_{CRE} interactions on TSS selection, we considered TSS positions where TSS+1NT is included within the randomized region of the MASTER template library (i.e., TSS positions 6, 7, 8, and 9). We first calculated %TSS values for each of these positions on the basis of the identity of TSS+1NT. Thus, for each TSS position, we averaged the %TSS values for the ~4,000 templates having A at

TSS+1NT, the ~4,000 templates having C at TSS+1NT, the ~4,000 templates having G at TSS+1NT, and the ~4,000 templates having T at TSS+1NT. Next, we calculated the difference in these %TSS values for reactions performed with RNAP-βWT vs. reactions performed with RNAP- βD446A. We observed that, for all four tested TSS positions (positions 6, 7, 8, and 9), the βD446A substitution decreased the %TSS when TSS+1NT was G (1.3–7.3% decreases; Fig. 2A, top row of table). In contrast, for three of the four tested TSS positions (positions 6, 7, and 8), the βD446A substitution did not decrease the %TSS when TSS+1NT was A, C, or T, and, for the fourth position (position 9), the βD446A substitution did not decrease the %TSS, or decreased the %TSS by smaller amounts, when TSS+1NT was A, C, or T (Fig. 2A, bottom three rows of table).

We identified 1,230 TSS positions (5.6% of the 21,872 above threshold TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element) that exhibited large, ≥20%, reductions in %TSS in reactions performed with RNAP-βD446A vs. reactions performed with RNAP-βWT. For these 1,230 TSS positions with large, ≥20%, CRE effects, ~90% contained G at TSS+1NT (Fig.2B, top row, Right), whereas, for the total sample of 21,872 TSS positions, there were no detectable sequence preferences at position TSS+1NT (Fig. 2B, top row, Left). Enrichment of G at TSS+1NT for TSS position with large, ≥20%, CRE effects was observed for TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element (TSS = 6, 7, 8, or 9) (Fig. 2B, bottom four rows). In summary, the overwhelming majority of TSS positions that exhibit large, ≥20%, CRE effects have G at TSS+1NT.

For each of these two TSS region sequences, we prepared templates containing G, A, C, or T at position TSS+1NT, performed in vitro transcription experiments with RNAP-βWT or RNAP-βD446A, and analyzed RNA products by primer extension. For each of the

two sets of constructs, the primer-extension results matched the MASTER results. A large, ~30%, CRE effect was observed when TSS+1NT was G but not when TSS+1NT was A, C, or T (Fig. 3). The results in Figs. 2 and 3 establish that disrupting sequence- specific RNAP–GCRE interactions affects TSS selection in vitro in a manner that correlates with the presence and position of GCRE in the TSS region. We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in vitro.

3.2.2. Sequence-specific RNAP-CRE interactions are a determinant of TSS selection in vivo.

Analysis of 47 (~16,000) consensus promoter derivatives. To define the contribution of sequence-specific RNAP–GCRE interactions to TSS selection in vivo, we used merodiploid native-elongating transcript sequencing (mNET-seq) (Vvedenskaya, Vahedian-Movahed et al., 2014). mNET-seq involves selective analysis of transcripts associated with an epitope-tagged RNAP in the presence of a mixed population of epitope-tagged RNAP and untagged RNAP (Fig. 4A). In prior work, we used mNET-seq to determine the effect of sequence-specific RNAP–GCRE interactions on pausing during elongation (Vvedenskaya et al., 2014). In this work, we used a variant of mNET-seq, 5′ mNET-seq, to determine the effect of sequence-specific RNAP–GCRE interactions on TSS selection (Fig. 4A). To do this, we introduced into cells a plasmid encoding 3xFLAG-tagged βWT or 3× FLAG-tagged βD446A, isolated RNA products associated with RNAP-βWT or RNAP-βD446A by immunoprecipitation, converted RNA 5′ ends to cDNAs, and performed high-throughput sequencing (Fig. 4A). To enable direct comparison of in vivo and in vitro results, we performed 5′ mNET-seq using the same MASTER template library of 47 (~16,000) consensus core promoter derivatives that we used for in vitro analysis.

The results of MASTER in vivo (Fig. 4 B and C) matched the results of MASTER in vitro (Fig. 2). For all four tested TSS positions (positions 6, 7, 8, and 9), the βD446A substitution decreased the %TSS when TSS+1NT was G (0.6–7.3% decreases) (Fig. 4B, top row of table). In contrast, for three of the four tested TSS positions (positions 6, 7, and 8), the βD446A substitution did not decrease the %TSS when TSS+1NT was A, C, or T, and, for the fourth position (position 9), the when TSS+1NT was A, C, or T (Fig. 4B, bottom three rows of table). Furthermore, we identified 860 TSS positions (4.3% of the 20,217 above-threshold TSS positions located 6-, 7-, 8-, or 9-bp downstream of the −10 element) with large, ≥20%, CRE effects. For these 860 TSS positions with large, ≥20%, CRE effects, ∼80% contained G at TSS+1NT (Fig. 4C, Right), whereas, for the total sample of 20,217 TSS positions, there were no detectable sequence preferences at position TSS+1NT (Fig. 4C, Left). The results establish that disrupting sequence-specific RNAP– GCRE interactions affects TSS selection in vivo in a manner that correlates with the presence and position of GCRE in the TSS region. We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in vivo.

Analysis of E. coli transcriptome. Having shown by MASTER that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in the context of a consensus core promoter in vivo, we next assessed the contribution of sequence-specific RNAP–CRE interactions to TSS selection in the context of natural promoters in vivo in E. coli. (The primers used in the in vivo MASTER analysis by 5′ mNET-seq shown in Fig. 4 provided information only about transcripts from the synthetic consensus promoter derivatives. This is because the primers used for synthesis of the first cDNA strand annealed only to transcripts produced from the synthetic consensus promoter derivatives. A separate

experiment, with primers that enable generation of cDNAs from transcripts produced from natural E. coli promoters, was necessary to provide information about transcripts from natural E. coli promoters. Therefore, to analyze transcripts from natural E. coli promoters, the primers used for synthesis of the first cDNA strand carried nine randomized nucleotides at the 3′ end.)

Using data from experiments performed with RNAP-βWT, we identified 1,500 above-threshold TSS positions associated with natural promoters in E. coli. Of these 1,500 TSS positions, we identified 44 TSS positions that exhibited large, ≥20%, CRE effects (Table S1); 39 of these 44 (∼90%) contained G at TSS+1NT (Fig. 5B, Right, and Table S1), whereas for the total sample of 1,500 above-threshold TSS, there were no detectable sequence preferences at TSS+1NT (Fig. 5B, Left).

To validate the 5′ mNET-seq results, we performed primer extension experiments with two E. coli promoters that contained a TSS that exhibited a large, ≥20%, CRE effect and contained G at TSS+1NT: PsecE and PhemC (Table S1). We generated linear templates carrying PsecE or PhemC, performed in vitro transcription assays using RNAP-βWT or RNAP-βD446A, and analyzed TSS selection by primer extension (Fig. 5C). For each promoter, two prominent start sites were observed in reactions with RNAP-βWT. In the case of PsecE, ∼60% of the transcripts started at an A located 7-bp downstream of the predicted −10 element (A7) and ∼40% of the transcripts started at a G located

8-bp downstream (G8) (Fig. 5C, Left). In the case of PhemC, ∼30% of the transcripts started at an A located 6-bp downstream of the predicted −10 element (A6) and ∼70% of the transcripts started at a G located 8-bp downstream (G8) (Fig. 5C, Right).

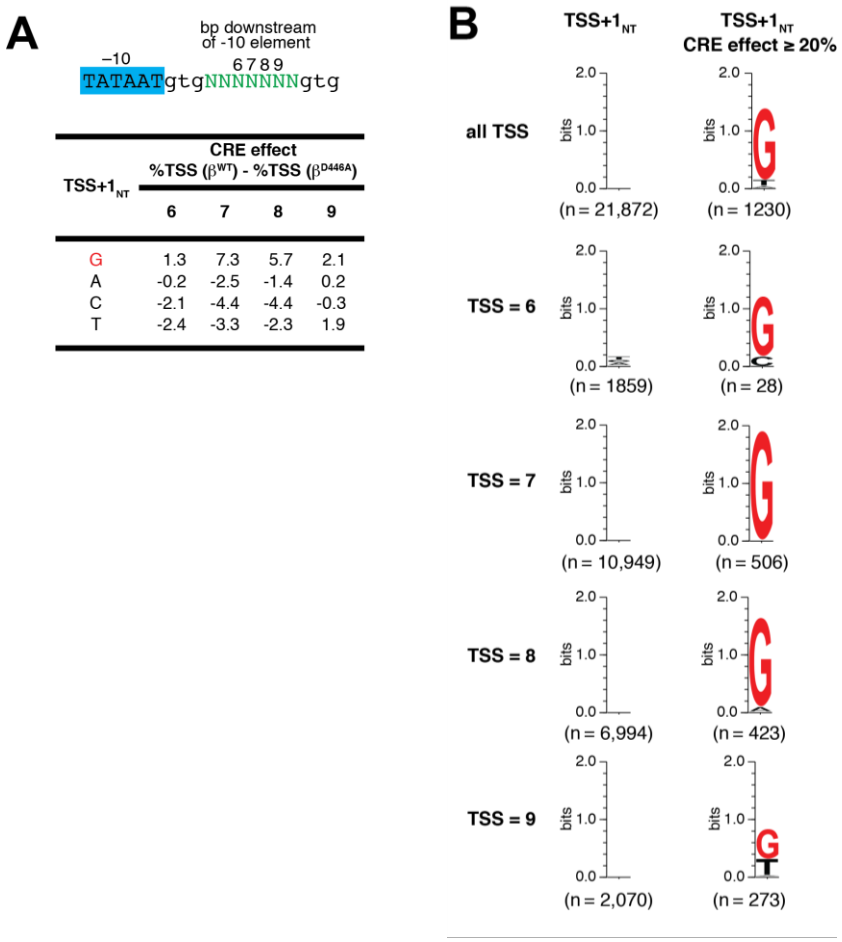
For each promoter, the percentage of transcripts starting at the position that contained G at TSS+1NT (A7 for PsecE and G8 for PhemC) was reduced by ∼30% when reactions were performed with RNAP-βD446A (Fig. 5C), consistent with results of 5′ mNETseq (Table S1). We conclude that sequence-specific RNAP–CRE interactions are a determinant of TSS selection in natural promoters in the E. coli genome.

Summary of sequence-Specific RNAP–CRE Interactions in TSS Selection. Here we show that sequence-specific interactions between RNAP and the downstream segment of the nontemplate strand of the transcription bubble (CRE) are a determinant of TSS selection. In particular, using high-throughput sequencing–based approaches, we define a role of sequence-specific recognition of a G at the most downstream position of the CRE (GCRE) during TSS selection in the context of a library of 47 (∼16,000) TSS region sequences of a consensus core promoter in vitro and in vivo (Figs. 2–4) and in the context of natural promoters in E. coli in vivo (Fig. 5 and Table S1). As discussed above, variability in TSS selection is believed to involve transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo (Fig. S1A) (Robb et al., 2013, Vvedenskaya, 2015, Winkelman et al., 2016a, Winkelman et al., 2016b). We propose that the observed effects of sequence-specific RNAP–CRE interactions on TSS selection occur by influencing transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo (Fig. S1B). Specifically, we propose that sequence-specific RNAP–CRE interactions favor TSS selection at sequences that contain G at TSS+1NT. According to this proposal, the role of sequence-specific RNAP–CRE interactions in defining the downstream edge of the transcription bubble concurrently defines the extent of transcription bubble expansion or contraction (scrunching or antiscrunching) in RPo and therefore modulates TSS selection

(Fig. S1B). The results of this work, together with results of previous work, establish that TSS selection involves at least four promoter sequence determinants: (i) position relative to the −10 element (preference for the position 7-bp downstream of the −10 element) (Aoyama & Takanami, 1985, Jeong & Kang, 1994, Lewis & Adhya, 2004, Liu & Turnbough, 1994, Sorensen et al., 1993, Vvedenskaya, 2015, Walker & Osuna, 2002); (ii) sequence of TSST and TSS-1T (strong preference for pyrimidine at TSST and preference for purine at TSS- 1T, which enable initiation with a purine NTP and maximize stacking between DNA bases and the initiating purine NTP) (Hawley & McClure, 1983, Jorgensen, Buch et al., 1969, Maitra & Hurwitz, 1965, Shultzaberger, Chen et al., 2007, Vvedenskaya, 2015); (iii) sequence of the discriminator element (preference for TSS selection at upstream positions for discriminator sequences that disfavor scrunching and preference for TSS selection at downstream positions for discriminator sequences that favor scrunching) (Winkelman et al., 2016a, Winkelman et al., 2016b); and (iv) sequence of the CRE (preference for G at TSS+1NT). In addition to these sequence determinants, DNA topology and NTP concentrations also influence TSS selection (Liu & Turnbough, 1994, Qi & Turnbough, 1995, Sorensen et al., 1993, Tu & Turnbough, 1997, Turnbough, 2008, Turnbough & Switzer, 2008, Vvedenskaya, 2015, Walker, Mallik et al., 2004, Walker & Osuna, 2002, Wilson, Archer et al., 1992). Thus, TSS selection is a multifactorial process, in which the ultimate outcome for a given promoter reflects the contributions of multiple promoter sequence determinants and multiple reaction conditions. Because sequence-specific RNAP–CRE interactions are only one of several determinants of TSS selection, their quantitative significance at different promoters differs. At some promoters, such as PsecE and PhemC, sequence-specific RNAP–CRE interactions have quantitatively large,

≥20%, effects on TSS selection (Fig. 5C and Table S1), whereas at other promoters, the quantitative effects of RNAP–CRE interactions are smaller.

Prospect. In prior work, our lab showed that sequence-specific RNAP–CRE interactions affect RPo formation during transcription initiation, RPo stability during transcription initiation, translocational bias during transcription elongation, and sequence-specific pausing during transcription elongation (Vvedenskaya et al., 2014, Zhang et al., 2012). Accordingly, our findings that sequence-specific RNAP–CRE interactions are a determinant of TSS selection add to an emerging view that sequence-specific RNAP–CRE interactions play functionally important roles during all stages of transcription that involve an unwound transcription bubble. A priority for future work will be to assess the roles of sequence-specific RNAP–CRE interactions in other steps of transcription that involve an unwound transcription bubble (e.g., transcriptional slippage, initial transcription, promoter escape, factor-dependent pausing, and termination). Another priority for future work will be to assess possible roles of sequence-specific RNAP–CRE interactions in eukaryotic transcription, noting that RNAP residues involved in sequence-specific RNAP–CRE interactions are conserved in bacteria and eukaryotes.



TSS+1NT	CRE effect %TSS (βWT) - %TSS (βD446A)			
	6	7	8	9
G	1.3	7.3	5.7	2.1
A	-0.2	-2.5	-1.4	0.2
C	-2.1	-4.4	-4.4	-0.3
T	-2.4	-3.3	-2.3	1.9

Figure 1. Analysis of effects of sequence-specific RNAP-CRE interactions by massively systematic transcript end readout [MASTER; (Vvedenskaya, Zhang et al., 2015)]. A. RPo for TSS at position 7. B. Top: DNA fragment carrying the MASTER template library lacCONS-N7. Bottom right: 5′ RNA-seq analysis of RNA products generated from the MASTER-N7 template library in vitro. The sequence of the barcode is used to assign the RNA product to an N7 region and the sequence of the 5′ end is used to define the TSS. Bottom left: Structural organization of downstream end of transcription bubble in RPo for promoter containing GCRE formed with wild-type RNAP (top) or RNAP derivative carrying the βD446A substitution (bottom).

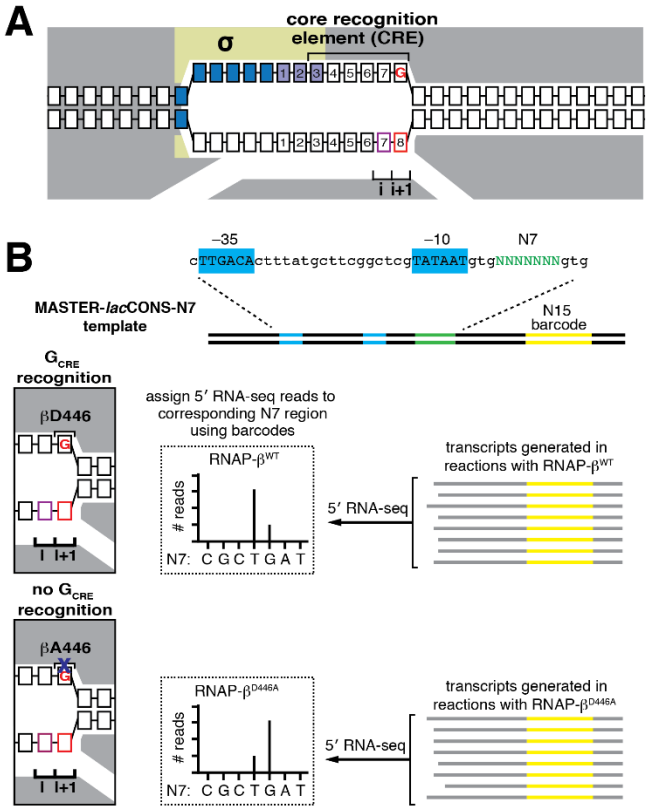
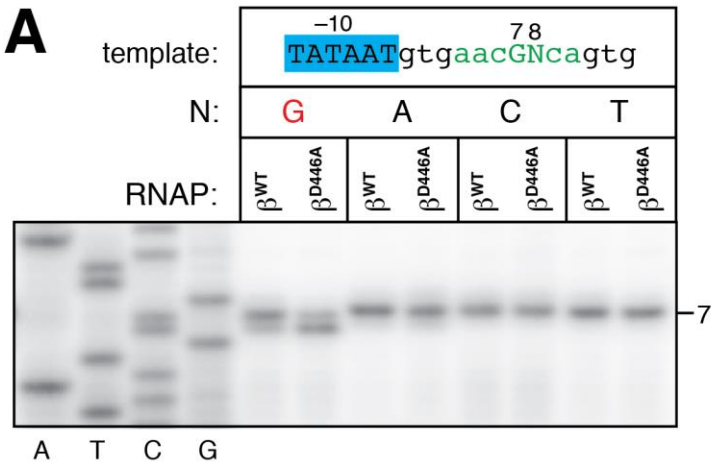


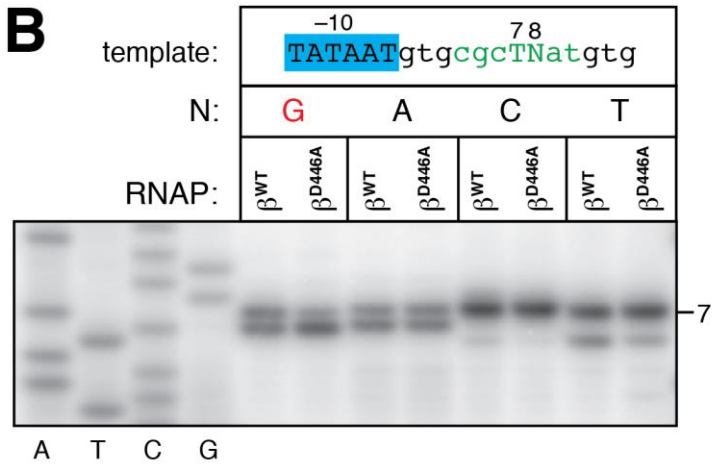
Figure 2. Effects of disrupting RNAP-GCRE interactions in vitro:analysis by MASTER

A. Effect of sequence at TSS+1NT on %TSS for RNAP-βWT vs. RNAP-βD446A.

B. Sequence preferences for TSS+1NT. Sequence logo (Crooks, Hon et al., 2004)for TSS+1NT of above-threshold TSS (left) and TSS that exhibited a large, ≥ 20%, reduction in %TSS in reactions performed with RNAP-βD446A vs. reactions performed with RNAP-βWT.



N	%TSS (β^{WT}) - %TSS (β^{D446A}) at TSS = 7	
	primer extension	MASTER
G	30	29
A	0	0
C	0	0
T	0	0



N	%TSS (β^{WT}) - %TSS (β^{D446A}) at TSS = 7	
	primer extension	MASTER
G	25	23
A	-7	-6
C	-4	-12
T	-11	-19

Figure 3. Effects of disrupting RNAP-GCRE interactions in vitro: analysis by primer extension.

Left: Primer-extension results. RNA products were generated in reactions performed with RNAP-βWT or RNAP-βD446A and placCONS templates carrying TSS-region sequences (in green) of AACGNCA (panel A) or CGCTNAT (panel B), where N is G, A, C, or T. Bands corresponding to a TSS at position 7 are indicated.

Right: Table lists the difference in %TSS (%TSS for reactions with RNAP-βWT - %TSS for reactions with RNAP-βD446A) at position 7 for templates carrying a G, A, C, or T at position 8 calculated by primer extension or calculated by MASTER.

To validate the MASTER results, we performed further analyses of two TSS region sequences that exhibited large, ≥20%, CRE effects, contained a TSS at the most common position (position 7), and contained G at TSS+1NT (position 8) (Fig. 3).

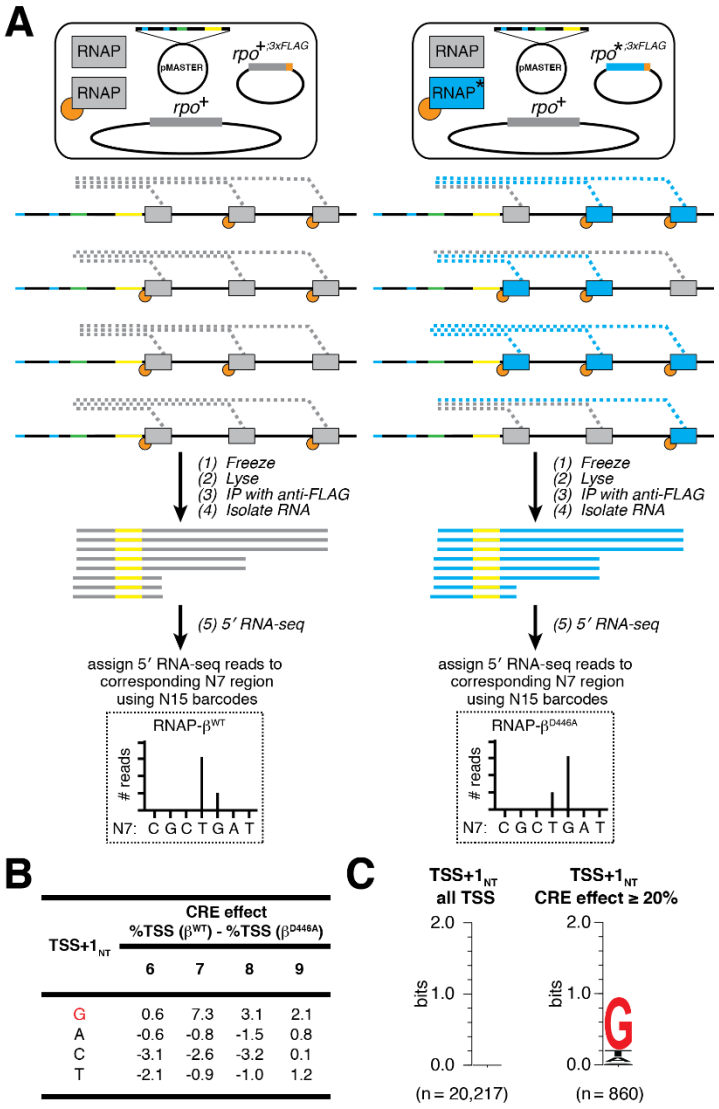


Figure 4. Figure 4. Effects of disrupting RNAP-G$_{CRE}$ interactions *in vivo*: 5′ mNET-seq analysis of 4^7 (~16,000) consensus promoter derivatives. (A) Steps in 5′ mNET-seq analysis of TSS selection from plasmid-borne MASTER template library: (top) RNAP derivatives in cells (the blue RNAP derivative with asterisk is RNAP-β^{D446A}); (middle) RNAPs on the same MASTER template in four cells (RNA products in blue are associated with RNAP-β^{D446A}); (bottom) isolation of RNA products after immunoprecipitation with anti-FLAG affinity gel and sequencing analysis of RNA 5′ ends. In this example, TSS selection at the T in the middle of the randomized TSS region is decreased with the mutant RNAP derivative. **(B)** Effect of sequence at TSS+1$_{NT}$ on %TSS for RNAP-β^{WT} vs. RNAP-β^{D446A}. Table lists the difference in %TSS (%TSS for RNAP-β^{WT} - %TSS for RNAP-β^{D446A}) at positions 6, 7, 8, or 9 for TSS-regions carrying G, A, C, or T at TSS+1$_{NT}$. **(C)** Sequence preferences for TSS+1$_{NT}$. Sequence logo (Crooks et al., 2004) for TSS+1$_{NT}$ of above-threshold TSS positions located 6-9 bp downstream of the -10 element (left) and TSS positions located 6-9 bp downstream of the -10 element that exhibited a large, ≥ 20%, reduction in %TSS in 5′ mNET-seq analysis of RNAP-β^{D446A} vs. 5′ mNET-seq analysis of RNAP-β^{WT}.

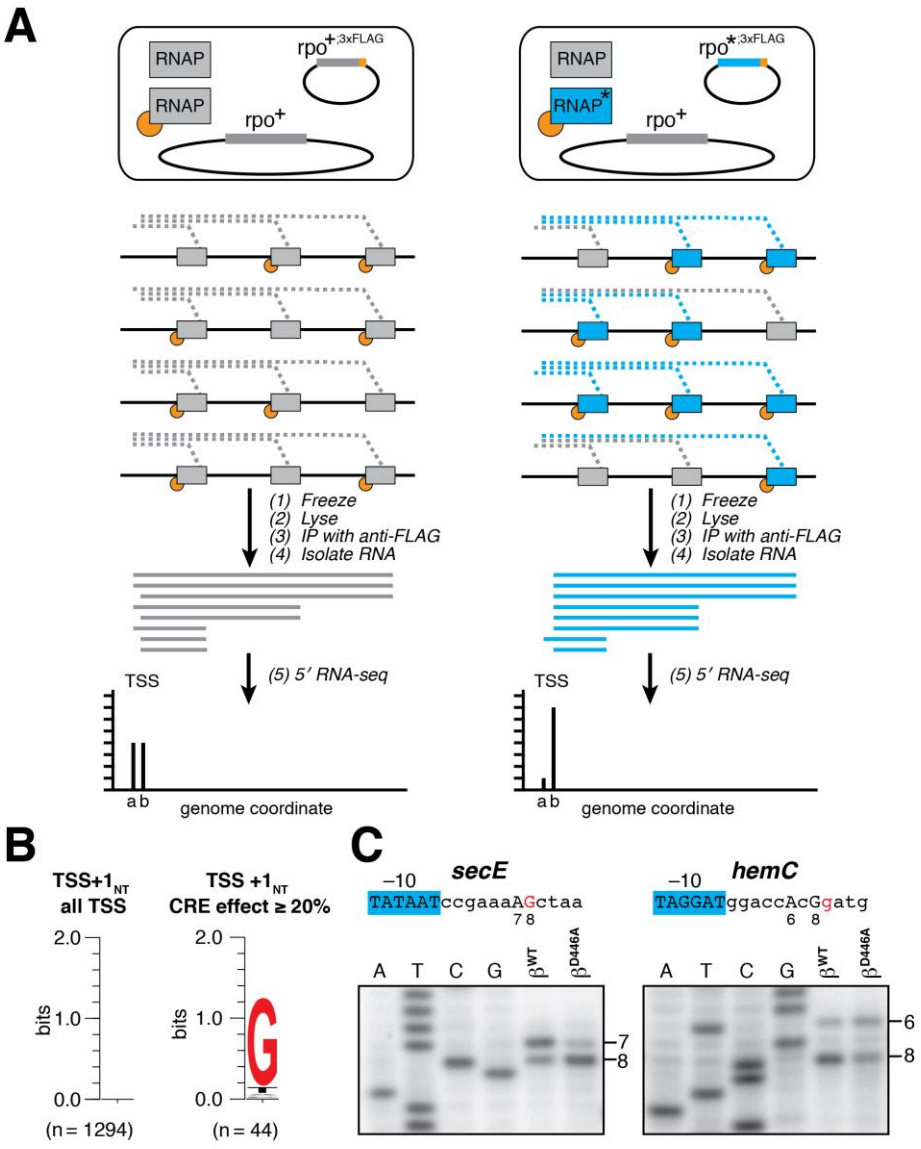


Figure 5. Effects of disrupting RNAP-G_{CRE} interactions *in vivo*: 5′ mNET-seq analysis of *E. coli* transcriptome

(A). Steps in 5′ mNET-seq analysis of natural promoters: (top) RNAP derivatives in cells (the blue RNAP derivative with asterisk is RNAP-β^{D446A}); (middle) RNAPs on the same transcription unit in four cells (RNA products in blue are associated with RNAP-β^{D446A}); (bottom) isolation of RNA products after mmunoprecipitation with anti-FLAG and sequencing analysis of RNA 5′ ends. In this example, TSS selection at genome coordinate labeled "a" is decreased with the mutant RNAP derivative.

(B). Sequence preferences for TSS+1_{NT}. Sequence logo (Crooks et al., 2004) for TSS+1_{NT} of above-threshold TSS associated with natural promoters (left) and TSS associated with natural promoters that exhibited a large, ≥ 20%, reduction in %TSS in 5′ mNET-seq analysis of RNAP-β^{D446A} vs. RNAP-β^{WT} (see Table 1).

(C). Primer-extension analysis of TSS selection *in vitro* from natural promoters. RNA products were generated in reactions performed with RNAP-β^{WT} or RNAP-β^{D446A} and templates carrying P*secE* (left) or P*hemC* (right). The sequence of each promoter, including -10 element and 12 downstream bp, is provided. In the case of P*secE*, bands corresponding to a TSS at A7 or G8 are indicated. In the case of P*hemC*, bands corresponding to a TSS at A6 or G8 are indicated. Base in red is G_{CRE} associated with the TSS at A7 of P*secE* or with the TSS at G8 of P*hemC*.

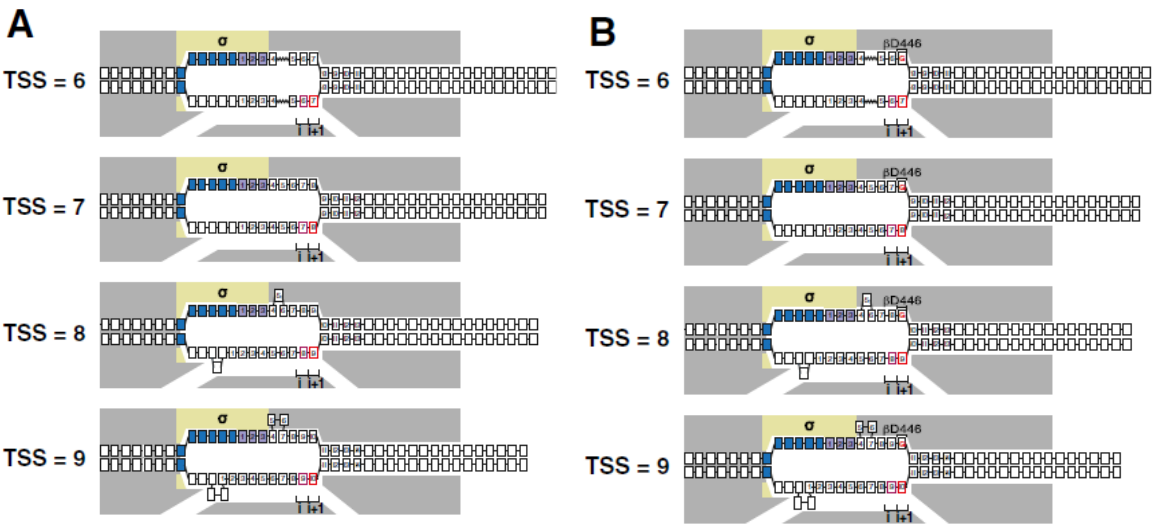


Fig. S1. Model for TSS selection and hypothesis for effects of RNAP–CRE interactions on TSS selection.
(A) Model for TSS selection: changes in TSS selection result from changes in DNA scrunching and antiscrunching in RPo (11–13). TSS = 6 through TSS = 9, RPo engaged in TSS selection at positions 6 through 9 nt downstream of −10 element; gray, RNAP; yellow, σ; blue boxes, −10 element nucleotides; purple boxes, discriminator nucleotides; black boxes, DNA nucleotides (non–template-strand nucleotides above template-strand nucleotides; nucleotides downstream of −10 element numbered); pink box, TSST; red box, TSS+1T; i and i+1, RNAP active-center initiating NTP binding site and extending NTP binding site. Scrunching is indicated by bulged-out nucleotides. Antiscrunching is indicated as a stretched nucleotide–nucleotide linkage. Exact positions and conformations of scrunched nucleotides of the nontemplate and template DNA strands remain to be determined.
(B) Hypothesis for effects of RNAP–CRE interactions on TSS selection: RNAP–CRE interactions modulate TSS selection by modulating the extent of scrunching and antiscrunching in RPo. TSS = 6 through TSS = 9, RPo engaged in TSS selection at positions 6 through 9 nt downstream of −10 element at promoters that contain GCRE (i.e., promoters that contain G at the non–template-strand position opposite TSS+1T). Red "G,"GCRE. Black "βD446," RNAP β-subunit residue that makes sequence-specific favorable interaction with GCRE. Other rendering and colors as in A.

Table S1. TSS positions in natural promoters in E. coli that exhibited large, ≥20%, CRE effects.

Promoter sequence (putative −10 element is underlined; TSS is capitalized; G_{CRE} is in bold)	$TSS{+}1_{NT}$	CRE effect (%)	Genome coordinate	Strand	Downstream gene	Distance (in bp) to start of downstream gene
ggagattgcccatcccgccatcctggtctaagcttggaaaGgatcaa	g	39	1490753	–	*gapC*	40
aagcattttcttatacccgttcagacgttattcttatttcAgatcat	g	36	1986797	+	*ftnB*	128
tttttacctacgcaggctatttcttcggtacaatcccgatGgttcag	g	35	2434193	–	*accD*	267
accctgcattgtgtcctctctttggtactaagcttacttGgagtaa	g	35	3441061	+		
atgaaagattaattagtcaagattatgatatctttttaacGgataat	g	33	2315698	+	*rcsB*	479
tatcccgagcggtttcaaaattgtgatctataattaacaaAgtgatg	g	33	4611152	+	*osmY*	244
cttaaccggagggtgtaagcaaacccgctacgcttgttacAgagatt	g	33	4330407	+	*proP*	95
cgcacaaatcatatgaaaaatgaatgcttatactgaagacCgcgctt	g	32	4439413	+	*ytfK*	174
ctatctttattgccagcctggcctttggtagcgtagatccAgaactg	g	31	18553	+	*nhaR*	162
GgcgcaaacgtctgggtgctcggtctgttactgttcttccAgcaaat	g	30	3371448	–	*nanE*	412
ggatgacgcggcaaatcaggtgctctgctactgttatgaaGgtaacc	g	28	671079	–	*nadD*	507
tctgtgtggtgcacgccgcacgggcgtctatattcttgttGgcgtgg	g	28	18122	+	*nhaR*	593
aagcgcggcaggctgttgtcgaccaggctaaactggaggaGgaaatg	g	27	2910110	–	*pyrG*	444
tgttgaaaaaatttttccccgttttgactaaaatgcgccaGgattga	g	27	850212	+	*ompX*	238
aggatctccgttgctttatgagtcatgatttactaaaggcTgcaact	g	27	519164	+	*ybbA*	569
acactgacatcactctggcaaggatgttaggatggaccacGgatgat	g	27	3990789	–	*hemC*	23
cctcgacgaagctgacatttattgcggtattattgctgatGgcctgc	g	27	702007	–	*nagC*	413
AagaatcgcgtcgattgctggtgaatgctaataatgtattGgctcgg	g	26	2522223	–		
tggtacgccgctcatcgcacaactgtttatgatctattacGgcctgc	g	25	1998251	–	*yecC*	437
cgccgcctggacaccgctctccgtctggtataatgatgccGgacagg	g	25	633069	–		
tgttgtttaaaaattgttaacaattttgtaaaatаccgacGgataga	g	25	127717	+	*lpd*	195
tgctggcgcagcaaatgcggctgaagtctataataaagatGgtaata	g	25	2034135	+		
attgctgagacaggctctgttgagggcgtataatccgaaaAgctaat	**g**	25	4177279	+	*secE*	79
gaacctggcaaaagagaccgttgatttctatgatttgaaaGgcgatc	**g**	25	1796090	–	*ihfA*	538
ccgccaccccgtacctctgataatggtctaaaatcattgaAgccact	**g**	25	2035631	+	*hchA*	204
gcaactggtcacgctgatcgacgaagggtacactagcggaAgtcagt	**g**	24	147583	+		
tgccctttaaaattcggggcgccgaccccatgtggtctcaAgcccaa	**g**	24	1862617	+	*gapA*	154
acatattgactaatttctgtaactgcataatctgatagacCgcgcct	**g**	24	3354064	+	*gltB*	661
ttcattcacaatactggagcaatccagtatgttcattctcTggtata	**g**	24	4161080	+	*fabR*	44
tattcttttaaaaaaaggggtaacaccgtaatctcataccGgtacgc	**g**	23	2939320	+	*fucR*	48
caacgtcaataaaatcaaaatcatcgtctattctctttgtGgtctgc	**g**	22	3935969	+	*rbsB*	309
ccgtgttgcgcaatttgtcaacgaaaacaataatgcgtaaGgtagaa	**g**	21	2520783	–	*gltX*	111
aacgtaatcacggacggtaaaatccgctacgctgtaatacAggctgt	**g**	21	158624	+		
tttccctcgatcccaacgagcgcattggtaaactgcgtcaGgatcag	**g**	21	856176	+		
tccttgctttaaaacgttataagcgtttaaattgcgcttcAggtgct	**g**	21	419421	+	*brnQ*	170
aaaccggatacgttccgccacagtggtgtacaatagaacaAgctatt	**g**	21	1323765	+	*yciO*	333
gcggattgacggatcatccgggtcgctataaggtaaggatGgtctta	**g**	21	3639984	+	*uspA*	127
cgctggcggcggcgcgtttaagcaggtattagtagatagcGgtgtcg	**g**	20	4491570	–	*idnR*	431
tcactttggtgatttcaccgtaactgtctatgattaatgaGgcggtg	**g**	20	3248888	+	*yqjC*	81
ccgctcgcatttttccctaagttaaatgagtaatctgatgGtgtgta	t	34	1498605	+	*ydcH*	46
actgcagcattctctttacctctgttgcagaatcttgatcCtgagct	t	24	3307652	–		
ttggttgacattcatatgaaaaaaatcataattccatcatGtttgtg	t	24	345252	+	*yahM*	152
cccgatctttttttgtcacttttgtataaaatgccagggtGatggtt	a	20	4074647	+	*yihW*	22
aacggtaaaggcgaagtcgatgatatcgaccacctcggcaAccgtcg	c	27	4182593	+		

Table S2. Plasmids used in this study

Plasmid name	Description	Source
pMASTER-*lac*CONS-N7	TSS-region library containing 16,295 (~99.5%) of a possible 16,384 sequences at positions 4- to 10-bp downstream of the −10 element of the *lac*CONS promoter	(11)
pRL706	Used for overexpression and purification of RNAP-β^{WT}	(34)
pRL706-β^{D446A}	Used for overexpression and purification of RNAP-β^{D446A}	(16)
pRL706-$\beta^{WT;3xFLAG}$	Used for overexpression of 3×FLAG tagged RNAP-β^{WT}	(16)
pRL706-$\beta^{D446A;3xFLAG}$	Used for overexpression of 3×FLAG tagged RNAP-β^{D446A}	(16)
pBEN 493	Plasmid derived from pACYC184; inserts can be cloned upstream of the tR′ terminator on a HindIII, BamHI fragment	(35)
pHV-S01	pBEN493 carrying *lac*CONS promoter derivative with CGCTGAT TSS-region sequence (in italics) between the HindIII site and BamHI site (used for experiments shown in Fig. 3*B*). Sequence of insert: AAGCTTGTTCAG-AGTTCTACAGTCCGACGATCaggcTTGACActttatgcttcggctcgTATAATgtg*CG-CTGAT*gtgagcggataacaatGGATCC	Present work
pHV-S02	pBEN493 carrying *lac*CONS promoter derivative with CGCTAAT TSS-region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3*B*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgcttcg-gctcgTATAATgtg*CGCTAAT*gtgagcggataacaatGGATCC	Present work
pHV-S03	pBEN493 carrying *lac*CONS promoter derivative with CGCTCAT TSS-region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3*B*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgcttcgg-ctcgTATAATgtg*CGCTCAT*gtgagcggataacaatGGATCC	Present work
pHV-S04	pBEN493 carrying *lac*CONS promoter derivative with CGCTTAT TSS-region sequence between HindIII and BamHI sites (used for experiments shown in Fig. 3*B*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgctt-cggctcgTATAATgtg*CGCTTAT*gtgagcggataacaatGGATCC	Present work
pHV-S05	pBEN493 carrying *lac*CONS promoter derivative with AACGGCA TSS-region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3A). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgctt-cggctcgTATAATgtg*AACGGC*Agtgagcggataacaat GGATCC	Present work
pHV-S06	pBEN493 carrying *lac*CONS promoter derivative with AACGACA TSS-region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3*A*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgctt-cggctcgTATAATgtg*AACGACA*gtgagcggataacaatGGATCC	Present work
pHV-S07	pBEN493 carrying *lac*CONS promoter derivative with AACGCCA TSS-region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3*A*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgctt-cggctcgTATAATgtg*AACGCCA*gtgagcggataacaatGGATCC	Present work
pHV-S08	pBEN493 carrying *lac*CONS promoter derivative with AACGTCA TSS region sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 3*A*). Sequence of insert: AAGCTTGTTCAGAGTTCTACAGTCCGACGATCaggcTTGACActttatgctt-cggctcgTATAATgtg*AACGTCA*gtgagcggataacaatGGATCC	Present work
pHV-S17	pBEN493 carrying P*secE* sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 5*C*). Sequence of insert (TSS positions are capitalized): AAGCTTcttttgcacgctttcgtaccagaacctggctc-atcagtgattttctttgtcataatcattgctgagacaggctctgttgagggcgtataatccgaaaAG-ctaatacgcgtttcGGATCC	Present work
pHV-S18	pBEN493 carrying P*hemC* sequence between the HindIII site and BamHI site (used for experiments shown in Fig. 5*C*). Sequence of insert (TSS positions are capitalized): AAGCTTtcaggatccactgccagacctcattttacgg-tttgcgcaggcgtctacgtttcaccacaacactgacatcactctggcaaggatgttaggatggaccAc-Ggatgataatgacgg GGATCC	Present work

Table S3. Oligonucleotides used in this study

Name	Description	Sequence
HV100	HindIII upstream primer	5′-TATAAAGCTTGTTCAGAGTTCTACAGTCCGACGA-3′
HV101	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacATCAGCGcacATTATAcgagccga-3′
HV102	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacATTAGCGcacATTATAcgagccga-3′
HV103	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacATGAGCGcacATTATAcgagccga-3′
HV104	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacATAAGCGcacATTATAcgagccga-3′
HV105	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacTGCCGTTcacATTATAcgagccga-3′
HV106	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacTGTCGTTcacATTATAcgagccga-3′
HV107	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacTGGCGTTcacATTATAcgagccga-3′
HV108	BamHI downstream primer	5′-TGCCGGATCCattgttatccgctcacTGACGTTcacATTATAcgagccga-3′
HV117	−100 HindIII upstream primer, *secE*	5′-TATAAAGCTTctttttgcacgctttcgtaccag-3′
HV118	+15 BamHI downstream primer, *secE*	5′-TGCCGGATCCgaaacgcgtattagcttttcg-3′
HV119	−100 HindIII upstream primer, *hemC*	5′-TATAAAGCTTtcaggatccactgccagacctc-3′
HV120	+15 BamHI downstream primer, *hemC*	5′-TGCCGGATCCccgtcattatcatccgtggt-3′
HV121	Biotinylated upstream primer for making IVT templates (anneals 35 base pairs upstream of HindIII site in BN493)	5′-/5Biosg/GTTGTAATTCTCATGTTTGACAGC-3′
HV122	Downstream primer for making IVT templates	5′-GGTCCTCGCCGAAAATGACCCAG-3′
HV123	For primer extension analyses of RNAs produced from promoter-tR′ fusions	5′-CCTCTCTGCCGGATCC-3′
BN436	Sequencing primer	5′-GATTTCAGTGCAATTTATCTC-3′
s1086	Illumina RA5+6N	5′-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3′(all bases are RNA)
s1082	Illumina RTP+9N	5′-GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3′
Illumina RP1		5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3′
Illumina RPI1		5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGC-ACCCGAGAATTCCA-3′
s1206	Illumina RA5+15N	5′-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNNNNNNNNNNN-3′(all bases are RNA)
s128	RT primer used for generation of MASTER libraries	5′-CCTTGGCACCCGAGAATTCC-3′
s1115	Custom primer used for Illumina sequencing	5′-CTACACGTTCAGAGTTCTACAGTCCGACGATC-3′

Table S3. Samples for high-throughput sequencing

Sample serial no.	Description
RNA libraries	
VV631	5′ mNET-seq, RNAP-β^{WT}, replicate 1
VV632	5′ mNET-seq, RNAP-β^{WT}, replicate 2
VV655	5′ mNET-seq, RNAP-β^{WT}, replicate 3
VV656	5′ mNET-seq, RNAP-β^{WT}, replicate 4
VV633	5′ mNET-seq, RNAP-β^{D446A}, replicate 1
VV634	5′ mNET-seq, RNAP-β^{D446A}, replicate 2
VV657	5′ mNET-seq, RNAP-β^{D446A}, replicate 3
VV658	5′ mNET-seq, RNAP-β^{D446A}, replicate 4
VV854	MASTER in vitro, RNAP-β^{WT}, replicate 1
VV855	MASTER in vitro, RNAP-β^{WT}, replicate 2
VV860	MASTER in vitro, RNAP-β^{D446A}, replicate 1
VV861	MASTER in vitro, RNAP-β^{D446A}, replicate 2
VV871	MASTER in vivo, RNAP-β^{WT}, replicate 1
VV872	MASTER in vivo, RNAP-β^{WT}, replicate 2
VV873	MASTER in vivo, RNAP-β^{WT}, replicate 3
VV874	MASTER in vivo, RNAP-β^{D446A}, replicate 1
VV875	MASTER in vivo, RNAP-β^{D446A}, replicate 2
VV876	MASTER in vivo, RNAP-β^{D446A}, replicate 3
DNA template libraries	
VV891	Used for analysis of VV854, VV855, VV860, VV861
VV782	Used for analysis of VV871
VV783	Used for analysis of VV871
VV914	Used for analysis of VV871
VV784	Used for analysis of VV872
VV904	Used for analysis of VV872
VV905	Used for analysis of VV872
VV786	Used for analysis of VV873
VV906	Used for analysis of VV873
VV907	Used for analysis of VV873
VV788	Used for analysis of VV874
VV790	Used for analysis of VV875
VV792	Used for analysis of VV876

3.3. Materials and methods

Plasmids and Oligonucleotides. Plasmids are listed in Table S2. Oligonucleotides are listed in Table S3.

Proteins. RNAP-βWT holoenzyme and RNAP-βD446 holoenzyme were prepared from E. coli strain XE54 (Tang et al., 1994) transformed with plasmids pRL706 or pRL706-βD446A, respectively, using procedures described in ref (Mukhopadhyay et al., 2003b). Also, discussed in chapter 1.

In Vitro Transcription Assays. For MASTER experiments shown in Fig. 2, single round in vitro transcription assays were performed essentially as described in ref. 11 using a linear DNA template containing the placCONS-N7 library (Fig. 1B, Upper). RNA products were purified and TSS selection was analyzed by 5′ RNAseq as described in ref. 11 (see Table S4 for list of samples). In vitro transcription assays shown in Figs. 3 and 5C were performed essentially as described in ref. (Goldman, Sharp et al., 2011). RNA products generated in these reactions were analyzed by primer extension as described in ref. (Goldman et al., 2011).

Analysis of TSS Selection in Vitro by MASTER.

(a) Preparation of template DNA. pMASTER-lacCONS-N7 plasmid DNA was diluted to ∼109 molecules/μL. One microliter of diluted DNA was amplified by emulsion PCR using a Micellula DNA Emulsion and Purification Kit (Chimerx) in detergent-free Phusion HF reaction buffer containing 5 μg/mL BSA, 0.4 mM dNTPs, 0.5 μM Illumina RP1 primer, 0.5 μM Illumina RPI1 primer, and 0.04 U/μL Phusion HF polymerase (Thermo Scientific). Emulsion PCR reactions were performed with an initial denaturation step of 10 s at 95 °C,

amplification for 30 cycles (denaturation for 5 s at 95 °C, annealing for 5 s at 60 °C, and extension for 15 s at 72 °C), and a final extension for 5 min at 72 °C. The emulsion was broken, and DNA was purified according to the manufacturer's recommendations. DNA was recovered by ethanol precipitation and resuspended in 30 μL nuclease-free water.

(b) Transcription reactions. In vitro transcription assays were performed by mixing 10 nM template DNA with 50 nM RNAP-βWT holoenzyme or 50 nM RNAP-βD446A holoenzyme in transcriptionbuffer [50 mM Tris·HCl (pH 8.0), 10 mM MgCl2, 0.01 mg/mL BSA, 100 mM KCl, 5% (vol/vol) glycerol, 10 mM DTT, and 0.4U/μL RNase OUT]. RNAP-promoter open complexes were allowed to form by incubation at 37 °C for 10 min. A single round of transcription was initiated by addition of a mixture of NTPs to a final concentration of 1 mM and heparin to a final concentration of 0.1 mg/mL. After 15 min, reactions were stopped by addition of EDTA (pH 8) to a final concentration of 10 mM. Nucleic acids were recovered by ethanol precipitation and resuspended in 30 μL nuclease-free water.

(c) Purification of RNA products. Nucleic acids recovered from the ethanol precipitation were treated with 2 U TURBO DNase (Life Technologies) at 37 °C for 1 h, mixed with an equal volume of 2×RNA loading dye [95% (vol/vol) deionized formamide, 18 mM EDTA, 0.25% (wt/vol) SDS, xylene cyanol, bromophenol blue, and amaranth], and separated by electrophoresis on 10% (wt/vol) acrylamide, 7 M urea slab gels (equilibrated and run in 1× TBE). The gel was stained with SYBR Gold nucleic acid gel stain (Life Technologies), bands were visualized on a UV transilluminator, and RNA transcripts ∼100 nt in size were excised from the gel. The excised gel slice was crushed and incubated in 300 μL 0.3 M NaCl in 1× TE buffer at 70 °C for 10 min. Eluted RNAs were separated from crushed gel

fragments using a Spin-X column (Corning). After the first elution, the crushed gel fragments were collected; the elution procedure was repeated; and nucleic acids were collected, pooled with the first elution, isolated by ethanol precipitation, and resuspended in 25 μL RNase-free water. Purified RNA products were analyzed by 5′ RNA-seq using the procedure described in the next section.

(d) 5′ RNA-seq. Before cDNA library construction 5′ triphosphate RNA products were converted to 5′ monophosphate RNA. To do this, ∼100 ng purified RNA was treated with 20 U 5′-RNA polyphosphatase (New England Biolabs). Samples were extracted with acid phenol:chloroform (pH 4.5). RNA products were recovered by ethanol precipitation and resuspended in 10 μL RNase-free water. Ligation of adaptor to 5′ end of RNA products. RNA products were combined with PEG 8000 [10% (wt/vol) final concentration], oligo s1206 (1 pmol/μL final concentration), ATP (1 mM final concentration), 40 U RNase OUT, 1× T4 RNA ligase 1 reaction buffer (New England Biolabs), and 10 U of T4 RNA ligase 1 (New England Biolabs) in a total volume of 30 μL. The mixture was incubated at 16 °C for 16 h. Size selection of adaptor-ligated RNA products. Adaptor-ligated RNA products were mixed with an equal volume of 2× RNA loading dye and separated by electrophoresis on 10% (wt/vol) acrylamide, 7 M urea slab gels (equilibrated and run in 1× TBE). The gel was stained with SYBR Gold nucleic acid gel stain, bands were visualized with UV transillumination, and species ranging from ∼80 to ∼300 nt were excised from the gel. RNA products were eluted from the gel using the procedure described above, isolated by ethanol precipitation, and resuspended in 10 μL nuclease-free water. cDNA synthesis. Ten microliters of gel-eluted RNA products was mixed with 0.3 μL s128 oligonucleotide (100 pmol/μL), incubated at 65 °C for 5 min, and cooled to 4 °C; 9.7 μL of a mixture containing

4 μL 5× First-Strand buffer (Life Technologies), 1 μL 10 mM dNTP mix, 1 μL 100 mM DTT, 1 μL (40 U) RNase OUT, 1 μL (200 U) SuperScript III Reverse Transcriptase (Life Technologies), and 1.7 μL nuclease-free water was added to the RNA/oligonucleotide mixture. The reactions were incubated in a thermal cycler with a heated lid at 25 °C for 5 min, followed by 55 °C for 60 min and 70 °C for 15 min. Reactions were cooled to room temperature, 10 U RNase H (Life Technologies) was added, and the reactions were incubated at 37 °C for 20 min.

(e) Size selection of cDNA products. An equal volume of 2× RNAloading dye was added, and nucleic acids were separated by electrophoresis on 10% (wt/vol) acrylamide, 7 M urea slab gels (equilibrated and run in 1× TBE). The gel was stained with SYBR gold nucleic acid gel stain, and ∼80 to ∼150 nt species were excised from the gel. cDNA products were recovered from the gel using the procedure described above and resuspended in 10 μL nuclease-free water.

(f) Amplification of cDNA products. Five microliters of gel-isolated cDNA products were added to a mixture containing 1× Phusion HF reaction buffer, 0.2 mM dNTPs, 0.25 μM Illumina RP1 primer, 0.25 μM Illumina index primer, and 0.02 U/μL Phusion HF polymerase. PCR was performed with an initial denaturation step of 30 s at 98 °C, amplification for 11 cycles (denaturation for 10 s at 98 °C, annealing for 20 s at 62 °C, and extension for 10 s at 72 °C), and a final extension for 5 min at 72 °C. Purification of cDNA products. Amplified cDNA products were separated by gel electrophoresis using a nondenaturing 10% (wt/vol) acrylamide slab gel (equilibrated and run in 1× TBE). The gel was stained with SYBR Gold nucleic acid gel stain, and species at ∼170 bp were excised

from the gel. cDNA products were eluted from the gel with 600 μL 0.3 M NaCl in 1× TE buffer at 37 °C for 2 h, precipitated, and resuspended in 13 μL nucleasefree water.

(g) High-throughput sequencing. Libraries were sequenced on an Illumina HiSeq 2500 platform in rapid mode using custom primer s1115.

(h) Data analysis. Sequencing of template DNA (sample VV891) (Table S4) was used to associate the 7-bp randomized sequence in the region of interest with a corresponding second 15-bp randomized sequence that serves as its barcode. The identity of the 15-bp barcode in each RNA product was used to determine the identity of bases at positions 4–10 of the lacCONS template from which the RNA product was generated. Sequences derived from the RNA 5′ end of reads that were perfect matches to the sequence of the template were used for analysis of TSS selection. Experiments were performed in duplicate (samples VV854 and VV855 for RNAP-βWT and samples VV860 and VV861 for RNAP-βD446A) (Table S4).

Analysis of TSS Selection in Vivo from 47 (~16,000) Consensus Promoter Derivatives.

(a) Cell growth. Escherichia coli DH10B-T1R cells (Life Technologies) containing plasmids pRL706-βWT;3xFLAG or pRL706-βD446A;3xFLAG were transformed with ~50 ng pMASTER-lacCONS-N7 library to obtain a 25-mL overnight culture representing cells derived from at least 20 million unique transformants; 0.5 mL of the overnight cell culture was used to inoculate 50 mL LB media containing 100 μg/μL carbenicillin and 25 μg/μL chloramphenicol. When the cell density reached anOD600 ~0.3, 1 mMIPTG was added, and cells were grownfor an additional 2 h. Cell suspensions were divided equally among 12 × 2-mL tubes (BioExcell) and centrifuged (1 min, 21,000 × g at room

temperature) to collect cells, and supernatants were removed. Cell pellets were then rapidly frozen on dry ice and stored at −80 °C. pMASTER-lacCONS-N7 plasmid DNA was isolated from these cells using a Plasmid Miniprep kit (Qiagen). Plasmid DNAwas used as template in emulsion PCR reactions to generate a product that was sequenced to assign barcodes (see below).

(b) RNA isolation. Cells pellets derived from 12 mL culture were resuspended in 1 mL lysis buffer (B-Per, Bacterial Protein Extraction Reagent; Thermo Scientific) supplemented with one quarter of a protease inhibitor mixture tablet (complete Mini EDTA-free; Roche), 1mMEDTA, 80UMurine RNase Inhibitor (NEB), 100 μg lysozyme (Thermo Scientific), and 150 U DNase I (Thermo Scientific) and incubated for 10 min. The lysate was then clarified by centrifugation (10 min, 21,000 × g), and NaCl was added to a final concentration of 150 mM. The lysate was added to 1 mL anti-FLAG M2 affinity gel (Sigma Aldrich) that had been washed three times with 3 mL 1× TBS and equilibrated in 3 mL wash buffer (B-Per solution containing 150 mM NaCl, 1 mM EDTA, 50 U/mL Murine RNase Inhibitor, and protease inhibitor mixture [complete EDTA-free (Roche); 1 tablet per 50 mL]). The lysate and affinity gel mixture was nutated at 4 °C for 2.5 h in a 1.7-mL centrifuge tube. The mixture was transferred to a 10-mL Econo-Pack disposable chromatography column (Bio-Rad), the flow through was collected, and the affinity gel was washed eight times with 5 mL wash buffer and three times with 250 μL elution buffer (B-Per solution containing 150 mM NaCl, 1 mM EDTA, 50 U/mL Murine RNase Inhibitor, and 2 mg/mL 3× FLAG peptide; GenScript). For the washes with elution buffer, the affinity gel was incubated for 30 min before collection of the fractions. The presence of epitope tagged βWT or βD446A was analyzed in each fraction by immunoblotting. To isolate the

RNA products associated with RNAP, pooled eluates from above were mixed with three volumes of TRI Reagent solution (Molecular Research Center), incubated at 70 °C for 10 min, and centrifuged (10 min, 21,000 × g) to remove insoluble material. The supernatant was transferred to a fresh tube, ethanol was added to a final concentration of 60.5% (vol/vol), and the mixture was applied to a Direct-zol spin column (Zymo Research). DNase I treatment was performed on-column according to the manufacturer's recommendations. RNA products were eluted from the column with three sequential portions of 30 μL nuclease-free water that had been heated to 70 °C. Before cDNA library construction, RNA products were treated with 4 U TURBO DNase (Ambion) at 37 °C for 1 h. Following DNase treatment, samples were extracted with acid phenol:chloroform, and RNA products were recovered by ethanol precipitation and resuspended in RNase free water.

(c) 5′ RNA-seq. Before cDNA library construction, 5′ monophosphate RNA products were first removed by treatment of 0.75–1.3 μg of RNA with 1 U Terminator 5′-Phosphate-Dependent Exonuclease (Epicentre). Samples were extracted with acid phenol:chloroform, RNA products were recovered by ethanol precipitation and resuspended in RNase-free water. Next, 5′ triphosphate RNA products were converted to 5′ monophosphate RNA products by treating samples with 20 U 5′-RNA polyphosphatase as described in ref. 29. Samples were extracted with acid phenol:chloroform, and RNA products were recovered by ethanol precipitation and resuspended in 10 μL RNase-free water. 5′ RNA-seq analysis was performed as described above.

(d) Data analysis. In vivo MASTER experiments were performed in triplicate (samples VV871, VV872, and VV873 for RNAP-βWT and samples VV874, VV875, and VV876

for RNAP-βD446A) (Table S4). pMASTER-lacCONS-N7 plasmid DNA isolated from each individual cell culture was used as template in emulsion PCR reactions to generate products that were sequenced to assign barcodes as described in ref. 11. For each RNAP-βWT sample, three emulsion PCR products were generated and sequenced (Table S4). For each RNAP-βD446A sample, one emulsion PCR product was generated and sequenced (Table S4). The identity of the 15-bp barcode in each RNA product was used to determine the identity of bases at positions 4–10 of the lacCONS template from which the RNA product was generated. Sequences derived from the RNA 5′ end of reads that were perfect matches to the sequence of the template were used for analysis of TSS selection.

5′ mNET-seq experiments.

For the in vivo MASTER experiments shown in Fig. 4, E. coli DH10B-T1R cells (Life Technologies) containing plasmids pRL706-βWT;3xFLAG or pRL706-βD446A;3xFLAG were transformed with ∼50 ng pMASTER-lacCONS-N7 library to obtain a 25-mL overnight culture representing cells derived from at least 20 million unique transformants; 0.5 mL of the overnight cell culture was used to inoculate 50 mL LB media containing 100 μg/μL carbenicillin and 25 μg/μL chloramphenicol. When the cell density reached an OD600 ∼0.3, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added, and cells were grown for an additional 2 h. RNA associated with RNAP was isolated using procedures described in ref. 16. For the experiments shown in Fig. 5, MG1655 cells containing plasmids pRL706-βWT;3xFLAG or pRL706-βD446A;3xFLAG were shaken at 220 rpm at 37 °C in 100 mL 4× LB (40 g Bacto tryptone, 20 g Bacto yeast extract, and 10 g NaCl per liter) containing 200 μg/μL carbenicillin in 500-mL DeLong flasks

(Bellco).When cell density reached an OD600 ~0.6, 1 mM IPTG was added, and cells were grown for an additional 4 h. RNA associated with RNAP was isolated using procedures described in ref. 16. RNA products associated with RNAP were analyzed by 5′ RNA-seq using procedures described in ref. (Vvedenskaya, Goldman et al., 2015) (see Table S4 for list of samples).

Analysis of TSS Selection in Vitro by Primer Extension.

(a) Preparation of template DNA. Linear DNA templates were generated by PCR using plasmids pHV-S01, pHV-S02, pHV-S03, pHV-S04, pHV-S05, pHV-S06, pHV-S07, pHV-S08, pHV-S17, or pHV-S18 as template and oligonucleotide primer HV121, which contains a 5′ biotin moiety, and oligonucleotide primer HV122. The biotinylated linearDNAtemplates generated by PCR were bound to streptavidin-coated paramagnetic beads [Streptavidin MagnaSphere Paramagnetic Particles (SA-PMPs); Promega]. To do this, 100 μL SA-PMP slurry per each DNA template was washed three times with 100 μL binding buffer [10 mM Tris (pH 8), 150 mM NaCl, and 100 μg/mL BSA]. The SA-PMPS were resuspended in 100 μL binding buffer, 2.5 μL 400 nM DNA template stock was added to each slurry, and the mixture was gently mixed for 30 min at 25 °C. The binding buffer was removed, SA-PMPs were washed three times with 1× TB [40 mM Tris (pH 8), 10mM MgCl2, 50mMKCl, 10mMβ-mercaptoethanol, 10 μg/mL BSA, and 5% (wt/vol) PEG-8000], and resuspended in 10 μL reaction buffer to obtain 100 nM SA-PMP–conjugated DNA templates stock solutions that were used for the transcription assays.

(b) Transcription reactions. In vitro transcription assays were performed by mixing 50 nM RNAP with 10 nM template (attached to beads) for 10 min at 37 °C in 1× TB. Transcription was initiated by adding NTPs to a final concentration of 100 μM. The total

reaction volume was 20 μL. Reactions were stopped after 10 min by adding 100 μL stop solution [0.5 mg/mL glycogen and 10 mM EDTA (pH 8.0)]. Magnetic beads were pelleted using a MagneSphere TechnologyMagnetic Separation Stand (Promega), and the supernatant was transferred to a fresh tube and extracted with acid phenol:chloroform. RNA transcripts were recovered by ethanol precipitation and resuspended in 12 μL water.

(c) Primer-extension reactions. Oligonucleotide primerHV123 was 32P-5′ end-labeled with T4 polynucleotide kinase in a 50-μL reaction containing 120 pmol of primer, 40 U of enzyme, and 100 μCi ofγ32P ATP (Perkin Elmer). The labeling reaction was incubated at37 °C for 1 h followed by an incubation at 95 °C for 10 min. Unincorporated nucleotides and salts were removed by passage over an Illustra G-25 microspin column (GE Healthcare). One microliter of labeled primer was mixed with 5 μL of the RNA recovered from the transcription reactions. This mixture was heated at 90 °C for 2 min and immediately transferred to ice. Reverse transcription was performed by adding 4 μL of a mixture containing 10 U AMV reverse transcriptase (New England Biolabs), AMV buffer, dNTPs (10 mM of each dNTP), and 10 U murine RNase inhibitor (New England Biolabs) to the annealed primer template mixture and incubating at 55 °C for 1 h, 5 min at 95 °C, and then cooled to 4 °C. Reactions were stopped by addition of 10 μL 98% (vol/vol) formamide containing 10 mM EDTA, 0.02% (wt/vol) bromophenol blue, and 0.02% (wt/vol) xylene cyanol. Samples were electrophoresed on an 8% (wt/vol) acrylamide, 7 M urea slab gel (equilibrated and run using a gradient buffer of 1× TBE in the upper reservoir and 1× TBE, 0.3 M NaOAc in the lower reservoir). Radiolabeled species were detected by storage-phosphor imaging. TSS assignments were made by comparison with a sequencing ladder prepared using the same radiolabeled primer used for the extension reactions and a

Sequenase Version 2.0 DNA sequencing kit (USB Corporation). Experiments were performed three independent times (one of the independent replicates for each template is shown in Fig. 3 and Fig. 5C). The values for %TSS (RNAP-βWT) −%TSS (RNAP-βD446A) reported in Fig. 3 were derived by averaging the results of the three experiments.

Chapter 4

RNAP-CRE interaction is a determinant of abortive transcription

Hanif Vahedian-Movahed and Richard H. Ebright

4.1. Background

During initial transcription, RNAP begin synthesis of RNA product by engaging in a series of short product synthesis and release. This product are referred to as abortive products (Hsu, 2009). Using in vitro abortive initiation assay, researchers have measured the rate constant for the formation of open complex for different promoters and found a direct correlation with the homology score, allowing these parameters to be considered as indicators of promoter strength in vitro. However, in most cases no correlation has been found between the in vivo promoter strength (measured as mRNA levels) and homology score (Hsu, 2009, Knaus & Bujard, 1990, Mulligan, Hawley et al., 1984) .This suggest that promoter strength cannot be solely judged based on the promoter binding and rate of RPo formation but it also depends on the steps after RPo formation including steps of abortive initiation and promoter escape. Abortive transcription is considered to be a kinetically significant off-pathway reactions that occur in transcription initiation and can compete with on-pathway productive transcription, (Buc & Strick, 2009, Hsu, 2009, Skancke et al., 2015).

During initial transcription, RNAP uses a "scrunching" mechanism, in which RNAP remains stationary on promoter DNA and unwinds and reels in downstream DNA in each

nucleotide-addition cycle (Kapanidis, Margeat et al., 2006, Revyakin, Ebright et al., 2004, Revyakin et al., 2006). During initial transcription, in each scrunching step, previous RNAP-CRE interactions are broken, and new RNAP-CRE interactions are made. Therefore, RNAP-CRE interactions are expected to affect the energetics of scrunching and therefore to play functional role in abortive transcription (probability of abortive product release).

4.2. Results and discussions

To test the hypothesis that sequence-specific RNAP-CRE interactions affect abortive transcription I used an in vitro abortive transcription radioactive assays to compare the abortive profiles of wt-RNAP to abortive profile by a mutant RNAP defective in sequence-specific RNAP-G_{CRE} interaction. This assay allow quantification of a series of important initiation parameter including: overall abortive yield (AY), productive yield (PY) and abortive-productive ratio (APR). Also, for each transcripts we can quantify the probability of abortive transcript release (Pi) by correcting for the fraction of RNAPs reaching at that template position (Hsu, 1996, Hsu, 2009). The higher the probability of transcript release Pi at each position, the higher the barrier to continuation of RNA synthesis or promoter escape at that template position.

The sequence of N25 promoter used in this work contain a G at position +9. When transcript length is 8 nt to add the next NTP, DNA moves 1 nt toward the active site of RNAP (scrunching) which places the non-template strand +9G into b-pocket to make favorable RNAP-CRE interactions. According to this model, we predict that this interaction should reduce the probability of abortive transcript release of 8 nt length (Fig. 1). In accordance to our prediction, my results in fig. 2 and fig. 3 show that disrupting

sequence-specific RNAP-G_{CRE} interactions increase the APR by a factor of 5 at position +8 which is 1 bp upstream of a consensus +9 G_{CRE}. Additionally, the probability of the abortive transcript release increased by ~20% at position 1 bp upstream of a consensus G_{CRE} at positon. Our results establish that sequence-specific RNAP–G_{CRE} interactions affects abortive transcription in vitro in a manner that correlates with the presence of G_{CRE}.

Next, we wanted to ask if we move the position of G_{CRE} at position +9, would we observe a corresponding effect in abortive transcription at the new positions. To do this I constructed N25 promoter derivatives with 2 base (CC) or 4 base insertion (CCCC) between positions +5 and +6, which places the +9G at position +11 and +13. According to our model described earlier and in fig. 1, we predict that an RNAP derivative defective in sequence-specific RNAP-G_{CRE} interaction would exhibit increased in abortive transcript release at position 1 bp upstream of the G_{CRE} (here at positions +10 and +12). According to our prediction my results in fig 2 and 4 show that disrupting sequence-specific RNAP-G_{CRE} interactions increase the APR by a factor of 5 at position +10 and +12. Additionally, the probability of the abortive transcript release (P_{+10} and P_{+10}) show increased by ~9% and ~6% at position +10 and +12 (upstream of +11 G_{CRE} and +12 G_{CRE}). We note that the CRE effect diminishes as we get further away from the transcription start site. This could be due to the natural tendency of RPitc complexes to get more successful in entering the productive stage of transcription as the transcripts get longer. This pattern can be clearly observed from the Pi values listed in Fig 3 (RNA length seems to be inversely correlated with the probability of abortive transcript release). My results establish that disrupting sequence-specific RNAP–G_{CRE} interactions affects abortive transcription in vitro in a manner that correlates with the position of G_{CRE} along the template.

I also constructed N25 promoter derivatives with 2 base deleted (+4 to +5) or 4 base deleted (+4 to +7) which places the +9G at position +7 and +5. Then, I used analogous in vitro abortive transcription radioactive assays to compare the abortive profiles of wt-RNAP to abortive profile by a mutant RNAP defective in sequence-specific RNAP-G_{CRE} interaction. According to our model described earlier and in fig. 1, we predict that an RNAP derivative defective in sequence-specific RNAP-G_{CRE} interaction would exhibit increase in abortive transcript release at position 1 bp upstream of the G_{CRE} (here at positions +6 and +4). According to our prediction my results in fig 2 and 5 show that disrupting sequence-specific RNAP-G_{CRE} interactions increase the APR by a factor of 5 and 4 at position +6 and +4, respectively. Additionally, the probability of the abortive transcript release (P_{+6} and P_{+4}) show increased by ~9% and ~34% at position +6 and +4 (upstream of +7 G_{CRE} and +5 G_{CRE}).

My results establish that disrupting sequence-specific RNAP–G_{CRE} interactions affects abortive transcription in vitro in a manner that correlates with both the presence and position of G_{CRE} along the template. We conclude that sequence-specific RNAP–CRE interactions are a determinant of abortive transcription, reducing the probability of abortive release of transcript of i-1 length.

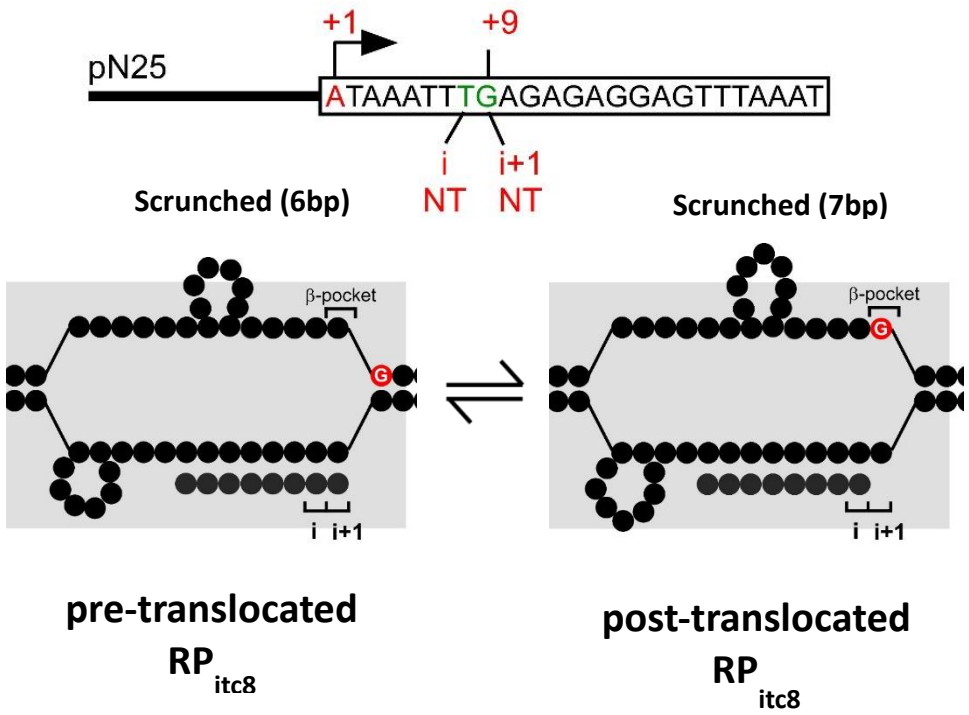


Figure 1. Hypothesis for effects of RNAP–CRE interactions on abortive transcription. RNAP–CRE interactions modulate abortive transcription by modulating the probability of RNAP-DNA complex survival at each template position. The favorable sequence-specific RNAP-CRE interactions with a G at the non–template-strand corresponding to i+1 site of the active site (i+1 NT) could make the RNAP-DNA complex more stable, RNAP could dwell longer at that position and therefore increase the probability of next NTP binding and bound formation and reducing the probability of the transcript release. Top panel shows the sequence of N25 promoter used in this work which contain a G at position +9. When transcript length is 8 nt to add next NTP, DNA moves 1nt toward the active site of RNAP (scrunching) which places the NT strand +9G into b-pocket to make favorable RNAP-CRE interactions and therefore should reduce the probability of abortive transcript of 8 nt length.

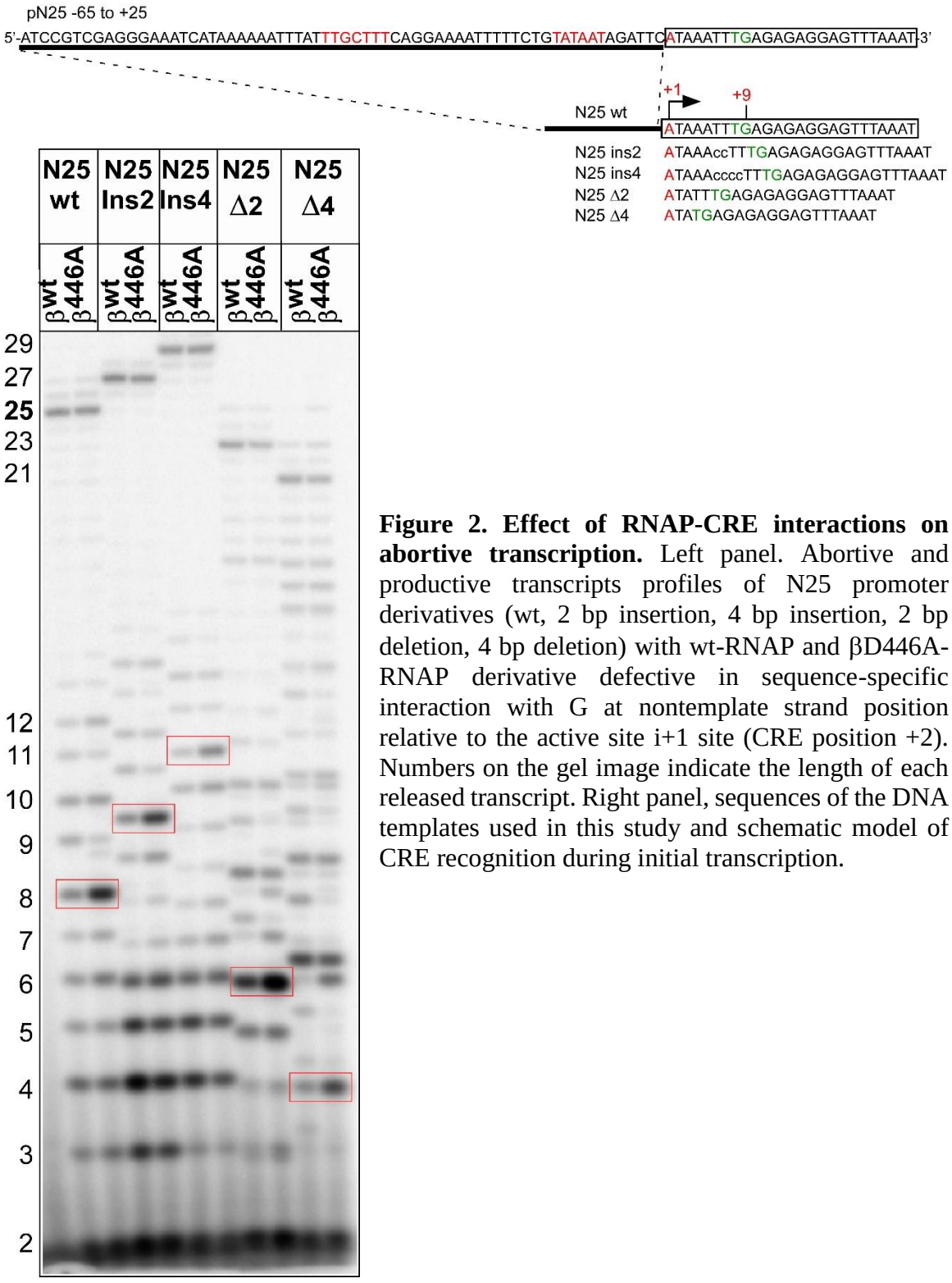


Figure 2. Effect of RNAP-CRE interactions on abortive transcription. Left panel. Abortive and productive transcripts profiles of N25 promoter derivatives (wt, 2 bp insertion, 4 bp insertion, 2 bp deletion, 4 bp deletion) with wt-RNAP and βD446A-RNAP derivative defective in sequence-specific interaction with G at nontemplate strand position relative to the active site i+1 site (CRE position +2). Numbers on the gel image indicate the length of each released transcript. Right panel, sequences of the DNA templates used in this study and schematic model of CRE recognition during initial transcription.

A

pN25 [wt]		wt-RNAP		[Ala446]β-RNAP			
seq	RNA length	APR	SEM	APR	SEM	mut/wt	SEM
T	2	245	40	394	93	2	0.2
A	3	230	101	304	76	2	0.2
A	4	344	86	418	90	1	0.0
A	5	101	22	135	24	1	0.1
T	6	103	18	209	25	2	0.1
T	7	35	6	65	7	2	0.1
T	8	158	25	709	39	5	0.5
G	9	54	9	24	1	0	0.0
A	10	67	10	78	6	1	0.1
G	11	22	2	14	1	1	0.1
A	12	17	1	34	1	2	0.2
G	13	6	0	5	0	1	0.1
A	14	1	0	3	0	3	0.2

pN25 [wt]		%Pi (probability of abortion)		CRE effect
seq	RNA length	βWT	βD446A	$\%P_i(\beta^{WT}) - \%P_i(\beta^{D446A})$
A	3	48	44	-5
A	4	42	37	-4
A	5	11	11	0
T	6	10	14	4
T	7	3	4	1
T	8	13	32	19
G	9	4	1	-3
A	10	5	3	-2
G	11	2	1	-1
A	12	1	1	0
G	13	0	0	0
A	14	0	0	0

B

	wt-RNAP	SEM	[Ala446]β-RNAP	SEM
AY	0.51	0.05	0.65	0.02
PY	0.49	0.05	0.35	0.02
APR	1.06	0.21	1.84	0.17

Table 1. Effect of disrupting RNAP-G_{CRE} interaction on abortive transcription: abortive initiation parameters. A left panel, abortive-productive ratio (APR) values of each abortive transcript. A, right panel, % of the probability of abortive transcript release of each transcript. CRE effect id shown as the differences in %Pi for reaction with WT-RNAP and for reactions with [Ala446A]□-RNAP. B. overall initiation parameters for the N25 promoter, AY: abortive yield, PY: productive yiled, APR: abortive-productive ratio.

A

pN25 [ins 2]		wt-RNAP		[Ala446]β–RNAP			
seq	RNA length	APR	SEM	APR	SEM	mut/wt	SEM
T	2	635	148	611	122	1	0.1
A	3	622	126	765	249	1	0.2
A	4	1593	497	1395	442	1	0.0
A	5	577	121	609	136	1	0.0
C	6	261	30	431	26	2	0.1
C	7	14	2	40	4	3	0.2
T	8	11	2	22	1	2	0.3
T	9	33	3	95	9	3	0.2
T	10	98	10	506	25	5	0.3
G	11	26	1	17	1	1	0.1
A	12	19	1	21	0	1	0.1
G	13	6	0	5	0	1	0.1
A	14	13	0	20	3	1	0.2
G	15	3	0	2	0	1	0.1
A	16	1	0	1	0	1	0.2

pN25 [ins 2]		CRE effect
seq	RNA length	$\%P_i(\beta^{WT}) - \%P_i(\beta^{D446A})$
A	3	6
A	4	-6
A	5	1
C	6	4
C	7	1
T	8	0
T	9	2
T	10	9
G	11	0
A	12	0
G	13	0
A	14	0
G	15	0
A	16	0

B

pN25 [ins 4]		wt-RNAP		[Ala446]β RNAP			
seq	RNA length	APR	SEM	APR	SEM	mut/wt	SEM
T	2	162	113	93	41	16	15.9
A	3	165	45	65	17	0	0.2
A	4	990	319	498	110	1	0.0
A	5	633	138	379	57	1	0.0
C	6	291	26	307	21	1	0.0
C	7	46	6	72	3	2	0.2
C	8	17	2	36	3	2	0.1
C	9	6	1	16	3	3	0.1
T	10	13	1	24	1	2	0.1
T	11	35	2	64	4	2	0.2
T	12	25	1	112	5	5	0.3
G	13	10	1	8	1	1	0.1
A	14	16	1	13	2	1	0.1
G	15	5	0	2	0	0	0.1
A	16	2	0	2	0	1	0.2
G	17	0	0	0	0	1	0.4
A	18	0	0	0	0	0	0.7

pN25 [ins 4]		CRE effect
seq	RNA length	$\%P_i(\beta^{WT}) - \%P_i(\beta^{D446A})$
A	3	-9
A	4	1
A	5	4
C	6	10
C	7	3
C	8	2
C	9	1
T	10	1
T	11	3
T	12	6
G	13	0
A	14	0
G	15	0
A	16	0
G	17	0
A	18	0

Table 2. Sequence insertion analysis of effect of RNAP-CRE interactions on abortive transcription.

A

pN25 [Δ 2]		wt-RNAP		[Ala446]β–RNAP			
seq	RNA length	APR	SEM	APR	SEM	mut/wt	SEM
T	2	361	245	1615	652	45	41.8
A	3	103	36	573	122	7	2.0
T	4	198	86	396	118	2	0.6
T	5	486	148	922	232	2	0.1
T	6	1385	257	6262	1134	5	0.2
G	7	48	8	279	15	6	0.7
A	8	90	14	55	5	1	0.0
G	9	253	37	323	30	1	0.1
A	10	49	5	31	1	1	0.1
G	11	75	9	126	6	2	0.1
A	12	18	1	15	2	1	0.2
G	13	5	0	15	0	3	0.5
G	14	4	1	6	1	1	0.4
A	15	1	0	1	1	1	0.9
G	16	2	1	4	1	2	0.8
T	17	2	0	4	1	2	0.3
T	18	16	1	29	3	2	0.1
T	19	4	1	7	2	2	0.1
A	20	5	1	6	2	1	0.0

B

pN25 [Δ 4]		wt-RNAP		[Ala446]β–RNAP			
seq	RNA length	APR	SEM	APR	SEM	mut/wt	SEM
T	2	159	89	15	117	-4	4.3
A	3	8	10	-30	14	-1	3.0
T	4	340	121	1314	378	4	0.3
G	5	92	28	5	10	0	0.1
A	6	227	76	687	168	3	0.3
G	7	919	146	848	110	1	0.0
A	8	132	6	35	4	0	0.0
G	9	196	19	229	20	1	0.0
A	10	53	1	30	0	1	0.0
G	11	61	4	116	6	2	0.0
G	12	14	0	16	0	1	0.0
A	13	25	0	12	1	0	0.0
G	14	20	1	37	5	2	0.2
T	15	6	0	6	0	1	0.0
T	16	27	1	40	1	1	0.0
T	17	26	2	41	3	2	0.0
A	18	24	2	33	2	1	0.0
A	19	10	2	11	1	1	0.0
A	20	8	0	8	1	1	0.0

pN25 [Δ 2]		CRE effect
seq	RNA length	% $P_i(\beta^{WT})$ - % $P_i(\beta^{D446A})$
A	3	4
T	4	-15
T	5	-16
T	6	9
G	7	1
A	8	-3
G	9	-6
A	10	-1
G	11	-1
A	12	0
G	13	0
G	14	0
A	15	0
G	16	0
T	17	0
T	18	0
T	19	0
A	20	0

pN25 [Δ 4]		CRE effect
seq	RNA length	%$P_i(\beta^{WT})$ - %$P_i(\beta^{D446A})$
A	3	198
T	4	34
G	5	-15
A	6	7
G	7	-23
A	8	-6
G	9	-2
A	10	-2
G	11	1
G	12	0
A	13	-1
G	14	0
T	15	0
T	16	0
T	17	0
A	18	0
A	19	0
A	20	0

Table 3. Sequence deletion analysis of effect of RNAP-CRE interactions on abortive transcription.

4.3. Material and methods

Oligos and DNA templates preparation. The oligonucleotides were purchased from IDT (table X) and were solubilized in nuclease free water. The duplex DNA template were formed by annealing of nontemplate-strand oligonucleotide (2 uM) and template-strand oligonucleotide (2 uM) in 50 µl 5 mM Tris-HCl (pH 7.7), 200 mM NaCl, and 10 mM MgCl2, heated 2min at 95°C, and cooled to 25°C in 2°C steps with 60 s per step in a thermal cycler (Applied Biosystems).

Proteins. *E. coli* wild-type RNAP holo and mutants-RNAP holo were prepared from E. coli strain XE54 (Tang et al., 1994) that were transformed with either plasmids pRL706 or pRL706-βD446A derivatives. Protein purification procedures were essentially as in (Mukhopadhyay et al., 2003b). The purified proteins included RNAP-βWT and RNAP-βD446A.

Abortive transcription assays. In vitro abortive assays were performed by mixing 50 nM RNAP with a mixtures of 30 nM N25 promoter derivatives, 100 uM NTP, 5 µCi 32P-γ-ATP (Perkin Elmer; 6000 Ci/mmol)], 200 mM KCl and 1x reaction buffer (50 mM Tris HCl pH 8.0, 10 mM MgCl2, 10 mM 2-mercaptoethanol, acetylated BSA 10 ug/ml). the protein were diluted before use to nM in protein dilution buffer (10 mM Tris HCl pH 8.0, 10 mM 2-mercaptoethanol, 10 mM KCl, 100 uM EDTA, 0.4 mg/ml BSA, 5% glycerol and 0.1% v/v Triton X-100). The total reaction volume was 10 µL. The reactions were incubated for 10 min at 37 °C. Reactions were stopped after 10 min by adding 100 µL stop solution [1 mg/mL glycogen, 10 mM EDTA (pH 8.0), 300 mM NaAc]. Add 330 ul ethanol, mix and incubate at -20 C overnight. Centrigue for 15 min at 13000 rpm. Discard the supernatant and use speed vac to dry the pellete for 15 min. Add 10 ul formamide loading

buffer (80% formamide, 1x TBE buffer, 10 mM EDTA, 0.08% xylene cyanol, 0.08% amaranth) to the pellet. Vortex and spin three time. Samples were heated at 90°C for 2 min, centrifuge at 5000 RPM for 1 min. 4ul of sample was used to run on a 22% TBE-Urea polyacrylamide gels (UreaGel system, National Diagnostics) using Bio-Rad 21x40 cm gel casting cassettes. The gel was run at 35 W and until the amaranth dye reached 1 cm from the bottom of the gel. A gradient buffer system was used: upper reservoir, TBE; lower reservoir, TBE containing 0.3 M NaOAc. Autoradiography of gels was performed using storage phosphor screens and a Typhoon 9400 variable mode imager (GE Life Science) and quantified using ImageQuant software. Abortive transcriptions parameter were calculated as described in (Hsu, 1996, Hsu, 2009). Values shown in Figs. are the averages of three independent experiments.

Chapter 5

RNAP-CRE interaction modulate pausing and translocational bias during transcription elongation

[1]Hanif Vahedian-Movahed, Irina O. Vvedenskayaa, Jermey Bird, Jared G. Knoblauch, Seth R. Goldman, Bryce E. Nickels and Richard H. Ebright

5.1. Background

5.1.1. Rationale for occurrence of RNAP-CRE interaction during transcription elongation.

RNAP-CRE interactions involve the RNAP core and the DNA transcription bubble, both of which are present at each stage of transcription initiation, elongation and termination. As such, RNAP-CRE interactions may occur at each stage of transcription to potentially play functional roles not only at initiation but also in post-initiation stage of transcription including transcription elongation and transcription termination. This

[1] Author contributions: I obtained the initial evidence showing CRE modulate pausing (pause assays with class I and class II pause sequences). I also performed the translocation bias assays showing RNAP-CRE interactions stabilizes post-translocation. IOV performed NET-seq experiments and I assisted her by generating reagents. JB performed in vitro pause assays from sequences identified in NET-Seq and I assisted him by generating reagents. Additionally, I performed complementary scaffold-based pause assays to assess the effect of substitutions at +1, -1 and -10. SRG generated reagents for NET-Seq assays and assisted in data analysis. JGK also assisted in data analysis. Part of this chapter appears as published:

*Vvedenskaya IO, *Vahedian-Movahed H, *Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE (2014) Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. Science 344: 1285-9.

proposal is striking as until very recently the prevailing view has been that the sequence-specific RNAP-DNA interactions are mainly limited to transcription initiation, except with respect to sigma-factor retention during the initiation-elongation transition stage (Mukhopadhyay et al., 2001, Nickels et al., 2004) or when sigma-factor (as a trans regulatory element) associates with an elongating RNAP core (Goldman et al., 2015). In such cases, the sequence-specific interactions during the elongation stage (which is mediated by a sigma-factor and not the RNAP core) can cause σ-dependent pausing, which would have important functional significance (e.g., in recruiting a transcription factor to the elongation complex) (Strobel & Roberts, 2015). This assumption have been challenged by our observations which shows core subunit makes sequence-specific interaction with DNA of transcription bubble (see chapter one). In principle, the sequence-specific RNAP-CRE interactions delineated in the crystal structure and functional studies of RPo also may occur in RDe (Fig. 1). More than thirty crystal structures of bacterial and eukaryotic RDe have been determined to date. However, most contain nucleic-acid scaffolds that do not include transcription-bubble nontemplate-strand nucleotides and thus cannot provide information about RNAP-CRE interactions, and, of the minority that do include transcription-bubble nontemplate-strand nucleotides, most do not contain G at CRE position +2 and thus cannot provide information about sequence-specific RNAP-CRE interactions. Only one crystal structure of RDe has G at CRE position +2: namely, PDB 3PO2, a structure of a yeast RNAP II backtracked and arrested RDe (Cheung & Cramer, 2011, Zhang et al., 2012). Strikingly, in this structure, the G at CRE position +2 is inserted into a pocket formed by RNAPII amino acids residues equivalent to βR151, βI445, βD446, βR451, βT539, and βV547, adopting the same conformation, and making the same

interactions, as for G at CRE position +2 in the crystal structure of RPo (Fig. 1). This observation, together with other observations (Kireeva, Domecq et al., 2011), implies that the RNAP-CRE interactions that mediate recognition of G at CRE position +2 in RPo also can occur in RDe (at least in a backtracked, arrested RDe). This observation further implies that these RNAP-CRE interactions can be made not only by bacterial RNAP but also by eukaryotic RNAP II (I will cover this in chapter 6). In principle, sequence-specific RNAP-CRE interactions during elongation could play functionally important roles during transcription elongation, potentially affecting sequence-dependent translocational bias (the probability of RNAP translocation between pre-translocated and post-translocated states) and sequence-dependent pausing.

5.1.2. Translocation bias and rationale for role of CRE in translocation bias.

During transcription elongation, after NTP binding, catalysis and pyrophosphate release, the nucleotide at 3'-end of the nascent RNA occupies the active site i+1 site (pre-translocation state). In order to add the next NTP, the RNAP translocate relative to the DNA, where the nucleotide at 3'-end of the nascent RNA occupies the active site i-site, and therefore the i+1-site is available for the incoming NTP (post-translocation state) (Fig. 2) There are two model for translocation: 1) power stroke models, which postulates a power stroke tightly coupled to pyrophosphate release drives motion; 2) Brownian ratchet models, which suggest reversible diffusion of the enzyme along the DNA template between its pre- and post-translocated states is directionally rectified through the binding of the incoming NTP. The latter model is the acceptable model in the field. Based on this model prior to the addition of each nucleotide, RNAP oscillates between the pre- and post-translocation states, and at this stage intrinsic or extrinsic factors could bias probability of the presence

of one registers over the other. The translocation bias effects are fundamental to transcription as the fraction of times that each states are available could affect NTP binding, pyrophosphate binding, entrance to pause states, arrest state and termination. Since RNAP-CRE interactions are mediated by core subunit and transcription bubble and both of each are present in all stages of elongation and termination, we predict that RNAP-CRE interactions could occur at each of these stages and play functional roles by affecting the translocation bias between pre-translocation state and post-translocation state.

5.1.3. Transcriptional pausing and rational for the role of CRE in pausing.

Regulation of gene expression during transcription elongation often involves sequences in DNA that cause the transcription elongation complex (TEC) to pause. Pausing can affect gene expression by facilitating engagement of regulatory factors, influencing formation of RNA secondary structures, and enabling synchronization of transcription and translation. Several lines of evidence suggest that pausing involves specific sequence signals that inhibit nucleotide addition (Aivazashvili, Bibilashvili et al., 1981, Chan & Landick, 1993, Dangkulwanich, Ishibashi et al., 2014b, Hein, Palangat et al., 2011, Herbert, La Porta et al., 2006b, Kireeva & Kashlev, 2009, Kyzer, Ha et al., 2007b, Landick, 2009, Larson, Landick et al., 2011, Neuman, Abbondanzieri et al., 2003b, Weixlbaumer, Leon et al., 2013). Sequence-dependent transcriptional pauses include "Class I" pauses (RNA-hairpin-stabilized pauses), "Class II" pauses (backtracking-stabilized pauses), and "ubiquitous"/"elemental" pauses (unstabilized pauses) (Artsimovitch & Landick, 2000a, Dalal et al., 2006, Herbert et al., 2006a, Herbert, Greenleaf et al., 2008, Larson et al., 2011, Neuman et al., 2003a). All three classes are proposed to arise from the same sequence-dependent intermediate. We note that the two prototypical examples of Class I pauses (his

and trp; (Artsimovitch & Landick, 2000a, Lee, Phung et al., 1990)), the two prototypical examples of Class II pauses (pheP ops and rfaQ ops; (Artsimovitch & Landick, 2000a)), and the two highest-efficiency, longest-lifetime ubiquitous/elemental pauses characterized to date ("a" and "d"; (Herbert et al., 2006a)) all contain a nontemplate-strand G at the position immediately downstream of the pause site. We note further that, for the his and trp pauses, the nontemplate-strand G immediately downstream of the pause site has been shown to be important for pausing (Lee et al., 1990). Moreover, at the time unpublished data from Jonathan Weissman lab and our lab, identified a pause consensus sequence in *Escherichia coli* cells using native elongating transcript sequencing (NET-Seq) (fig. 3) (Larson, Mooney et al., 2014, Vvedenskaya et al., 2014). The occupancy of the TEC at a given position is correlated with the tendency of the TEC to pause at the position. Accordingly, NET-seq analysis enables identification of pause sites (Churchman & Weissman, 2011). Strikingly, again, the identified pause consensus sequence contained a nontemplate-strand G at the position immediately downstream of the pause site. We hypothesized that, in the proposed common intermediate for factor-independent, sequence-dependent pausing (Artsimovitch & Landick, 2000a, Dalal et al., 2006, Herbert et al., 2006a, Herbert et al., 2008, Larson et al., 2011, Neuman et al., 2003a), the nontemplate-strand G immediately downstream of the pause site corresponds to, and is recognized as, G at CRE position +2. Furthermore, we expect that this sequence-specific interaction could play functional roles in pausing.

5.2. Results and discussions

5.2.1. In vitro preliminary analysis of the effect of RNAP-CRE interactions on transcriptional pausing at class I and class II pause site.

To test the hypothesis that sequence-specific RNAP-CRE interactions play functionally important roles in pausing--recognizing the conserved G immediately downstream of the pause site, I carried out in vitro pause assays with WT-RNAP and βD446A-RNAP defecting in RNAP-G_{CRE} with a DNA template carrying a class I pause sequence (his pause) and two DNA templates carrying two of the class II pause sequence (*rfaq* Ops and *phep* Ops). My data showed that RNAP-CRE interactions modulate the pause capture efficiency for the three tested pause sequences (methods as in (Artsimovitch & Landick, 2000a, Kyzer, Ha et al., 2007a)). My data in Fig. 2 show that disruption of RNA-CRE interaction increase the pause capture efficiency at class I pause sites (RNA-hairpin-stabilized pauses) and decrease the pause capture efficiency at a class II pause sites "Class II" pause (backtracking-stabilized pauses). In other words, RNAP-CRE interactions naturally decrease pausing at class I pause site and increase pausing at class II pause site. This can be understood in terms of the stage at which RNAP-G_{CRE} interactions are formed and the translocation state at each class of pause. In transcription initiation complexes, RNAP core enzyme makes sequence specific interactions with G corresponding to active site i+1 (G_{CRE}). A transcription elongation complex at a pause site, TEC in post-translocated state will contain an unpaired nontemplate-strand G at downstream end of transcription bubble. An elongation complex at a consensus PE in a post-translocated state, but not in a pre-translocated state, is positioned to make favorable RNAP-CRE interactions. Therefore, RNAP-CRE interaction is expected to modulate pausing at a consensus PE: should decrease pausing if the paused state is pre-translocated or back-tracked should increase pausing if the paused state is post-translocated (Fig. 2). The current models predicts that the class I RNA-hairpin-stabilized pauses are caused by conditions that favor

hypertranslocation state while class II backtracking-stabilized pauses are caused by conditions that favor reverse-translocated states. Therefore we propose that possibly the conserved G immediately downstream of the pause site is recognized as G_{CRE} at i+1 NT and this interaction stabilizing the post-translocation state. This can explain our observations of RNAP-CRE interactions decreasing pausing at class I pause site and increasing pausing at class II pause site. I will show data supporting this hypothesis at the end of this chapter.

5.2.2. Genome-wide analysis of the effect of RNAP-CRE interactions on transcriptional pausing.

In vivo NET-Seq analysis. Given the preliminary results that showed RNAP-CRE interactions modulate transcription pausing, and the availability of technologies to study pausing at genome-wide scale in vivo we sought to investigate this phenomenon at genome-wide scale in live *E. coli* cells. As the first step, we sought to define the key sequence determinant of pause site, using NET-Seq.

To perform NET-seq in Escherichia coli, cells carrying a chromosomal rpoC-3xFLAG gene, encoding RNAP b′ subunit with a C-terminal 3xFLAG tag were grown to midexponential phase; cells were flash-frozen and lysed; 3xFLAG-tagged TECs were immunoprecipitated with an antibody against FLAG; RNAs were extracted from TECs; and RNA 3′ ends were converted to cDNAs and analyzed using high-throughput sequencing. We defined pause sites as positions where TEC occupancy exceeded TEC occupancy at each position 25 base pairs (bp) upstream and downstream. We identified 15,553 pause sites, which corresponds to ~19,800 total pause sites, given the estimated ~78% saturation of the analysis (tables S1 to S7). Alignment of pause-site sequences

revealed a clear consensus pause element (PE): G–10Y–1G+1, where position –1 corresponds to the position of the RNA 3′ end (Fig. 1A and fig. S2). Of the identified pause sites, ~35% exhibited a 3-of-3 match to the consensus PE, and ~42% exhibited a 2-of-3 match to the consensus PE (tables S4 and S5). Comparing the total number of pause sites with a 3-of-3 match (~6900) to the total number of sequences in the transcribed portion of the genome with a 3-of-3 match (~43,500 to 58,000) (Haas, Chin et al., 2012) indicates that, under these conditions, functional pausing occurs at ~12 to 16% of sequences with a 3-of-3 match.

In vitro pause analysis. To validate the NET-seq results, we selected two consensus PEs for analysis in vitro. One, located in yrbL, has –1C, and the other, located in gltP, has –1T (Fig. 5). Transcription assays show that the yrbL PE and gltP PE induce pausing in vitro (Fig. 5) and that consensus base pairs at positions +1, –1, and –10 are required for efficient pause capture in vitro. Next, Jeremy Bird and I performed experiments to determine the contribution of individual bases of the consensus PE to pausing at the yrbL PE (6). Jeremy Bird used a template which include a promoter site, halt-site and pause site. In parallel I performed analogous experiments but using a reconstituted elongation complex halted 1bp upstream of the pause site which allow quantification of the effect of CRE on pausing while reducing the effect contributed by the presence of sigma or sequences prior to the pause site, and allow analysis of hetero-duplex mutations. In both assays the results indicate that introduction of each of the three nonconsensus base pairs at position +1 (C, A, and T) reduces pausing; introduction of each of the two nonconsensus base pairs at position –1 (A and G) reduces pausing; and introduction of two of three nonconsensus base pairs at position –10 (A and T) reduces pausing (Fig. 5). In each case,

the effect on pausing is manifest at the level of pause capture efficiency. The consensus PE comprises sequence determinants that previously have been implicated as important for pausing: G at +1 (Herbert et al., 2006b, Lee et al., 1990), T or C at –1 (Aivazashvili et al., 1981, Chan & Landick, 1993, Hein et al., 2011, Herbert et al., 2006b), and G at –10 (Herbert et al., 2006b, Kyzer et al., 2007a). The consensus PE is especially similar to a consensus pause–inducing sequence identified by Herbert et al. in single-molecule studies (Herbert et al., 2006b).

Hypothesis for the mechanism of pausing. The sequence determinants in the consensus PE (G–10Y–1G+1) can be understood in terms of the structure and mechanism of the TEC (Fig. 7). In each nucleotide-addition cycle in transcription elongation, RNAP translocates between a "pretranslocated state," in which the RNAP active center "i" and "i + 1" sites interact with RNA positions –2 and –1, and a "posttranslocated state," in which the RNAP i site interacts with RNA position –1 and the RNAP i + 1 site is unoccupied and available for binding of a nucleoside triphosphate (NTP) (Dangkulwanich et al., 2014b, Dangkulwanich, Ishibashi et al., 2013, Larson et al., 2011). Translocation requires breaking the DNA base pair at position +1 and breaking the RNA-DNA base pair at position –10 (Bai, Shundrovsky et al., 2004, Tadigotla, D et al., 2006). Because the DNA base pair at position +1 must be broken for forward translocation, G/C at +1 will disfavor forward translocation relative to the less stable A/T. Similarly, because the RNA-DNA base pair at position –10 must be broken for forward translocation, G/C at –10 will disfavor forward translocation relative to the less stable A/T. Furthermore, because 5′-rGrN-3′/3′-dC-dN-5′ is more stable than 5′-rC-rN-3′/3′-dG-dN-5′ (Sugimoto, Nakano et al., 1995), rG:dC at position –10 also will disfavor translocation over rC:dG at –10. Translocation requires

movement of the template-strand DNA nucleotide and base-paired RNA nucleotide at position –1 from the RNAP active center i + 1 site to the i site and movement of the template-strand DNA nucleotide at position +1 to the i + 1 site. Because available evidence indicates that the RNAP active center i and i + 1 sites preferentially interact with 5′-rR rY-3′/3′-dY-dR-5′(3), a nontemplate-strand Y at position –1 and nontemplate-strand G at position +1 will disfavor forward translocation relative to all other sequences (the former by stabilizing the pretranslocated state; the latter by destabilizing the posttranslocated state). Thus, each position of the consensus sequence is predicted to favor the pretranslocated state over the posttranslocated state (–10 through effects on duplex stability, –1 through effects on active-center binding, and +1 through both). Accordingly, each position of the consensus PE would be predicted to increase the opportunity for the TEC to enter an "elemental pause" state (a state that, according to one view, is accessed from the pretranslocated state and serves as an obligatory intermediate for pausing) (Churchman & Weissman, 2011, Herbert et al., 2006b, Landick, 2009, Neuman et al., 2003a, Weixlbaumer et al., 2013, Zhou, Ha et al., 2011) or a "backtracked" state (a state that, according to another view, is accessed from the pretranslocated state and serves as the primary state for pausing) (Dangkulwanich et al., 2014b, Dangkulwanich et al., 2013).

In vivo mNET-Seq analysis of the effect of RNAP-CRE interactions. A TEC in a posttranslocated state at a PE will contain an unpaired G at the downstream end of the nontemplate strand of the unwound "transcription bubble" (Fig. 1). In transcription initiation complexes, RNAP core enzyme makes sequence-specific interactions with an unpaired G at the downstream end of the transcription bubble ("core recognition element" CRE) (Zhang et al., 2012) (Fig. 3A). In transcription initiation, interaction between RNAP

and the unpaired G of the CRE (G_{CRE}) facilitates promoter unwinding to form a stable, catalytically competent, RNAP-promoter open complex (RPo) (see chapter one). It has been proposed that RNAP-G_{CRE} interaction may be functionally important in transcription elongation as well as in transcription initiation (Zhang et al., 2012), but this has not been previously documented. The observation that the consensus PE contains the sequence feature required for establishment of RNAP-G_{CRE} interaction raises the possibility that RNAP-G_{CRE} interaction may mediate or modulate pausing. To test this possibility, we constructed and analyzed a mutant RNAP defective in sequence-specific recognition of G_{CRE}. The crystal structure of RPo shows that RNAP βD446 makes H bonds with Watson-Crick atoms of G_{CRE} and suggests that D446 reads the identity of G_{CRE} (Zhang et al., 2012). Replacement of βD446 by alanine results in the loss of the ability to distinguish G, A, T, or an abasic site at the position corresponding to G_{CRE} (see chapter one), confirming that βD446 reads the identity of G_{CRE} and providing a reagent to assess functional significance of sequence specific RNAP-G_{CRE} interactions.

To establish whether RNAP-CRE interactions affect pausing, we used a variant of NET-seq, "merodiploid NET-seq" (mNET-seq), that enables analysis of mutant RNAP derivatives, including mutant RNAP derivatives that do not support viability in haploid (fig. S1). mNET-seq uses merodiploid cells containing a plasmid-encoded, epitope tagged RNAP and a chromosome-encoded, untagged RNAP, and it involves selective analysis of transcripts associated with epitope-tagged RNAP in the presence of a mixed population of epitope tagged RNAP and untagged RNAP. We introduced into cells a plasmid encoding 3xFLAG-tagged wildtype RNAP β subunit (βWT) or 3xFLAG-taggedRNAP β subunit containing D446A (βD446A), we isolated RNAs associated with RNAP-βWT or RNAP

βD446A by immunoprecipitation, and we identified pause sites. For RNAP-βWT, alignment of pause sites revealed a consensus sequence matching the consensus PE, which validated mNET-seq as an effective system for analysis of pausing (Figs. 1A and 3C; figs.S2 and S6; and tables S4 to S9). For RNAP-βD446A we identified ~60 to 90% more pause sites than with RNAP-βWT (tables S4 to S10). Alignment of the pause sites revealed that ~30% more of the pause sites carried a G at position +1 (Fig. 3C, fig. S6, and table S6). We conclude that RNAP-βD446A is more susceptible than RNAP-βWT to pausing at sites with G at position +1.

In vitro analysis of the effect of RNAP-CRE interactions. We next compared pausing properties of RNAP-bD446A and RNAP-βWT in vitro, using templates carrying the yrbL PE and gltP PE (Fig. 7). RNAP-βD446A enhances pausing at yrbL PE and gltP PE. RNAP-βD446A also enhances pausing at other positions where the next nucleotide to be added to the transcript is G (positions with asterisks in Fig. 7). The results indicate that a substitution that disrupts sequence-specific RNAP-CRE interaction increases pausing, both in vivo and in vitro, at positions where the posttranslocated state contains G_{CRE}. We conclude that sequence-specific RNAP-G_{CRE} interaction occurs during elongation and counteracts pausing.

Analysis of the effect of RNAP-CRE interactions on translocation bias. To explore why disruption of sequence-specific RNAP-G_{CRE} interaction enhances pausing, we assessed whether RNAP-G_{CRE} interactions affect the translocational register of the TEC by assessing sensitivity of TECs to pyrophosphorolysis (Fig. 9). Sensitivity to pyrophosphorolysis provides a measure of TEC translocation because a TEC in a pretranslocated state is sensitive to pyrophosphorolysis but a TEC in a posttranslocated

state is resistant (Fig. 4A) (3). We performed assays with RNAP-βWT or RNAP-βD446A on templates containing G or T at position +1 (Fig. 4B).We found that TECs with RNAP-βWT were ~5 times as sensitive to pyrophosphorolysis when the template contained +1G as when the template contained +1T, which indicated that a greater proportion of TECs on templates containing +1G were in a pretranslocated state than of TECs on templates containing +1T (Fig. 4C). The results directly demonstrate the effect of G at position +1 on translocation bias. TECs with RNAP-βD446A were~4 times as sensitive to pyrophosphorolysis as TECs with RNAP-βWT when the template contained +1G (Fig. 4C), whereas, in contrast, TECs with RNAP-βD446A and RNAP-βWT exhibited identical sensitivities to pyrophosphorolysis when the template contained +1T (Fig. 4C). The results indicate that, on templates containing +1G, a greater proportion of TECs with RNAP-βD446A are in a pretranslocated state as compared to TECs with RNAP-βWT. We conclude that sequence specific RNAP-G_{CRE} interactions stabilize the TEC posttranslocated state, which provides a mechanistic explanation for the finding that RNAP-G_{CRE} interactions counteract pausing.

Discussion. Our findings define the key sequence determinants of transcriptional pausing as G–10Y–1G+1 (Figs. 1 and 2). The consensus PE promotes pausing by disfavoring translocation of the TEC to the posttranslocated state (Fig. 2B), which increases the opportunity for the TEC to enter an "elemental pause" state and/or a "backtracked" state. We further show that sequence-specific RNAP-G_{CRE} interactions counteract pausing by stabilizing the TEC in a posttranslocated state (Figs.3 and 4). Because residues of RNAP core that mediate sequence-specific RNAP-CRE interaction are conserved in RNAP from all living organisms, we suggest that RNAP-CRE interaction

counteracts pausing in all multisubunit RNAPs (see chapter 6). The consensus PE will, on average, be encountered by RNAP every ~32 bp during transcription elongation for organisms, such as E. coli, with ~50% G/C content and will be encountered even more frequently for organisms with higher G/C content. RNAP-G_{CRE} interactions may help overcome a barrier to forward translocation that occurs each time RNAP encounters a G-10Y-1G+1 sequence during transcription elongation. A major function of RNAP-GCRE interactions may be to suppress noise during transcription elongation by smoothing the sequence dependent energy landscape for transcription elongation.

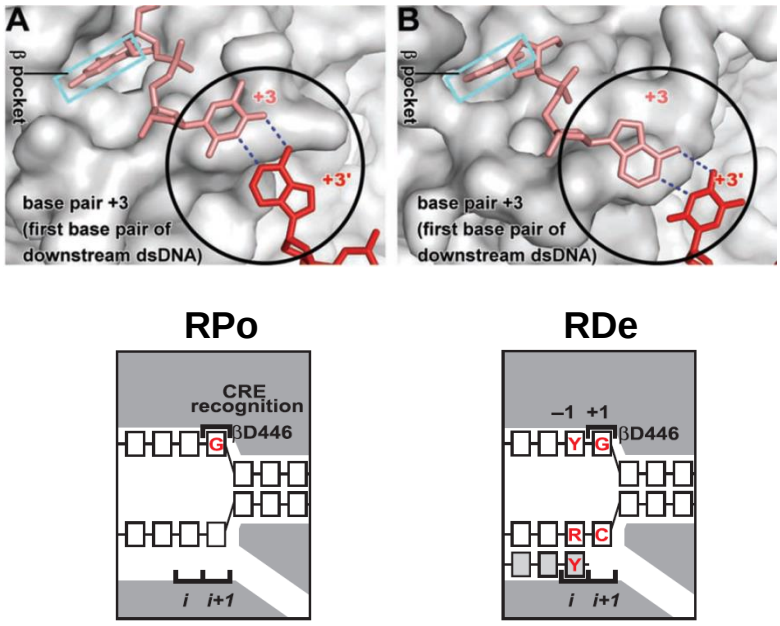


Figure 1. RNAP-CRE interactions in initiation and elongation. (A) the crystal structure of RPo-GpA. In transcription initiation complexes, RNAP core enzyme makes sequence specific interactions with G corresponding to active site i+1. (B) a crystal structure of a yeast RNAP II backtracked and arrested RDe (PDB 3PO2; 140). RDe in post-translocated state will contain an unpaired nontemplate-strand G at downstream end of transcription bubble. The presence in each case of an unpaired G at downstream end of nontemplate strand of transcription bubble suggests the possibility of equivalent sequence-specific interactions between RNAP core and the G during elongation.

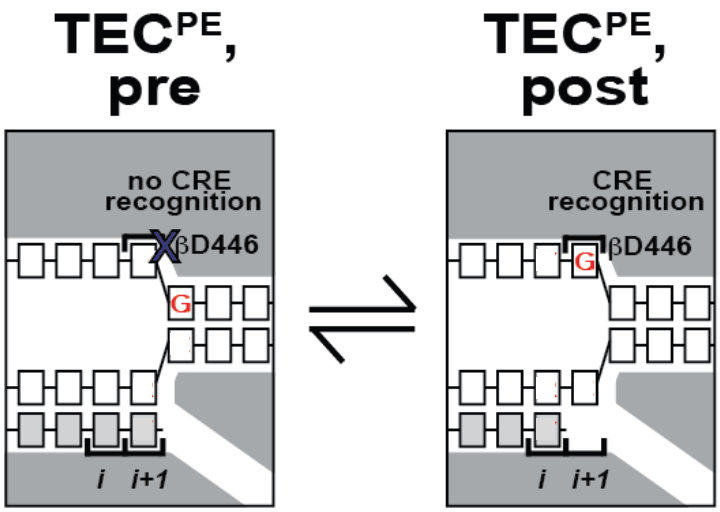


Figure 2. Model for RNAP-CRE interactions in the context of an elongation complex and hypothesis for the mechanism of pause modulation by RNAP-CRE interactions. An elongation complex at a consensus PE in a post-translocated state, but not in a pre-translocated state, is positioned to make favorable RNAP-CRE interactions. RNAP-CRE interaction is expected to modulate pausing at a consensus PE: should decrease pausing if the paused state is pre-translocated or back-tracked and should increase pausing if the paused state is post-translocated.

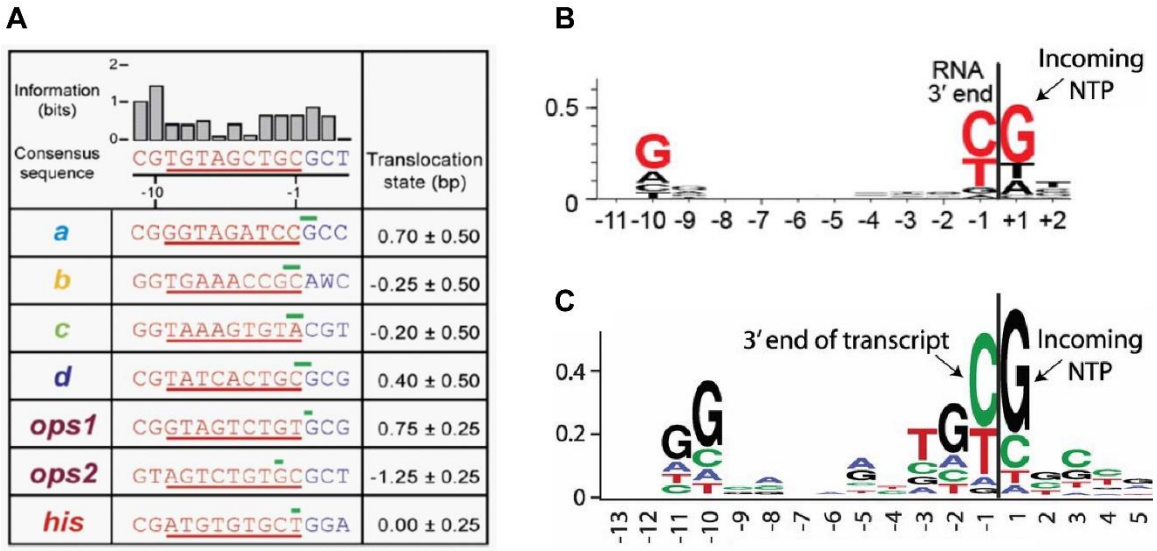


Figure 3. Identified pause sequences of bacteria contain a nontemplate-strand G at the position immediately downstream of the pause site. (A) a consensus pause sequence identified from in vitro single-molecule optical tweezer assay by Herbert et al., Cell 2006 (Steven Block lab). (B) at the time unpublished consensus pause sequences identified from genome-wide NET-Seq assay in E. coli cell by Vvedenskaya*Vahedian-Movahed*, Bird* et al., Science 2014 (our labs) and (C) by Larson et al, 2014 Science (Jonathan Weissman lab).

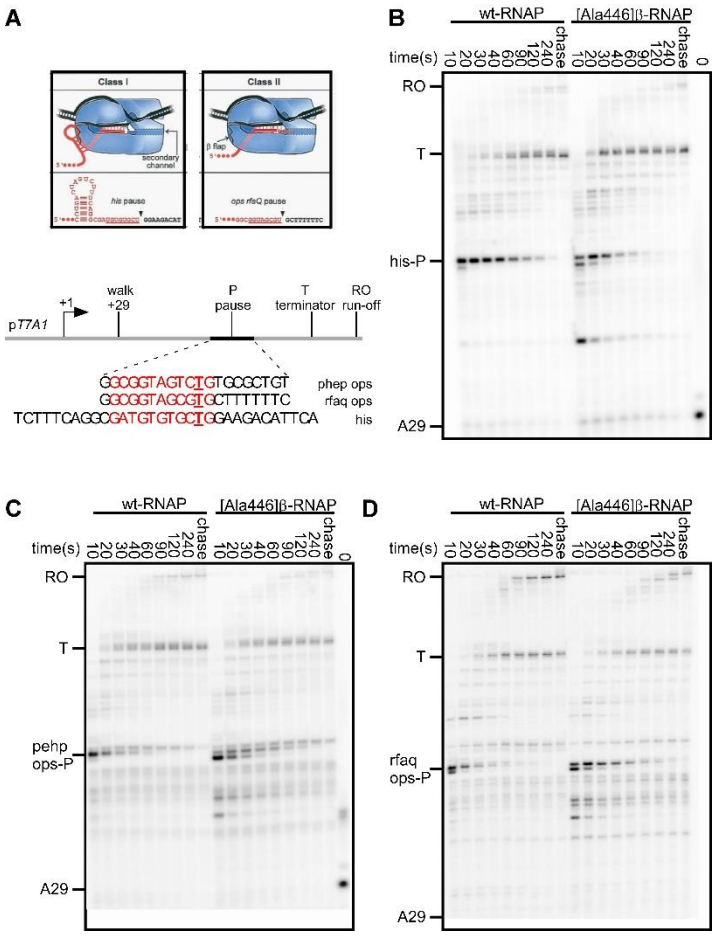


Figure 4. Effect of RNAP-CRE interactions on class I and class II stabilized pause sitese. (A) Schematic of the mechanism of class I and class II pause sites and sequences of the template used for pause assays. (B) Pause-capture efficiencies of RNAP β^{WT} and RNAP β^{D446A} at *his* pause. (C) Pause-capture efficiencies of RNAP β^{WT} and RNAP β^{D446A} at *phep ops* pause. (D) Pause-capture efficiencies of RNAP β^{WT} and RNAP β^{D446A} at *rfaq ops* pause.

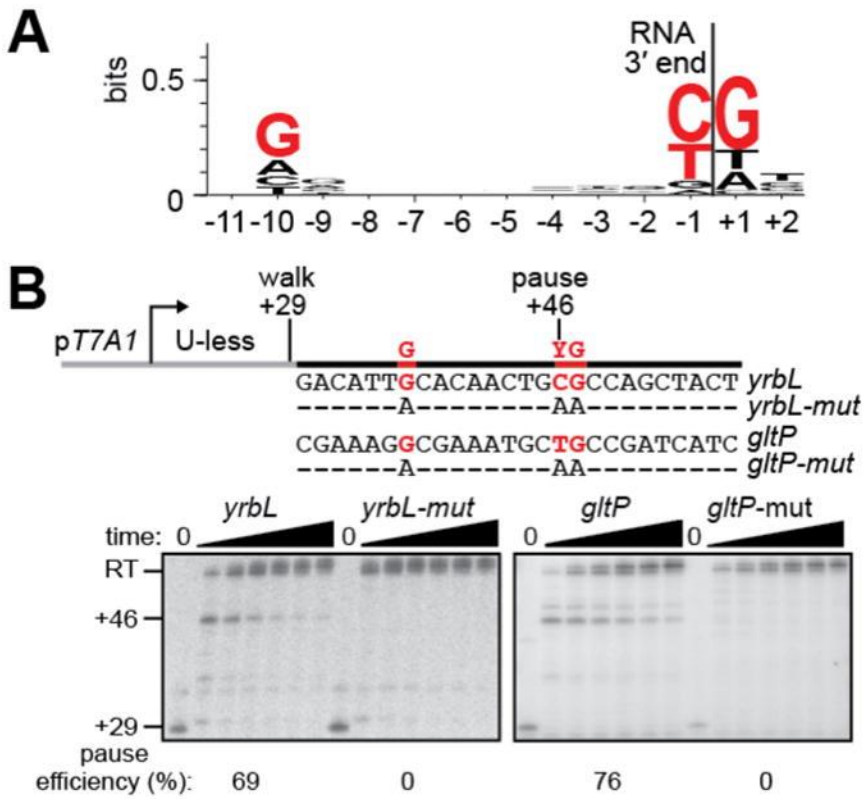


Figure 5. Identification of consensus PE. (A) Sequence logo for consensus PE from NET-seq. Red, bases with ≥0.2 bit sequence-information content. (B) In vitro transcription assays with consensus PEs and mutant PEs. Top, templates. Bottom, results. +29, RNA before addition of UTP; +46, RNA in TEC at PE; RT, read-through RNA; red, consensus PE bases.

.

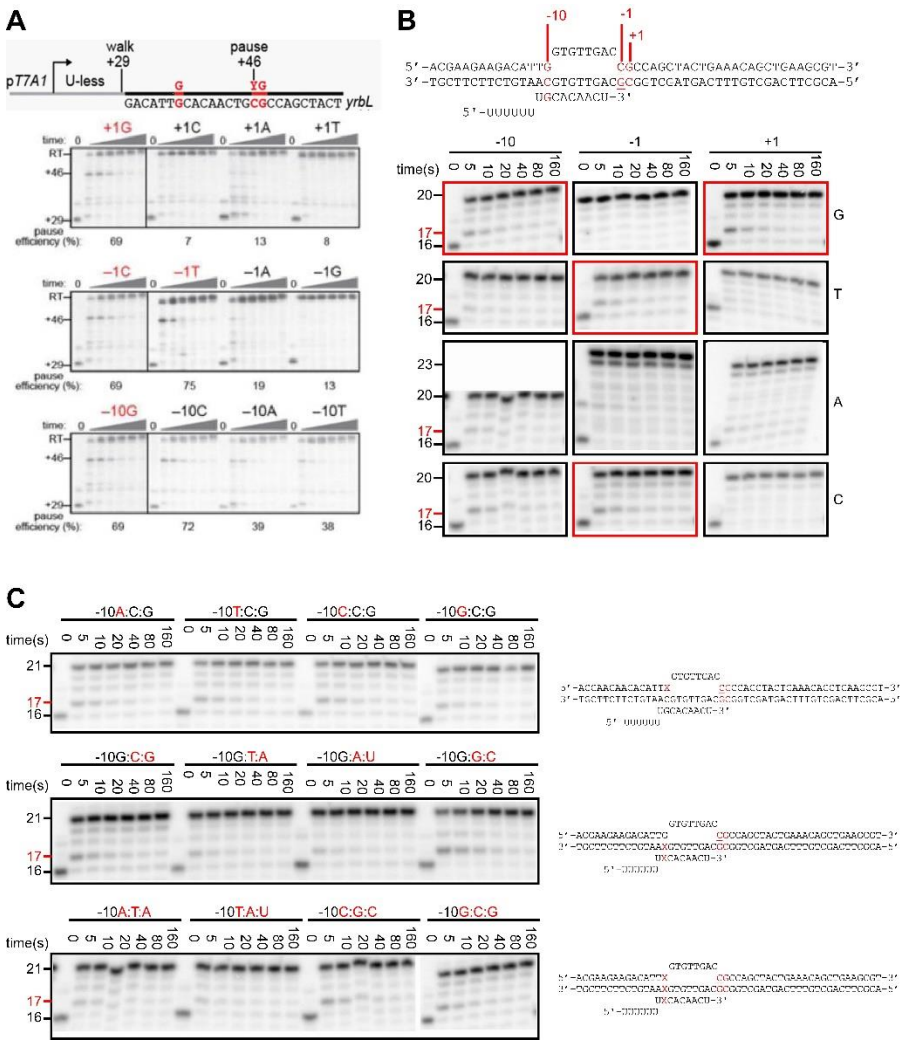


Figure 6. Contributions of individual base pairs of consensus PE. (A) *In vitro* transcription assays with yrbL PE derivatives with bb substitutions at posiions +1, -1 and -10. Red, consensus PE bases. (B) *In vitro* transcription assays with yrbL PE derivatives with bp substitutions at positions +1, -1 and -10 using a scaffold based-pause assays. (C) analysis of the contribution of base pairs of RNA:DNA hybrid and DNA: DNA using a scaffold based-pause assays.

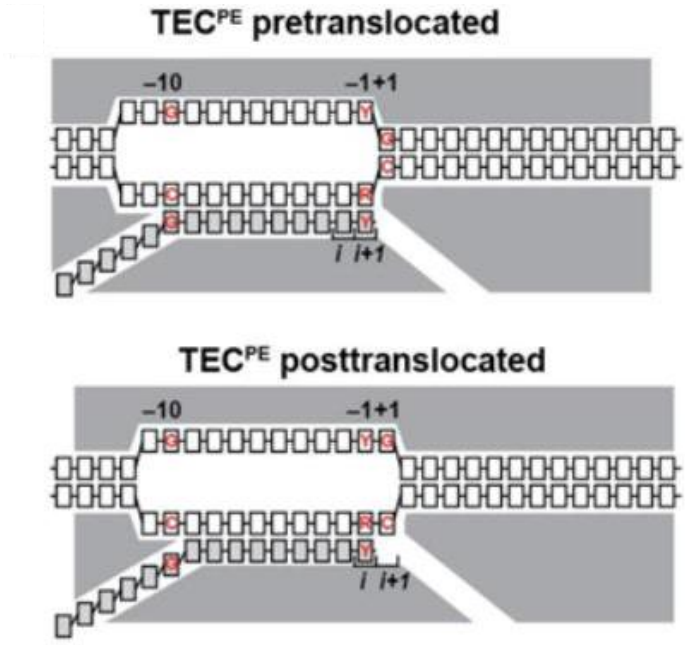


Figure 7. Hypothesis for how PE sequence induces pausing. Schematic representation of TEC in pre-translocated state at consensus PE (top) and TEC in post-translocated state at consensus PE (bottom). Consensus pause sequence is predicted to favor pre-translocation state over post-translocation state: first, at positions +1 and -10 G/C will disfavor forward translocation relative to the less stable A/T. second, At position -10, rG:dC also will disfavor translocation over rC:dG Nontemplate-strand Y at position –1 and nontemplate-strand. Third, G at position +1 will disfavor forward translocation relative to all other sequences. Because available evidence indicates that the RNAP active center i and i + 1 sites preferentially interact with 5′-rR-rY-3′/3′-dY-dR-5′. White boxes, DNA; gray boxes, RNA; gray shading, RNAP; red, consensus PE bases; i and i+1, RNAP active-center i and i+1 sites.

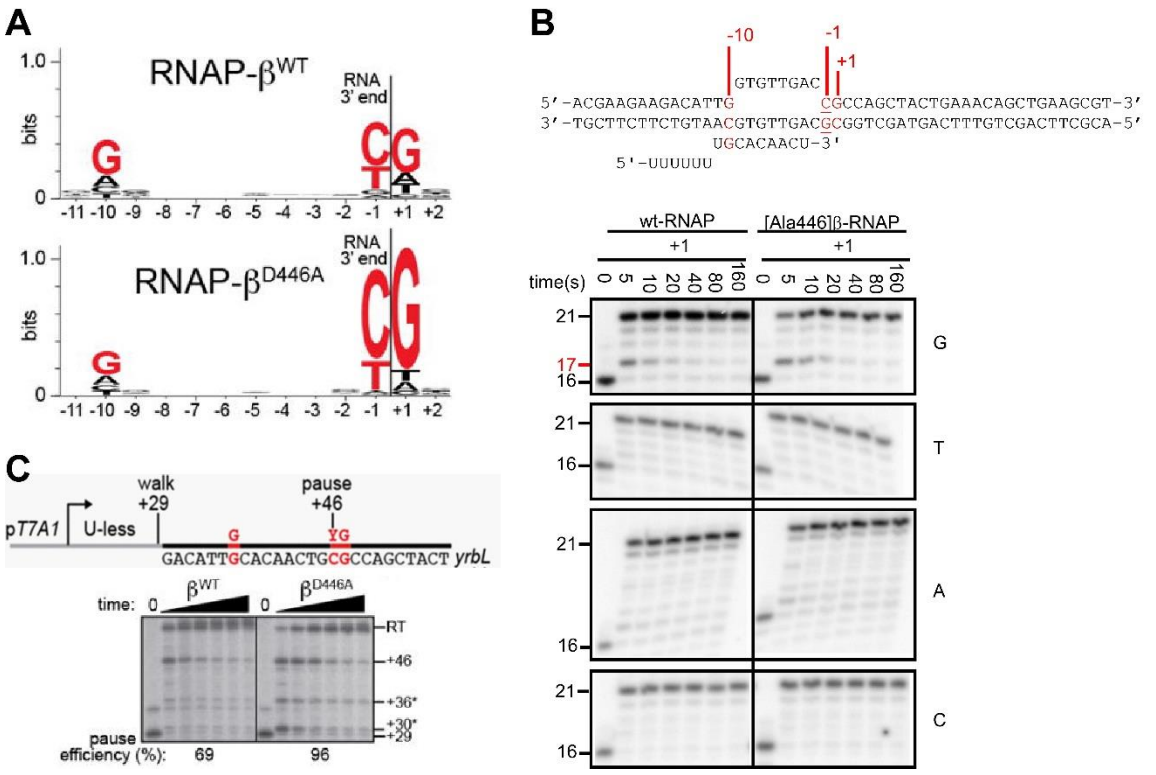


Figure 8. Sequence-specific RNAP-G$_{CRE}$ interactions modulate pausing. (A) Sequence logos for consensus PE with RNAP-β^{WT} (top; 0.4 bit sequence-information content at G$_{CRE}$) and RNAP-β^{D446A}(bottom; 1 bit of sequence-information content at G$_{CRE}$), as defined by mNET-seq. Red, bases with ≥0.2 bit sequence information content. (B) Pause-capture efficiencies of RNAP β^{WT} (left panels) and RNAP β^{D446A} (right panels) at *yrbL* PE. (C) Pause-capture efficiencies of RNAP β^{WT} (left panels) and RNAP β^{D446A} (right panels) at *yrbL* PE with all base substitutions at position +1 using scaffold-based pause assay.

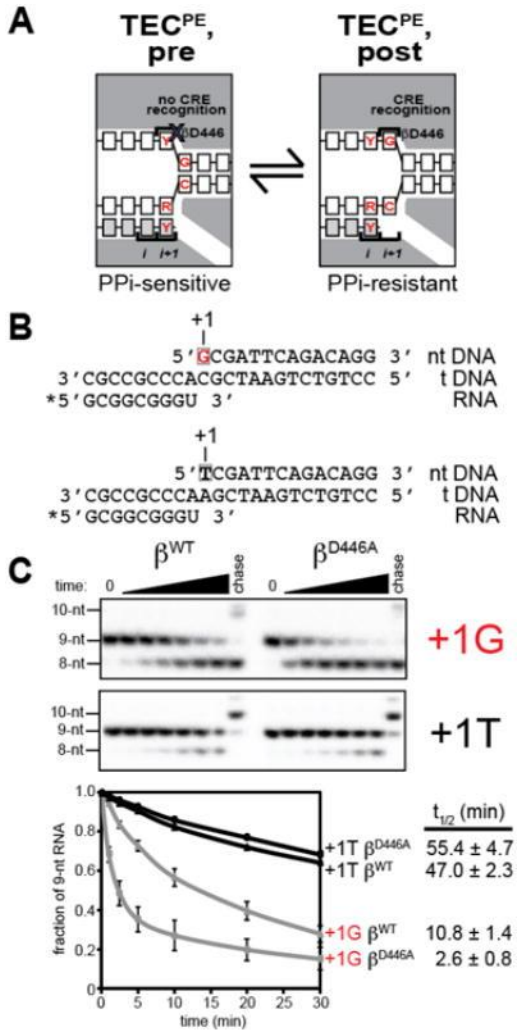


Figure 9. Sequence-specific RNAP-G_{CRE} interactions modulate translocation bias. (A) Structural organization of TEC in pre-translocated state at consensus PE (left; unfavorable RNAP-CRE interaction) and TEC in post-translocated state at consensus PE (right; favorable RNAP-CRE interaction). PPi, pyrophosphate. (B) Nucleic-acid scaffolds used for translocation-bias assays. Asterisk, radiolabel on RNA 5′ end; box, position corresponding to G_{CRE}; red, consensus PE base. (C) Translocational bias for RNAP β^{WT} and RNAP β^{D446A} on nucleic-acid scaffolds containing G or T at position corresponding to G_{CRE}. Gel images show pyrophosphorolysis reaction progress from 0-30 min. 9-nt, scaffold; 8-nt, product of pyrophosphorolysis; 10-nt, product of "chase" reaction with GTP (+1G template) or UTP (+1T template). Graph shows fraction of unaltered scaffold (mean±SEM; 3 measurements) as function of time.

5. 3. Material and methods

Strains. A complete list of strains is provided in Table S11. To construct strain SG110 (MG1655 *rpoC-3xFLAG*) PCR was performed on FLP-KanR-FLP cassette vector pKD4 using primers s861 [P1 primer from [Datsenko, 2000 #208] with 60 nt of homology to *rpoC* starting immediately upstream of the T in the TAA stop codon] and s862 [P2 primer from [Datsenko, 2000 #208] with 3xFLAG encoding sequence followed by 60 nt of homology to the genome starting 16 bp downstream of the A in the TAA stop codon of *rpoC*]. The PCR product generated in this reaction was transformed into electrocompetent BW25113 cells containing pKD46 as described *(Datsenko & Wanner, 2000)*. A kanamycin resistant, carbenicillin sensitive clone, SG99, was isolated. A P1 lysate prepared from SG99 was used to transduce the kanamycin resistance gene into MG1655 to generate SG101A. SG101A was confirmed to carry *rpoC-3xFLAG* by sequence analysis of a PCR product that encompassed the 3' region of *rpoC-3xFLAG* and by Western blot analysis using antibodies to FLAG (Sigma, mouse anti- FLAG M2) and β' (Neoclone) according to the manufacturer's suggested protocol for the Odyssey Imager (Licor). The kanamycin resistance cassette was removed from SG101A via introduction of plasmid pCP20 that expresses the FLP recombinase as described (Datsenko & Wanner, 2000) to create strain SG110.

Plasmids. Plasmids pRL706 (encoding βWT; (Severinov, Mooney et al., 1997)) and pRL706-βD446A (encoding βD446A; prepared from pRL706 by site-directed mutagenesis) were digested with StuI and XhoI. A DNA fragment corresponding to a Gly-Gly-Gly-Ser spacer followed by a 3x-FLAG sequence was generated by annealing 5’ phosphorylated oligos HV1 and HV2, and this fragment was ligated into StuI-XhoI

digested pRL706 or pRL706-βD446A to generate pRL706-βWT;3xFLAG or pRL706-βD446A;3xFLAG. pIA171 (Artsimovitch & Landick, 2000b), containing a T7A1 promoter followed by a 29 base pair U-less cassette, was a gift from Irina Artsimovitch.

TMR211-σ70. σ70 was prepared as in (Mukhopadhyay, Mekler et al., 2003a), but using plasmid pGEMD(-Cys)211C(Mekler et al., 2002). Yields were 4-60 mg/l, and purities were > 95%. A σ70 derivative labeled at position 211 with tetramethylrhodamine was prepared as in (Mekler et al., 2011a, Zhang et al., 2012).

RNAP. RNAP-βWT and RNAP-βD446A core and holoenzyme were prepared from *E. coli* strain XE54 (Tang et al., 1994) transformed with plasmids pRL706 or pRL706-βD446A. Protein purification procedures were essentially as in (Mukhopadhyay et al., 2003a). Purities were ≥ 95%. [TMR211-σ70]-RNAP. [TMR211-σ70]-RNAP-βWT and [TMR211-σ70]-RNAP-βD446A holoenzymes were prepared as described (Zhang et al., 2012). Wild-type or βD446A-containing RNAP core (3 μM) and TMR211-σ70 (12 μM) were equilibrated in 1.5 ml 20 mM Tris-HCl, pH 7.9, 200 mM NaCl, 0.1 mM EDTA, 5 mM 2-mercaptoethanol, and 5% glycerol for 2 h at 4°C. The reaction mixture was applied to a HiLoad 16/60 Superdex S200 column (GE Healthcare) equilibrated in 20 mM Tris-HCl (pH 7.9), 200 mM NaCl, 0.1 mM EDTA, 5 mM 2-mercaptoethanol, and 5% glycerol, and the column was eluted with 180 ml of the same buffer. RNAP holoenzyme derivatives typically eluted between 45 and 60 ml. Fractions containing RNAP holoenzyme were pooled, concentrated to ~3 mg/ml using 30 kDa MWCO Amicon Ultra-15 centrifugal ultrafilters (EMD Millipore), and stored in 10 mM Tris-HCl (pH 7.9), 100 mM NaCl, 0.05 mM EDTA, 2.5 mM 2-mercaptoethanol, and 50% glycerol at -80°C.

Oligonucleotides. Sequences of oligonucleotides used in this work are presented in Figs. 3B, 4B, S5A,C and Table S12.

Nucleic-acid scaffolds. Non-template oligodeoxyribonucleotides (ntDNA), template oligodeoxyribonucleotides (tDNA) and oligoribonucleotide (RNA) were resuspended in RNase-free water (Ambion). Nucleic-acid scaffolds for fluorescence-detected RNAP-DNA interaction assays were prepared essentially as in (Zhang et al., 2012). Nontemplate-strand oligonucleotide (0.5 mM) and template-strand oligonucleotide (0.55 mM) were combined in 25 μl 5 mM Tris-HCl (pH 7.7), 200 mM NaCl, and 10 mM MgCl2, heated 2min at 95°C, and cooled to 25°C in 2°C steps with 60 s per step in a thermal cycler (Applied Biosystems). Nucleic-acid scaffolds for translocational-bias assays were prepared essentially as in (Hein et al., 2011). Nontemplatestrand oligonucleotide (1 μM), template-strand oligodeoxyribonucleotide (1 μM), and 5'-32P-labelled RNA oligonucleotide [0.5 μM; 5'-32P-labelled using T4 polynucleotide kinase (New England Biolabs) and [γ32P]-ATP (PerkinElmer)] were combined in 50 μl 10 mM Tris-HCl (pH 7.9), 40 mM KCl, and 5 mM MgCl2, heated 2 min at 95°C, cooled to 45°C in 30 s, and cooled to 25°C in 2°C steps with 120 s per step, in a thermal cycler (Applied Biosystems).

Cell growth for NET-seq. Biological replicates of SG110 (MG1655 *rpoC-3xFLAG*) cells were shaken at 220 RPM at 37°C in 30 ml of LB [10 g Bacto tryptone (Becton Dickinson & Co), 5 g Bacto yeast extract (Becton Dickinson & Co) and 10 g NaCl per liter] in 125 ml DeLong flasks (Bellco) to OD600 ~0.6. Cell suspensions were divided equally among 12 x 2 ml tubes (BioExcell) and centrifuged (1 min, 21,000 x g at room temperature) to collect cells, cell pellets were washed by addition of 0.5 ml of TBS (50 mM Tris HCl pH

7.5, 150 mM NaCl), centrifuged (30 sec, 21,000 x g at room temperature) and supernatants were removed. Cell pellets were then rapidly frozen on dry ice and stored at –80°C.

Cell growth for mNET-seq. Biological replicates of MG1655 cells containing plasmids pRL706-βWT;3xFLAG or pRL706-βD446A;3xFLAG were shaken at 220 RPM at 37°C in 100 ml 4x LB (40 g Bactotryptone, 20 g Bacto yeast extract and 10 g NaCl per liter) containing 200 μg/μL carbenicillin in 500 ml DeLong flasks (Bellco). When cell density reached an OD600 ~0.6, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added, and cells were grown for an additional 4 h. Cells were harvested and stored as described above. Immunoblotting of cell extracts indicated that the amounts of 3xFLAG-βWT and 3xFLAG-βD446A were similar in all replicates (Fig. S8).

NET-seq and mNET-seq: affinity purification of RNAP. All manipulations were performed at 4°C. Cells pellets derived from 12 ml of culture were resuspended in 1 ml of lysis buffer [B-Per, Bacterial Protein Extraction Reagent (Thermo Scientific)] supplemented with one quarter of a protease inhibitor cocktail tablet [complete Mini EDTA-free (Roche)], 1 mM EDTA, 80U Murine RNase Inhibitor (NEB), 100 μg lysozyme (Thermo Scientific), 150U DNase I (Thermo Scientific) and incubated for 10 min. The lysate was then clarified by centrifugation (10 min, 21,000 x g) and NaCl was added to a final concentration of 150 mM. A small portion of the lysate (50 μl) was removed for isolation of total RNA. The remainder of the lysate was added to 1 ml of anti-FLAG M2 affinity gel (Sigma Aldrich) that had been washed three times with 3 ml of TBS and equilibrated in 3 ml of Wash buffer [B-Per solution containing 150 mM NaCl, 1 mM EDTA, 50U/ml Murine RNase Inhibitor, and protease inhibitor cocktail (Complete EDTAfree, Roche; 1 tablet per 50 ml). The lysate and affinity gel mixture was incubated

2.5 h at 4°C in a 1.7 ml centrifuge tube with gentle rocking. The mixture was transferred to a 10 ml Econo-Pack disposable chromatography column (Bio-Rad), the flow through was collected, the affinity gel was washed 8 times with 5 ml Wash buffer and 3 times with 250 μL of Elution buffer [B-Per solution containing 150 mM NaCl, 1 mM EDTA, 50U/ml Murine RNase Inhibitor and 2 mg/ml 3xFLAG peptide (GenScript)]. For the washes with Elution buffer the affinity gel was incubated for 30 min prior to collection of the fractions. The presence of epitope tagged β′, βWT or βD446A was analyzed in each fraction by immunoblotting (Figure S9). This analysis indicated that > 90% of epitope tagged β′ and βWT and > 70% of epitope tagged βD446A bound to the affinity gel.

NET-seq and mNET-seq: RNA isolation and purification. All manipulations with RNAs were performed at 4°C unless otherwise specified and nuclease-free water (Life Technologies) was used for all solutions. To isolate total RNA, 250 μl of TRI Reagent solution (Molecular Research Center) was added to 50 μl of clarified lysate. Samples were incubated at 70°C for 10 min, centrifuged (10 min, 21,000 x g) to remove insoluble material. The supernatant was transferred to a fresh tube, ethanol was added to a final concentration of 60.5%, and the mixture was applied to a Direct-zol spin column (Zymo Research). DNase I treatment was performed on-column according to the manufacturers recommendations. RNA was eluted from the column with 3 sequential portions of 30 μl nuclease-free water that had been heated to 70°C. Typical yields of total RNA from 50 μl of clarified total lysate were 50 to 60 μg. To isolate RNAs associated with RNAP, pooled eluates from above were mixed with 3 volumes of TRI Reagent solution and processed as described above. The RNAs associated with RNAP constituted 0.4-0.6% of the total RNA.

NET-seq and mNET-seq: cDNA library construction. Prior to cDNA library construction RNAs were treated with 4U TURBO DNase (Ambion) at 37°C for 1 h. Following DNase treatment, samples were extracted with acid phenol:chloroform and RNA transcripts were recovered by ethanol precipitation and resuspended in RNase-free water.

Ligation of adaptor to 3′ end of RNAs. 1 µg of RNA in 30 µl was incubated at 80°C for 2 min, then cooled to 4°C and incubated for 5 min. The RNA was then combined with PEG 8000 (12% final concentration), SOLiD-specific oligo s897 or Illumina-specific JB2 [5′ adenylated and 3′-end blocked with a 3' C3 spacer group oligo] (8.3 ng/µl final concentration), T4 Rnl2 reaction buffer (NEB) and 300U of T4 Rnl2, truncated (NEB). The mixture was incubated for 3 h at 37°C followed by 15 h at 16°C, and EDTA (20 mM final concentration) was added to stop the reactions.

Fragmentation of adaptor-ligated RNAs. 60 µl of 2x alkaline fragmentation solution (2 mM EDTA, 10 mM Na2CO3, 90 mM NaHCO3, pH 9.3) was added to the ligation reactions and this mixture was incubated at 95°C for 80 min. The majority of fragments were between ~50 and ~225 nt. The fragmentation reaction was stopped by the addition of 232 µl of ice-cold nuclease-free water, 40 µl 3M NaOAc (pH 5.5) and 4 µl of 10 mg/ml RNase-free glycogen. 500 µl of isopropanol was added and samples were placed at -80°C for 16 h, centrifuged (30 min, 21,000 x g at 4°C). The pellet was washed twice with 80% ethanol, air dried at room temperature for 5 min and re-suspended in 10 µl of nucleasefree water.

Size selection of fragments. Fragmented adaptor-ligated RNAs were mixed with 10 µl 2x RNA loading dye (95% deionized formamide, 18 mM EDTA, 0.25% SDS, xylene cyanol, bromophenol blue, amaranth) and fragments were separated by electrophoresis on 10% 7M

urea slab gels (equilibrated and run in TBE). The gel was stained with SYBR Gold nucleic acid gel stain (Life Technologies) and species ranging from ~75 to ~130 nt (for SOLiD libraries) and from ~60 to ~110 nt (for Illumina libraries) were excised from the gel. The gel was crushed and incubated in 400 µl of nuclease-free water at 70°C for 10 min. The gel debris were removed using a Spin-X column (Corning) and nucleic acids were isolated by ethanol precipitation and re-suspended in 10 µl of nuclease-free water.

cDNA synthesis. 10 µl of gel-eluted fragments were mixed with either 0.5 µl of SOLiD-specific s898a or Illumina-specific JB1 oligonucleotide (100 pmol/µl) and incubated at 65°C for 5 min then cooled to 4°C. 9.5 µl of a cocktail containing 4 µl of 5x First-Strand buffer (Life Technologies), 1 µl of 10 mM dNTP mix, 1 µl of 100 mM DTT, 1 µl (40U) RNase OUT (Life Technologies), 1 µl (200U) of SuperScript III Reverse Transcriptase (Life Technologies) and 1.5 µl of nuclease-free H2O. The reactions were incubated in a thermal cycler with a heated lid at 48°C for 60 min, followed by 15 min at 70°C. Reactions were cooled to room temperature, 20U of RNase H (Ambion) was added, and the reactions were incubated at 37°C for 40 min.

Size selection of cDNAs. An equal volume of 2x RNA loading dye was added and nucleic acids were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in TBE). The gel was stained with SYBR Gold nucleic acid gel stain (Life Technologies) and ~100 to ~150 nt speces (SOLiD) and ~90 to 140 nt (Illumina) species were excised from the gel. cDNAs were recovered from gel and resuspended in 10 µl of water as described above.

Circularization of cDNAs and PCR amplification. Circularization of the RT products was performed withssDNA CircLigase (Epicentre) in 20 µl reaction at 60°C for 1 h

according to the manufacturer's recommendations. Reactions were then incubated at 80°C for 10 min to inactivate the ligase. PCR was performed directly on the circularized products without additional purification using 5' and 3' primers from the SOLiD RNA Barcoding Kit. After amplification, cDNAs were separated by electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1xTBE), and cDNAs between ~125 bp and ~175 bp (SOLiD) and between ~155 bp and ~205 bp (Illumina) were isolated by gel excision. Amplified cDNAs were recovered from gel and resuspended in 10 μl of water as described above. The amplified cDNA libraries were sequenced using an Applied Biosystems SOLiD system (5500 XL) or Illumina HiSeq 2500 (run in rapid mode).

NET-seq and mNET-seq: data analysis. Bowtie version 1.0.0 was used to align the first 15 bases of each raw read to the *E. coli* reference genome (NC_000913.3) and identify those reads that were perfect matches. Alignment statistics are provided in Table S7. Among these reads, we used those that aligned to a unique position in the genome to identify pause sites. This results in omission of non-unique sequences and excludes from analysis rRNA genes and some tRNA genes. Read counts in each replicate were normalized. Pause sites were identified as genomic coordinates where the read count was at or above a threshold value of 25 'normalized' reads and that represented a local maximum in a 51 bp window centered on the coordinate (Tables S1 and S8). Sequence logos of the pause sites were generated from pauses identified in both replicates using Weblogo 3 (Crooks et al., 2004). Results of analysis using the SOLiD highthroughput sequencing platform are shown in Figs. 1 and 3 and Tables S1-S10. Results of analysis using the Illumina high-throughput sequencing platform are shown in Figs. S2, S6 and Table S7.

NET-seq and mNET-seq: data deposition. Raw reads have been deposited in the NIH/NCBI Sequence Read Archive under the study accession number SRP039384. Immunoblotting. Proteins in LDS Loading Dye (Life Technologies) were separated on a gradient NuPAGE 4-12% Bis-Tris gel [equilibrated and run in NuPAGE MOPS buffer (Life Technologies)] and transferred to a low fluorescence PVDF membrane (Immobilon-FL, EMD Millipore). Membranes were blocked in 5% non-fat dry milk in PBS, and then probed using either a mouse anti-FLAG M2 antibody (Sigma Aldrich) or mouse anti-β′ antibody (Neoclone, clone NT73) as a primary and a goat anti-Mouse IR-Dye 680LT (Li-Cor Biosciences) as a secondary antibody. Imaging was performed with the Odyssey Infrared Imaging System (Li-Cor Biosciences) according to the manufacturer's protocol.

***In vitro* transcription assays.** Linear transcription templates containing a T7A1 promoter, followed by a 29 bp U-less cassette, sequences extending from position –16 to +10 of each PE, and a 9 bp spacer sequence were synthesized by PCR from pIA171 using oligos JB49, JB105, JB111-120 and JB126. 40 nM PCR template was mixed with 50 nM RNAP holoenzyme in transcription buffer [20 mM Tris HCl pH 8.0, 20 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA, 2% glycerol] and incubated at 37°C for 5 minutes to form open complexes. Complexes were then walked to position +29 by the addition of 150 μM ApU, 10 μM ATP, 10 μM CTP [cold CTP + 5 μCi 32P-α-CTP (Perkin Elmer; 3000 Ci/mmol)], 2.5 μM GTP and incubation at 37°C for 15 minutes. NTPs and heparin were then added to final concentrations of: 150 μM ATP, 150 μM UTP, 150 μM CTP, 10 μM GTP (For assays shown in Figs. 1B, 2A, 3D, and S4) and 40 μg/ml heparin. Assays shown in Figure S7 were done using 40 μM GTP, 80 μM GTP, or 160 μM GTP. Aliquots were taken at 0, 10, 20, 30, 45, 60, and 120 seconds and mixed with an equal

volume of gel loading buffer [95% formamide, 100 mM Tris-HCl pH 8.0, 18 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue] to stop reactions. Samples were heated at 95°C for 5 min, cooled to room temperature, and run on 12% TBE-Urea polyacrylamide gels (UreaGel system, National Diagnostics). A gradient buffer system was used: upper reservoir, TBE; lower reservoir, BE containing 0.3 M NaOAc. Autoradiography of gels was performed using storage phosphor screens and a Typhoon 9400 variable mode imager (GE Life Science) and quantified using ImageQuant software. Pause-capture efficiency was calculated as described in (Landick, Wang et al., 1996). Values of pause-capture efficiency shown in Figs. 1-3, S7 are the averages of three independent experiments. To prepare the RNA ladders shown in Figure S4, +29 complexes were formed as described above and 200 μM of 3′-deoxy-ATP, 3′-deoxy-UTP, 3′-deoxy-GTP, or 3′-deoxy-CTP was added along with 200 μM ATP, 200 μM UTP, 200 μM CTP, 200 μM GTP and 40 μg/ml heparin.

RNAP-DNA interaction assays: fluorescence-detected equilibrium binding assays. Equilibrium binding experiments were performed using a σ70-molecular-beacon assay as described in chapter one and (Mekler et al., 2011a, Mekler et al., 2011b, Zhang et al., 2012).

RNAP-DNA interaction assays: fluorescence-detected kinetic assays. High-salt-induced-dissociation experiments were performed using a σ70-molecular-beacon assay as described in chapter one and (Mekler et al., 2011a, Mekler et al., 2011b, Zhang et al., 2012).

Translocation bias assays. Translocation bias assays were performed essentially as in (Hein et al., 2011). Transcription elongation complexes were reconstituted by incubating 100 nm *E. coli* RNAP core enzyme and 500 nM nucleic-acid scaffold in 100 μl 25 mM HEPES-KOH, pH 8.0, 130 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.15 mM EDTA, 5%

glycerol, and 25 μg/ml acetylated bovine serum albumin for 15 min at 37°C. Pyrophosphorolysis was initiated by addition of 1 μl 0.05 U/μl apyrase (New England Biolabs) and 1 μl 50 mM sodium pyrophosphate; reaction mixtures were incubated at 37°C, and 10 μl aliquots were withdrawn after 0, 1, 2.5, 5, 10, 20, and 30 min and quenched by mixing with 10 μl 98% formamide, 10 mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol. To confirm that transcription elongation complexes were catalytically active, a "chase" reaction was performed after the last time point, adding 11.4 μl reaction mixture to 0.6 μl 20 mM GTP (upper scaffold of Figure 4B) or 0.6 μl 20 mM UTP (lower scaffold of Figure 4B), incubating 5 min at 37°C, and withdrawing and quenching an aliquot as above. Products were applied to 7 M urea 20% polyacrylamide gels (19:1 acrylamide:bisacrylamide) slab gels, electrophoresed in TBE, and analyzed by storage-phosphor scanning (Typhoon; GE Healthcare). The fraction remaining of 9 nt RNA at each time point from each of three independent experiments was fitted to a single exponential for a pseudo-first-order reaction, and the half-life was determined (Hein et al., 2011).

Chapter 6

RNAP-CRE interactions plays functional roles in other domains of life: eukaryote and archaea

[1]Hanif Vahedian-Movahed, Fin Werner, Craig Kaplan, Bryce Nickels, Richard H. Ebright

6.1 Background

Residues of the RNAP core that mediate sequence-specific RNAP-CRE interactions are conserved in RNAP across three domains of life (Fig 1 A). Additionally, the structures of eukaryotic RNAP II elongation complex and bacterial RNAP initiation complex show similarity in mode of interactions with an unwound G of non-template strand at end of the transcription bubble (G_{CRE}) (Cheung & Cramer, 2011, Zhang et al., 2012). The G at CRE position +2 is inserted into a pocket formed by RNAPII amino acids residues equivalent to βR151, βI445, βD446, βR451, βT539, and βV547, adopting the same conformation, and making the same interactions, as for G at CRE position +2 in the crystal structure bacterial RNAP initiation complex (Fig. 1B,C). Based on the evidence we suggest that RNAP-CRE interactions occur in, and counteract pausing during transcription elongation by all multi-subunit RNAPs. To test this hypothesis, I sought to compare the pausing properties and translocation states of transcription elongation complexes reconstituted with bacterial RNAP (Escherichia coli RNAP), archaeal RNAP (Methanococcus janaschii RNAP), and

[1] Author contributions: The archaeal (*M. jannaschii*) RNAPs were provided by laboratory Prof. Finn Werner (University College London). The eukaryotic (*S. cerevisiae*) Pol IIs were provided by laboratory of Prof. Craig Kaplan (Texas A&M University). I performed all the experiments.

eukaryotic RNAP II (Sacchharomyces cerevisiae), in each case using the wild-type RNAP and a set of substituted RNAP derivatives containing substitutions of residues that, in bacterial RNAP, mediate RNAP-CRE interactions.

6.2. Result

6.2.1. RNAP-CRE interactions occurs in and modulate pausing by a eukaryote (*S. cerevisiae)* RNAP II.

To test the hypothesis that RNAP-CRE interactions occurs in eukaryote, I sought to investigate the pausing behaviors of *S. cerevisiae* wild-type Pol II and a substituted Pol II derivative containing substitutions of residues that, in bacterial RNAP, mediate RNAP-CRE interactions [Ala339]RPB2-Pol II. To enable direct comparison of pausing behaviors with *E. coli* RNAPs, I used a promoter-less scaffold-based pause assay which allow reconstitution of elongation complexes on similar DNA templates without the requirement of organism-specific promoter or transcription factors. In addition, I chose the sequence of yrbL pause which we have previously shown (in chapter 5) it can be modulated by RNAP-CRE interactions.

My data in Fig 2A, B shows that the *S. cerevisiae* WT-Pol II show a very weak pause signal at yrbL PE (position 17) which contain consensus base +1G (the next nucleotide to be added to the transcript is G) compared to E. coli WT-RNAP. The data suggest that the pause sequence in E. coli may not be conserved in S. cerevisiae. The data in Fig 2A, B further shows that, although *S. cerevisiae* WT-Pol II show a very weak pause at position 17, the S. cerevisiae [Ala339]RPB2-Pol II, increases the previously this weak pause at yrbL PE which contain consensus base +1G (the next nucleotide to be added to the transcript is G) in analogous manner to E. coli [Ala446]β-RNAP. Additionally, the data show that

[Ala339]RPB2-Pol II does not induce pausing when +1 base is a T or C and does not modulate a very weak pause observed when +1 base is a T. This data suggest that effect of sequence-specific RNAP-CRE interaction is conserved in *S. cerevisiae* where it play functional role by counteracting inherent tendency of RNAP to pause where the next incoming base is a G. We conclude that sequence-specific RNAP-G_{CRE} interaction that occur in E. coli during elongation also occurs in S. cerevisiae and play analogous functional roles during elongation by counteracting the inherent tendency of RNAP to pause in a manner that correlate with the presence of a G as the next incoming base prior to translocation. We hypothesis that CRE modulate pausing in yeast in analogous manner as in E. coli by modulating translocational bias and regardless of the observations that the pause sequence identified in E. coli does not seems to be conserved in S. cerevisiae.

My data in Fig 2A, also reveal interesting information about the empirically defined pause site observed with pol II at one bp downstream of the yrbL PE (pause at potion 18 on yrbL template). As can be seen in Fig. 2, *S. cerevisiae* WT-pol II mainly pause at position 18 on yrbL PE. This is not in agreement with the pause position observed with bacteria RNAP where they pause one nucleotide downstream at position 17. But this is in agreement with paused observed for Calf Thymus pol II which also pauses at position 18 (Larson et al 2014). One might conclude that the pause consensus sequence observed in bacteria is conserved in yeast but with one bp shifted downstream. However, the data in Fig. 2 is in disagreement with this conclusion. Fig. 2 shows that all base substations at position +1 assay with *S. cerevisiae* WT-pol II do not eliminate the pause at position 18. This results indicate that position yrbL PE+1 is probably not the determinant of the pause at positon 18 observed in *S. cerevisiae* pol II. Additionally, comparing the pause behavior

at this position for wild type and [Ala339]RPB2-Pol II, show no significant effect by CRE. This data indicate that the pause observed at position 18 is neither caused nor modulated by CRE. We speculate that the pause at position 18 might be irrelevant to the consensus PE element or might be a secondary pause.

The above discussed empirically defined pause site provide a useful tool to test our hypothesis that predicts regardless of the components of a PE, if a PE occurs immediately prior to a G_{CRE}, the RNAP-CRE interactions will modulate the pausing behavior by stabilizing the post-translocation state (unless the pause is of backtracking nature and the kinetic of entering to backtracking is faster than the kinetic of stabilization by CRE). To test this hypothesis we decided to introduce a consensus G_{CRE} immediately downstream of the empirically defined pause site at position 18 (Fig. 3) and compare the pausing for WT and [Ala339]RPB2-Pol II. In agreement with our prediction the data in Fig. 3 shows that the [Ala339]RPB2-Pol II increases the pause at the empirically defined pause site at position 18. In order to further confirm this observations, we sought to use similar scaffold-based pause assays but using a longer template in which the RNAP could encounter larger number of pause site and G_{CRE}. To do this I designed a strategy (Fig) in which an elongation complex is reconstituted on a short DNA scaffold and ligated to a longer downstream DNA duplex. The downstream DNA is immobilized on magnetic beads prior to the ligation, allowing high-salt wash to select for only fully ligated elongation complex. I chose a DNA sequence which have been previously used for pause studies and include several define pause site for E. coli RNAP. To allow single bp resolution I chose two different orientation for the DNA. One DNA start with pause a, b, c, d, and his. The second template start with pause d, his, a, b and c.

6.2.2. RNAP-CRE interactions modulate translocational bias in *S. cerevisiae* RNAP II.

In previous section we proposed that regardless of the components of a PE in S. cerevisiae, if a PE occurs immediately upstream of a G_{CRE}, the RNAP-CRE interactions will modulate the pausing behavior by stabilizing the post-translocation state in analogues manner to E. coli. Moreover, in previous section we found supporting evidence for this proposal by studying the effect of RNAP-CRE interaction disruption on the pause capture efficiency of the empirically defined pause site at position 18 that had a G_{CRE} inserted immediately at its downstream. Here we sought to assessed whether RNAP-G_{CRE} interactions affect the translocational register of the TEC by assessing sensitivity of TECs to pyrophosphorolysis

(Fig. 4). Sensitivity to pyrophosphorolysis provides a measure of TEC translocation because a TEC in a pretranslocated state is sensitive to pyrophosphorolysis but a TEC in a posttranslocated state is resistant (Fig. 4A) (3). We performed assays with WT-Pol II and [Ala339]RPB2-Pol II on templates containing G or T at position +1 and compared to the analogous results obtained with E. coli RNAPs (Fig. 4A, B).We found that TECs with *E. coli* RNAP-βWT and *S. cerevisiae* WT-Pol II were ~5 and ~7 times as sensitive to pyrophosphorolysis when the template contained +1G as when the template contained +1T, respectively, which indicated that a greater proportion of TECs on templates containing +1G were in a pretranslocated state than of TECs on templates containing +1T (Fig. 4B,C). The results directly demonstrate the effect of G at position +1 on translocation bias. TECs with *E. coli* RNAP-βD446A and *S. cerevisiae* [Ala339]RPB2-Pol II were ~4 and ~2 times as sensitive to pyrophosphorolysis as TECs with *E. coli* RNAP-βWT and *S. cerevisiae* WT-

Pol II when the template contained +1G, respectively (Fig. 4C). Whereas, in contrast, TECs with *E. coli* RNAP-βD446A and *S. cerevisiae* RNAP-βWT exhibited identical sensitivities to pyrophosphorolysis when the template contained +1T (Fig. 4C). The results indicate that, on templates containing +1G, a greater proportion of TECs with RNAP defecting in RNAP-CRE interaction are in a pretranslocated state as compared to TECs with WT RNAPs which makes RNAP-CRE interactions. We conclude that sequence specific RNAP-G_{CRE} interactions stabilize the TEC posttranslocated state in both E. coli and S. cerevisiae, which provides a mechanistic explanation for the finding that RNAP-G_{CRE} interactions counteract pausing in both *E. coli* and *S. cerevisiae.*

6.2.3. RNAP-CRE interactions in archaeal RNAP (Methanococcus janaschii RNAP).

To test the hypothesis that sequence-specific RNAP-CRE interactions occurs in archaea, I sought to investigate the pausing behaviors of *M. janaschii* wild-type RNAP and a substituted RNAP derivative of which that contain substitution of residue that, in bacterial RNAP, mediate RNAP-CRE interactions [Ala337]RpoB RNAP. To enable direct comparison of pausing behaviors with *E. coli* RNAPs, I used analogous promoter-less scaffold-based pause assay as in previous section which allow reconstitution of elongation complexes on similar DNA templates without the requirement of organism-specific promoter or transcription factors. Here, I also chose the same sequence of yrbL pause which we have previously shown it can be modulated by RNAP-CRE interactions in *E. coli* and in *S. cerevisiae* (see previous section and chapter 5).

My data in Fig 2A, B shows that the *M. janaschii* WT-RNAP show a pause signal at yrbL PE (position 17), however, the pause signal is not eliminated by introductions of A, C, or T at position +1, and that the pause signal is strengthen by substation of the consensus

+1G to +1A. This data suggest that the pause sequence in E. coli may not be conserved in *M. janaschii*.

The data in Fig 2A, B further shows that, the *M. janaschii* [Ala337]RpoB RNAP, in analogous manner to E. coli [Ala446]β-RNAP, increases the pause capture efficiency at yrbL PE which contain consensus base +1G by 10%. Additionally, the data show that *M. janaschii* [Ala337]RpoB-RNAP does not modulate the pausing when +1 base is a T or A and only moderately increase the pause capture efficiency (by ~4%) when +1 base is a A. This data suggest that effect of sequence-specific RNAP-CRE interaction is also conserved in *M. janaschii* where it play functional role by counteracting inherent tendency of RNAP to pause where the next incoming base is a G. The data further suggest that the pause sequence in E. coli may not be conserved in *M. janaschii*. We conclude that sequence-specific RNAP-G_{CRE} interaction that occur in *E. coli* and *S. cerevisiae* during elongation also occurs in *M. janaschii* and play analogous functional roles during elongation by counteracting the inherent tendency of RNAP to pause in a manner that correlate with the presence of a G as the next incoming base prior to translocation. We hypothesis that CRE modulate pausing in *M. janaschii* in analogous manner as in *E. coli* and *S. cerevisiae* by modulating translocational bias, regardless of the observations that the pause sequence identified in E. coli does not seems to be conserved in *S. cerevisiae* and *M. janaschii*.

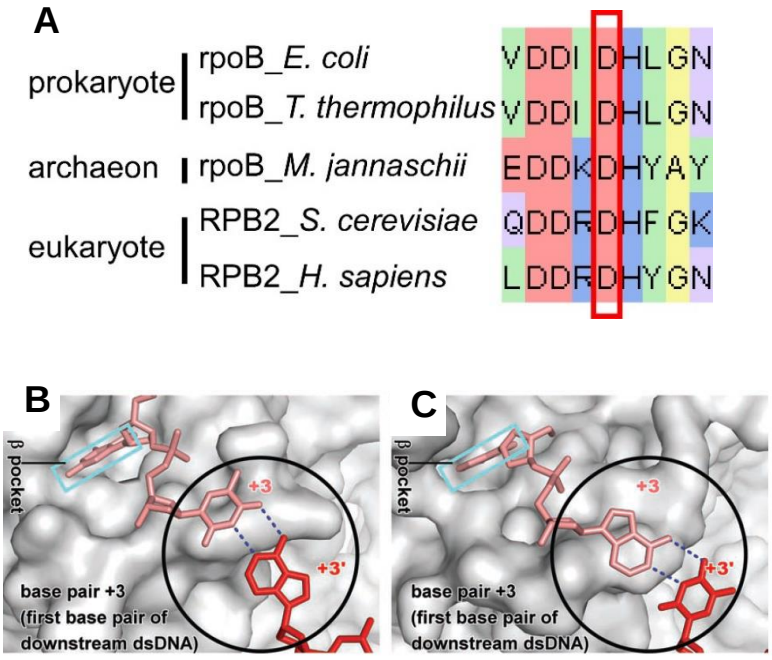
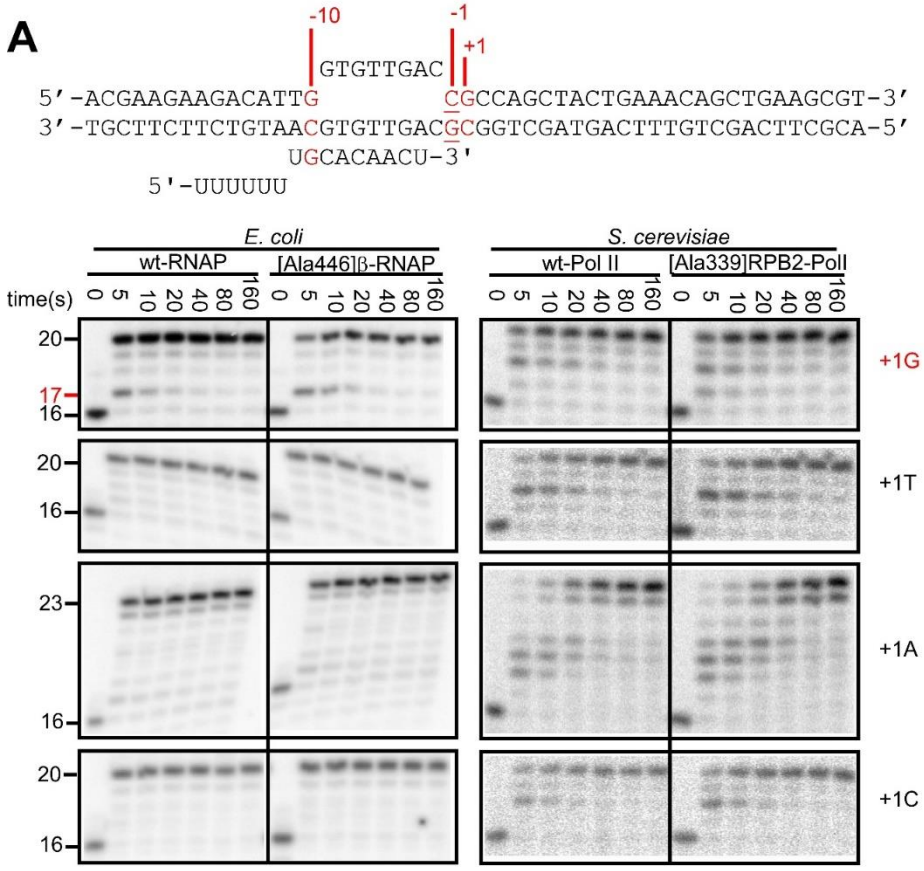


Figure 1. RNAP-CRE interactions in bacterial, eukaryote and archaea RNAPs. (A) Sequence conservation of specificity-determinant-residue of CRE position +2 in species from three domain of life. (B) Crystal structure of bacteria RPo PDB: 4G7H. (C) Crystal structure of a yeast RNAP II backtracked and arrested RDe (PDB 3PO2; 140).



B

DNA site	E. coli			S. cerevisiae		
	WT	D446A	(%D446A - %WT)	WT	D399A	(%D446A - %WT)
+17G	50%	66%	16%	16%	34%	16%
+17T	absent	absent	0%	absent	absent	0%
+17A	absent	absent	0%	<10%	<10%	0%
+17C	absent	absent	0%	absent	absent	0%

Figure 2. RNAP-CRE interactions occurs in and modulate pausing by a eukaryotic Pol II (A) in vitro pause assay with elongation complexes reconstituted with bacterial RNAP (Escherichia coli RNAP), and eukaryotic RNAP II (Sacchharomyces cerevisiae). (B) Effect of disruption of RNAP-CRE interactions on pause capture efficiencies.

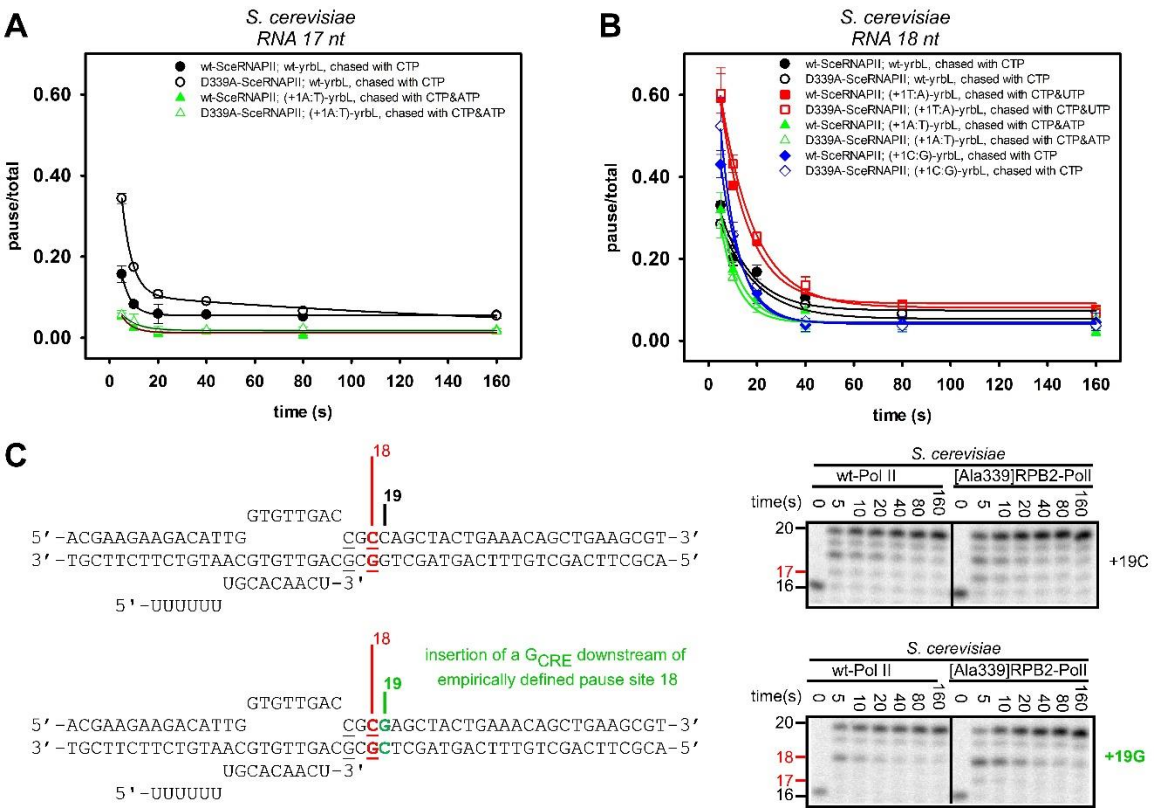


Figure 3. Analysis of the effect of RNAP-CRE interactions at pause sites with eukaryotic pol II at position 17 and 18 (A) comparison of the pause-decay graphs for *S. cerevisiae* WT- and [Ala339]RPB2-Pol II with template carrying G or A substitutions at position 17 (template with T and C substitutions were weak and not quantifiable). (B) comparison of the pause-decay graphs for *S. cerevisiae* WT- and [Ala339]RPB2-Pol II with template carrying G, A, T or C substitutions at position 18. (C). effect of the CRE when G_{CRE} is inserted downstream of the empirically defined pause site at position 18.

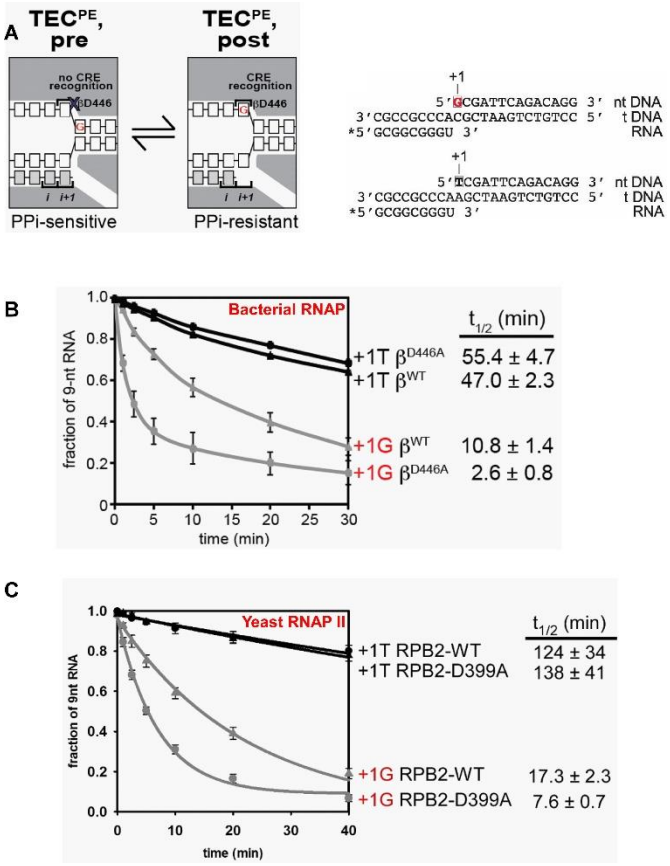
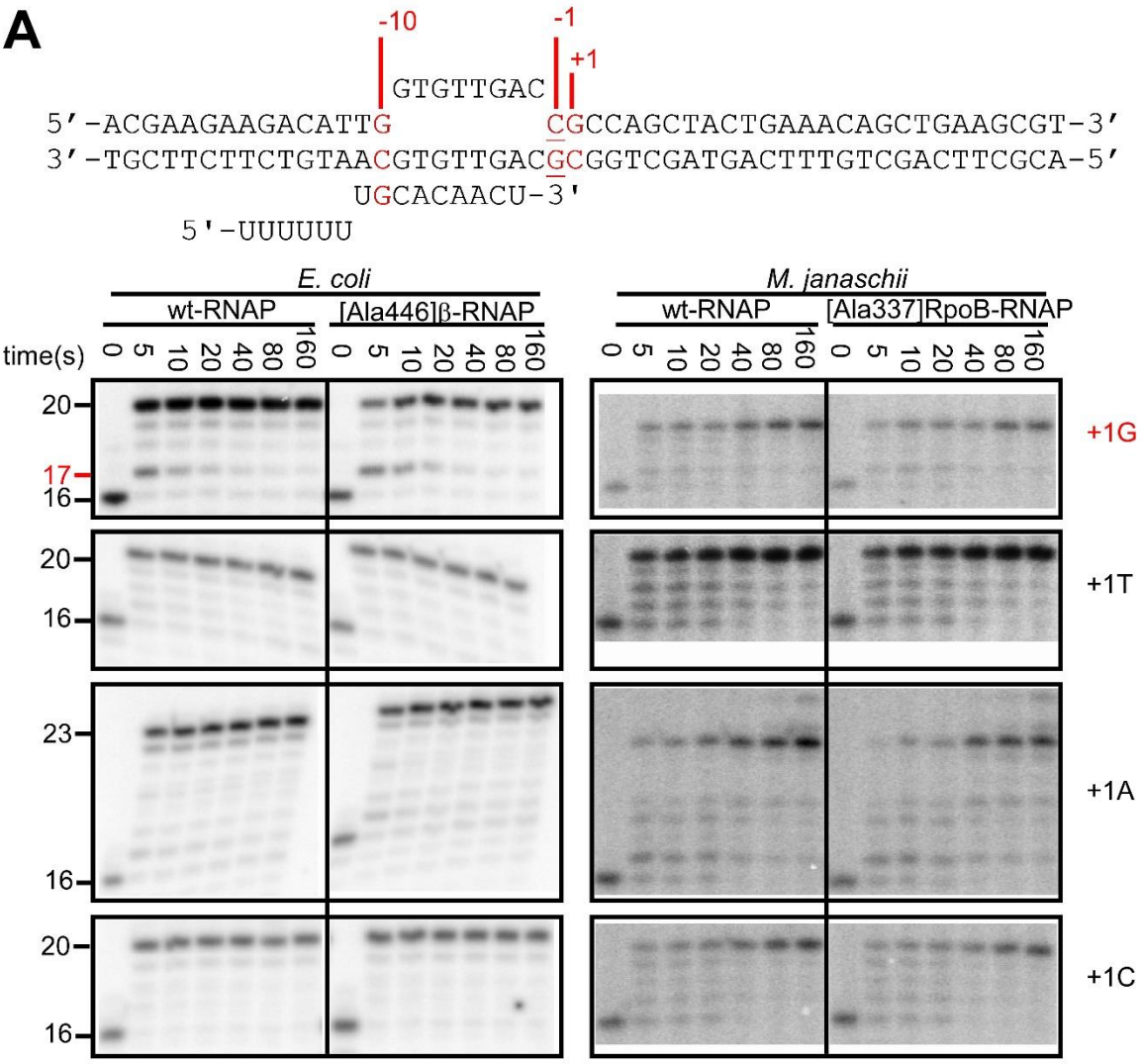


Figure 4. Sequence-specific RNAP-G$_{CRE}$ interactions modulate translocation bias of a eukaryotic Pol II. (A) Structural organization of TEC in pre-translocated state at consensus PE (left; unfavorable RNAP-CRE interaction) and TEC in post-translocated state at consensus PE (right; favorable RNAP-CRE interaction). PPi, pyrophosphate. Right panel, Nucleic-acid scaffolds used for translocation-bias assays. Asterisk, radiolabel on RNA 5′ end; box, position corresponding to G$_{CRE}$; red, consensus PE base. (B) Translocational bias for RNAP β^{WT} and RNAP β^{D446A} on nucleic-acid scaffolds containing G or T at position corresponding to G$_{CRE}$. Graph shows fraction of unaltered RNA (mean±SEM; 3 measurements) as a function of time.



B

DNA site	E. coli			M. janaschii		
	WT	D446A	(%D446A - %WT)	WT	D337A	(%D337A - %WT)
+17G	50%	66%	16%	17%	26%	9%
+17T	absent	absent	0%	15%	14%	-1%
+17A	absent	absent	0%	29%	32%	3%
+17C	absent	absent	0%	14%	18%	4%

Figure 5. RNAP-CRE interactions occurs in and modulate pausing by a archaea RNAP (A) in vitro pause assay with elongation complexes reconstituted with bacterial RNAP (Escherichia coli RNAP), and archaeal RNAP (Methanococcus janaschii RNAP) .(B) Effect of disruption of RNAP-CRE interactions on pause capture efficiencies.

6.3. Materials and methods

Proteins. Eukaryote *S. cerevisiae* WT-RNAP II and D339A-RNAPII were prepared and gifted by Craig Kaplan lab (University of Texas A&M). Archaea *M. janaschii* WT-RNAP and D337A-RNAP were prepared and gifted by Finn Werner lab (University College London). E. coli WT-RNAP and D446A-RNAP were purified as reported in chapter one.

Nucleic-acid scaffolds. Non-template oligodeoxyribonucleotides (ntDNA), template oligodeoxyribonucleotides (tDNA) and oligoribonucleotide (RNA) were resuspended in RNase-free water (Ambion). To prepare nucleic-acid scaffolds for pause assays of *E. coli* and *S. cerevisiae* we used the following strategy: nontemplate-strand oligonucleotide (2uM), template-strand oligonucleotide (2 uM) and RNA oligonucleotide (1 uM) were combined in 50 μl 10 mM Tris-HCl pH 7.9, 40 mM KCl, and 5 mM MgCl2, heated 2min at 95°C, and cooled to 25°C in 2°C steps with 60 s per step in a thermal cycler (Applied Biosystems). To prepare nucleic-acid scaffolds for pause assays of *M. janaschii* we used the following strategy: template-strand oligonucleotide (1 uM) and RNA oligonucleotide (9 uM) were combined in 50 μl 10 mM Tris-HCl pH 7.5, heated 2min at 95°C, and cooled to 25°C in 2°C steps with 60 s per step in a thermal cycler (Applied Biosystems). To prepare nucleic-acid scaffolds for translocational-bias assays nontemplate strand oligonucleotide (1 μM), template-strand oligodeoxyribonucleotide (1 μM), and 5'-32P-labelled RNA oligonucleotide [0.5 μM; 5'-32P-labelled using T4 polynucleotide kinase (New England Biolabs) and [γ32P]-ATP (PerkinElmer)] were combined in 50 μl 10 mM Tris-HCl (pH 7.9), 40 mM KCl, and 5 mM MgCl2, heated 2 min at 95°C, cooled to 45°C in 30 s, and cooled to 25°C in 2°C steps with 120 s per step, in a thermal cycler (Applied Biosystems).

Reconstitution of elongation complex (EC) with RNAP from three domain of life: ECs formation with *E. coli* RNAP: 50 nM *E. coli* RNAP core enzyme was mixed with 40 nM yrbL nucleic-acid scaffold in transcription buffer TB, 20 mM Tris HCl pH 8.0, 20 mM NaCl, 3 mM MgCl2, 0.1 mM EDTA, 14 mM 2-mercaptoethanol and incubated at 37°C for 10 minutes to form elongation complexes (EC). ECs formation with *S. cerevisiae* RNAP II: 200 nM *S. cerevisiae* RNAPII core enzyme was mixed with 40 nM yrbL nucleic-acid scaffold in transcription buffer TB and 0.25 mg/ml BSA, and incubated at 30°C for 10 minutes to form elongation complexes (EC). ECs formation with *M. janaschii* RNAP: 50 nM *M. janaschii* core enzyme was mixed with 40 nM yrbL RNA : template DNA hybrid in TB and 0.25 mg/ml BSA, and incubated at 65°C for 15 minutes. Then, DNA non-template strand was added to a final concertation of 40 nM and the mixture was incubated at 65°C for an additional 15 minutes to form elongation complex (EC).

Scaffold-based pause assay. After EC were formed, they were walked to position +16 (one base upstream of the pause site) by the addition of 10 μM GTP [9.9 uM cold CTP + 0.1 μM 32P-α-GTP (Perkin Elmer; 3000 Ci/mmol)], and incubation for 2 minutes at 37 °C, 30 °C or 65 °C for *E. coli*, *S. cerevisiae* or *M. janaschii ECs*, respectively. Reactions were then chased by addition into a tube containing either CTP for yrbL +1G or +1C scaffolds, CTP and ATP for yrbL +1A scaffold or CTP and UTP for yrbL +1T scaffold to a final concentrations of 10 μM NTP for each. Aliquots were taken at 0, 5, 10, 20, 40, 80, and 160 seconds and mixed with an equal volume of gel loading buffer [98% formamide, 10 mM EDTA, 0.04% xylene cyanol, 0.04% bromophenol blue] to stop reactions. Samples were heated at 95°C for 2 min, centrifuged at 5000 RPM for 1 min, and run on 20% TBE-Urea polyacrylamide gels (UreaGelsystem, National Diagnostics). Autoradiography of gels was

performed using storage phosphor screens and a Typhoon 9400 variable mode imager (GE Life Science) and quantified using ImageQuant software. Pause-capture efficiency was calculated. Values of pause-capture efficiency are the averages of three independent experiments.

Translocation bias assays. Translocation bias assays were performed essentially as in (Hein et al., 2011, Vvedenskaya et al., 2014) but conditions were adapted for S. cerevisiae RNAP. Transcription elongation complexes were reconstituted by incubating 100 nm *S. cerevisiae* RNAP core enzyme and 30 nM nucleic-acid scaffold in 100 μl 25 mM HEPES-KOH, pH 8.0, 130 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.15 mM EDTA, 5% glycerol, and 25 μg/ml acetylated bovine serum albumin for 15 min at 30°C. Pyrophosphorolysis was initiated by addition of 1 μl 0.05 U/μl apyrase (New England Biolabs) and 1 μl 50 mM sodium pyrophosphate; reaction mixtures were incubated at 30°C, and 10 μl aliquots were withdrawn after 0, 1, 2.5, 5, 10, 20, and 40 min and quenched by mixing with 10 μl 98% formamide, 10 mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol. To confirm that transcription elongation complexes were catalytically active, a "chase" reaction was performed after the last time point, adding 11.4 μl reaction mixture to 0.6 μl 20 mM GTP (upper scaffold of Figure 4B) or 0.6 μl 20 mM UTP (lower scaffold of Figure 4B), incubating 5 min at 37°C, and withdrawing and quenching an aliquot as above. Products were applied to 7 M urea 20% polyacrylamide gels (19:1 acrylamide:bisacrylamide) slab gels, electrophoresed in TBE, and analyzed by storage-phosphor scanning (Typhoon; GE Healthcare). The fraction remaining of 9 nt RNA at each time point from each of three independent experiments was fitted to a single exponential for a pseudo-first-order reaction, and the half-life was determined (*3, 21*).

Chapter 7

[1]RNAP core performs sequence-specific transcription in the absence of sigma factor

Hanif Vahedian-Movahed, Bryce Nickels and Richard H. Ebright

7.1. Background

Bacterial RNAP holo enzyme constitute of a multi-subunit catalytic core and a dissociable subunit called sigma factor which confer the ability on RNAP to recognize specific promoter DNA sequences and sequence-specific transcription initiation (Burgess, 1969, Feklistov, Sharon et al., 2014). Historically, the concept of sequence-specific transcription was postulated based on a simple argument of cell economy which necessitate the presence of specific/discrete sites of transcription (start and end) rather than fortuitous transcription (Roberts, 1970). This proposition was initially supported by the observations of similarity between in vivo and in vitro transcribed RNAs (Geiduschek, Tocchini-Valentini et al., 1964), and was further supported by the observation of a bias for purines as initiating nucleotides (Sugiura, Okamoto et al., 1969), the observation of altered transcription initiation by mutant promoter DNA (Roberts, 1969a), discovery of transcription factors that define the initiation site (sigma) (Burgess, Travers et al., 1969) and factors that define the termination site (Rho) (Roberts, 1969b) of transcription. It might be safe to say that a major breakthrough in establishing the "sequence-specific initiation"

[1] The data presented in this chapter are not systematic enough for publications. This data are mainly preliminary results.

of transcription, comes about in 1969 by the discovery of the sigma factor. It was shown that the RNAP purified on a phosphocellulose column (PC) was inactive with the T4-DNA. However, when a polypeptide (the sigma factor) purified from a different fraction of the same column was added, it regained its activity (Burgess et al., 1969). The follow up works established that sigma function at initiation stage (Travers & Burgessrr, 1969). The second important breakthrough, comes about in 1975, when a DNA binding site for RNAP just upstream of the RNA sequence was discovered (-10 element) and found to be common in all promoters (Pribnow, 1975a). Other works provide direct evidence that sigma-mediated recognition of conserved promoter element is responsible for sequence-specific recognition of promoter element as well as DNA unwinding in transcription initiation (Feklistov et al., 2014).

Despite above findings, accumulating evidence demonstrates that the RNAP core, in the absence of sigma, can still perform transcription. The first piece of evidence comes from the observation that, the RNAP core that was passed through the PC column and had lost sigma factor, still showed limited activity with T4 DNA and it showed relatively good activity with some other DNA, such as calf thymus DNA (Burgess et al., 1969). This type of transcription has been assumed to originate from: 1) RNAP binding at break sites in the DNA; 2) RNAP binding to the free ends of linear DNA; 3) RNAP binding to DNA region with low melting temperature (poly (dA-dT) DNA regions); 4) contamination from trace amount of sigma (Burgess et al., 1969, Dausse, Sentenac et al., 1972, Vogt, 1969, Wheeler, Woody et al., 1987). Therefore, transcription with the RNAP core alone has been thought to be at lowered level and to be “non-sequence-specific”.

As such, the prevailing view has been that the RNAP core requires binding to sigma factor for sequence-specific transcription. This is because it has been assumed that, the RNAP core does not include the determinants for sequence-specific recognition of the core promoter DNA (see chapter one for details of RNAP-DNA interactions). Additionally, because it has been assumed that, the RNAP core lacks the requirement for DNA unwinding and primarily relies on sigma factor to facilitate promoter DNA unwinding.

The first assumptions in the paragraph above, has been challenged by the data presented in chapter 1 indicating that the RNAP core exhibit sequence-specificity toward a DNA element at the downstream edge of the transcription bubble non-template-strand of the core promoter region corresponding to positions -4 to +2, the "core recognition element," CRE. The second assumptions in the paragraph above, has been challenged by the data presented in chapter 2 indicating that the RNAP core interactions with a G_{CRE} facilitate promoter unwinding as well as by other study that show RNAP core is capable of unwinding poly (dA-dT) duplex DNA (Wheeler et al., 1987). These observations, led us to investigate if the RNAP core is capable of performing sequence-specific transcription initiation in the absence of sigma factor. To test this hypothesis, I used a next-generation-sequencing-based approach to compare the sequence-specificity of the transcription initiation throughout the E. coli genome at base-pair resolution for the RNAP core to that of the RNAP holo enzyme. My results establish that RNAP core performs sequence-specific transcriptions initiation in the absence of sigma factor.

7.2. Results and discussions

7.2.1. Sequence-specific sigma-independent transcription by the RNAP core in vitro

It has often been presumed that the majority of transcription initiation observed with the RNAP core alone, are not sequence-specific and originates from: 1) RNAP binding at break sites in the DNA; 2) RNAP binding to the free ends of the linear DNA; 3) RNAP binding to DNA region with low melting temperature (poly (dA-dT) DNA regions), or 4) contamination from trace amount of sigma factor (Burgess et al., 1969, Dausse et al., 1972, Vogt, 1969, Wheeler et al., 1987). Therefore, to permit reliable assessments of the sequence-specificity of the transcription initiation by the RNAP core we need a strategy that circumvent these issues. To accomplish this I used a genome-wide sequencing-based strategy depicted in Fig 1. In this method, I generated a supercoiled plasmid template library of *E. coli* genomic DNA fragments (pGENOME) and used highly pure RNAP core to generate RNA transcripts from the template library in vitro, and analyzed the transcript 5'-ends using high throughput sequencing to define the transcription start sites (TSS) (Fig 1). This strategy accommodates the use of DNA template with high quality (lacking free ends or nicks and possessing high sequence complexity) as well as an RNAP core enzyme with high level of purity (using an RNAP core derivative containing the β'[R295Q; E275K] substitutions which disrupt the interaction between sigma factor and the RNAP core (Arthur, Anthony et al., 2000)).

To capture the full profile of transcription by the RNAP core, I performed in vitro transcription experiments with the pGENOME template library, using, in parallel, RNAP holo, the RNAP core and the β'[R275Q; E295K]-RNAP core derivative (this substitution

disrupt the interaction between core and sigma factor and provide highly pure RNAP core). RNA products generated in the transcription reactions were isolated and analyzed using high-throughput sequencing of 5′ ends (5′ RNAseq) to map for each RNA product, the genomic regions that produced the RNA and the TSS position (Fig 1). TSS where identified as coordinates where at least 50 reads align and that represent a local maxima (peak) within a window size of 11-bp centered on the coordinate. For each sequence, we calculated the percentage of reads starting at each position (5 bp upstream to 5 bp downstream of the coordinate) and for the analysis, we looked at the TSS which more than 80% of their reads originates from the centered positions.

To define if RNAP core performs sequence-specific transcription initiations, we compared in parallel the TSS sequences generated by RNAP holo, RNAP core and the β'[R275Q; E295K]-RNAP core with the pGENOME template library. The weblogo of the RNAP holo presented in Fig.1. B, demonstrate, a conserve sequence feature of a purine (R) as TSS position +1, a pyrimidine as the position prior to TSS (position -1), and a conserved -10 element, reflecting the known sequence feature of a sequence-specific sigma70-dependent transcription by RNAP holo. Therefore, establishing the validity of our assay by recapitulating the sigma 70 sequence-specific transcription from a library of genomic DNA fragments. Additionally, the weblogo of the RNAP core presented in Fig.1.B. , demonstrate, a conserve sequence feature of a G as the TSS position +1, a T as the position -1 (a TG-motif), adjacent to a 4 bp AT-rich region. The data for the β'[295Q;275K]-RNAP core matches the data obtained with the wt RNAP core, therefore ruling out the complication due to contamination with sigma factor. Additionally, the pGENOME template library consists of negatively supercoiled circular plasmid to eliminate or reduce

the complication due to transcription from free DNA ends or at nick sites. Therefore, this data suggest that the RNAP core is capable of sequence-specific transcription in the absence of sigma factor in a manner that correlates with the presence of an AT-rich region followed by a TG-motif.

7.2.1. Assessment of the transcription inhibitory effect of sigma factor throughout the *E. coli* genome.

It has been indicated that that sigma factor binding to the RNAP core could reduce the non-specific binding to the DNA sites while increasing the binding to specific sites (Hinkle & Chamberlin, 1972a, Hinkle & Chamberlin, 1972b). Furthermore, another study has shown that at high salt concentrations, the sigma factor binding to the RNAP core could inhibit transcription from a poly (dA-dT) template (Shorenstein & Losick, 1973). Having established that the RNAP core alone can perform sequence-specific transcription in vitro, using a pGENOME complex library, we next sought to investigate if binding of sigma to core could inhibit transcription genome-wide in vitro. To do this, we determined the frequency of the TSS across the E. coli genome in every 50,000 bp for the RNAP core and for the RNAP holo (Fig 2. A, B). Our data show a reduction in the frequency of TSSs across the E. coli genome in the presence of sigma factor. This data agree with previous observations. Additionally, analyzing the sequence region of the suppressed TSSs (TSS observed with the RNAP core alone that disappear in the presence of sigma factor) shows a decrease or a loss in the AT-richness (Fig 2. C). This data could be interpreted as different level of specificity for sigma-independent transcription, and that the sigma factor preferably reduce the transcription from the more non-specific sites and more abundant sites (the probability of occurrence of TSSs with only a single TG motif could be higher

than TSSs that include both AT-rich and TG-motifs. This could explain why the suppressed TSSs are mainly enriched for only 1 out of 2 of the motifs and not both (TG vs both AT-rich and TG). It is plausible that the TSSs which include both motifs constitute smaller portion of the sigma-independent TSSs, while presenting a higher strength as AT regions, have lower melting temperature. Therefore, we next sought to investigate if there is a correlation between the level of the AT-richness and the promoter strength. To do this we ranked the TSSs based on the number of read counts and analyzed the sequence feature of their TSS regions. As it can be seen from Fig.3, there seems to be a correlation between the presence of the AT-rich motif and the number of read counts. In addition, we can notice an inverse correlation between the TSS read counts and the presence of the AT-rich motif. This data could explain our above observations that the majority of the suppressed sigma-independent transcription sites were the highest frequent ones and include a single TG-motif.

7.2.3. Assessments of the sequence-specific sigma-independent transcription in vivo

To define if the RNAP core can perform sequence-specific and sigma-independent transcription in vivo, we used merodiploid native-elongating transcript sequencing (mNET-seq) (Vvedenskaya et al., 2014). mNET-seq involves selective analysis of transcripts associated with an epitope-tagged RNAP in the presence of a mixed population of epitope-tagged RNAP and untagged RNAP (Fig. 2A). To do this, we generated an artificial condition that prevent association of sigma factor to epitope-tagged RNAP core in living cell: using a rpoN null strain to eliminate sigma54-dependent transcription in vivo and constructing an rpoC mutant defective in interaction with sigma factor (rpoC [275Q;295K] substitution). To do this, we introduced into rpoN null cells a plasmid

encoding 3xFLAG-tagged β'WT or 3× FLAG-tagged β'[275Q;295K], isolated RNA products associated with RNAP-β'WT or RNAP- β'[275Q;295K] by immunoprecipitation, converted RNA 5′ ends to cDNAs, and performed high-throughput sequencing (Fig. 4).

To define if the RNAP core perform sequence-specific transcription initiations in vivo we performed similar analysis as above, but found no discernible differences between transcription by the RNAP holo and the RNAP core. This could be: (a) due to a limitation in our methodology in completely abolishing the binding of sigma to the RNAP core; (b) due to the presence of a regulatory process that prevent the sigma-independent transcription in vivo under the tested physiological condition, for example it could be possible that some suppressor molecule binds to the DNA sites for sigma-independent transcription and prevent RNAP core binding to those sites, therefore leaving only the promoter sites available for the RNAP core to bind; or it could be possible that intrinsically RNAP core is efficient in binding to all sites (promoter and non-promoter sites) and an activator molecule or the genome structure surrounding the promoter sites, makes the RNAP core binding and transcription more efficient from promoter sites than non-promoter sites under tested in vivo condition. To address issue (a), we sought to look at the efficiency of our method in recovering if any, of the TSS that are observed in vitro. To do this, we instead analyzed only the TSSs identified that are common in both conditions in vitro and in vivo. To define if RNAP core perform sequence-specific transcription initiations that occurred in vitro also occurs in vivo, we intersected the list of TSS sequences observed with RNAP holo and the β'[275Q;295K]-RNAP core in vitro and in vivo, and obtained a list of TSS that are observed in both condition. In each case, we looked at the sequences in which more than 80% of their reads is generated from the centered position of the TSS peaks. From the

data shown in Fig.2 we observed a similar pattern for sigma-independent transcription matching closely with the results obtained in vitro with pGENOME template. However, the number of identified TSS with RNAP core in vivo that matched with the in vitro condition are very low (n=6) and the more concerning issue is that they seem to show a weak -10 element signal. Therefore, these data are not very conclusive at this point. It would be useful to optimize the experimental conditions to allow higher concentration of RNAP core in vivo or to deplete the concertation of sigma factor. In a second note, it could be possible that in vivo sigma-independent transcription is regulated and only occurs at certain physiological states. It would be useful to combine the current in vivo experimental strategy with single-gene knock out of some activators, suppressor or nucleoid-associated proteins. Additionally, it is possible that, sigma-independent transcription is not regulated, but still occurs at certain physiological states that free RNAP core concentration are higher than normal physiological state. For example, during sporulation when sigma factor is lost (Linn, Greenleaf et al., 1973) or during stringent response when ribosomal transcription are shutdown, and therefore, a large fraction of RNAP previously engaged in transcription elongation of rRNA genes becomes available. Under this condition, the sigma-independent transcription might become beneficial by stochastic production of new products among the cell in the population. Therefore, it would be useful to use the current in vivo experimental strategy at different physiological states such as during stringent response.

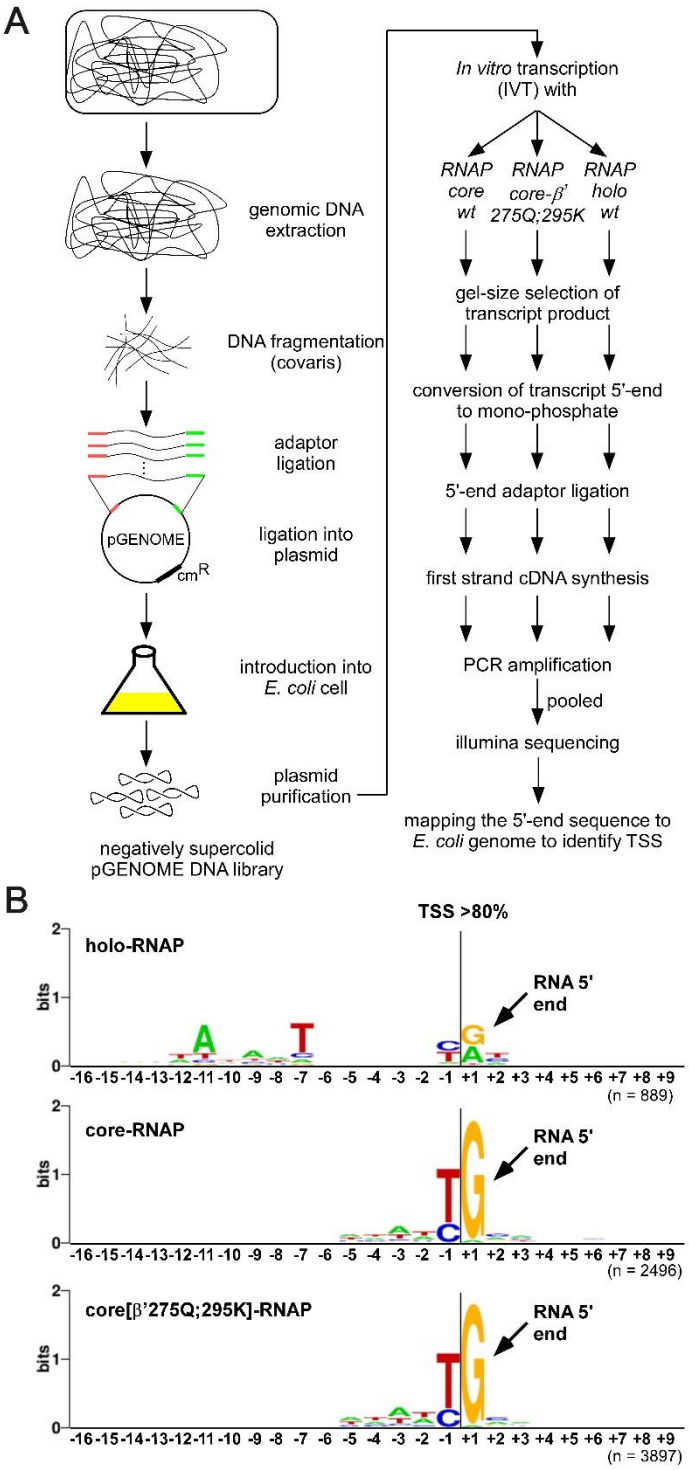


Figure 1. Sequence-specific sigma-independent transcription by the RNAP core in vitro

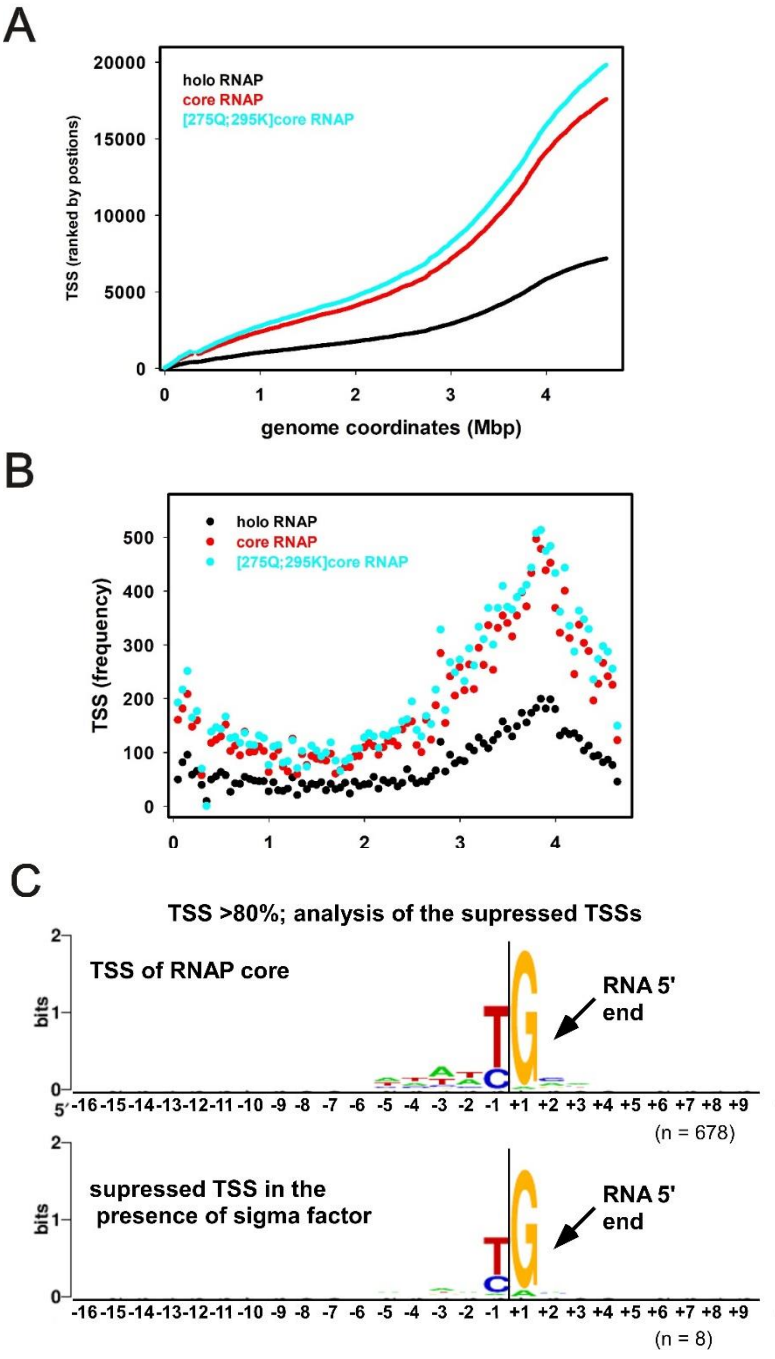


Figure 2. Sigma factor binding to the RNAP suppresses the RNAP core mediated transcription genome-wide in vitro.

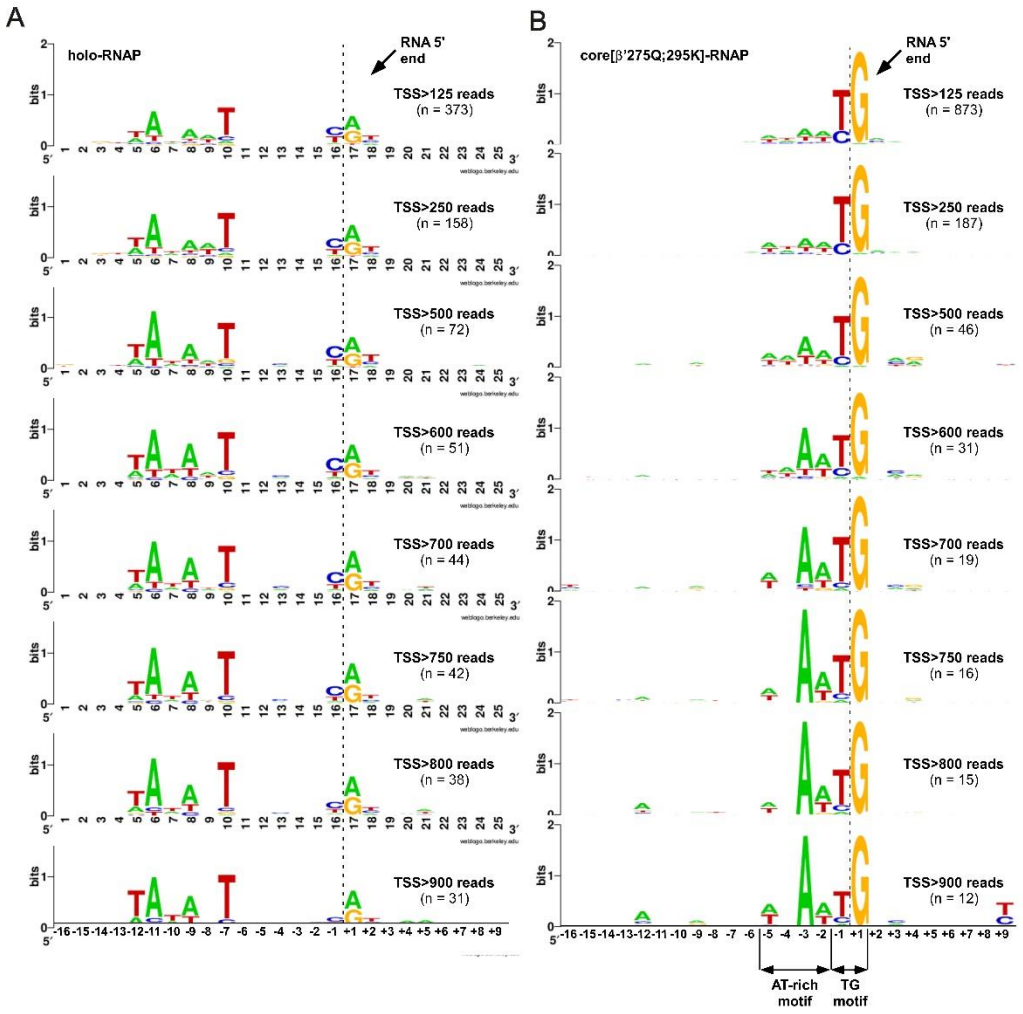


Figure 3. Correlation between the number of reads and AT-richness.

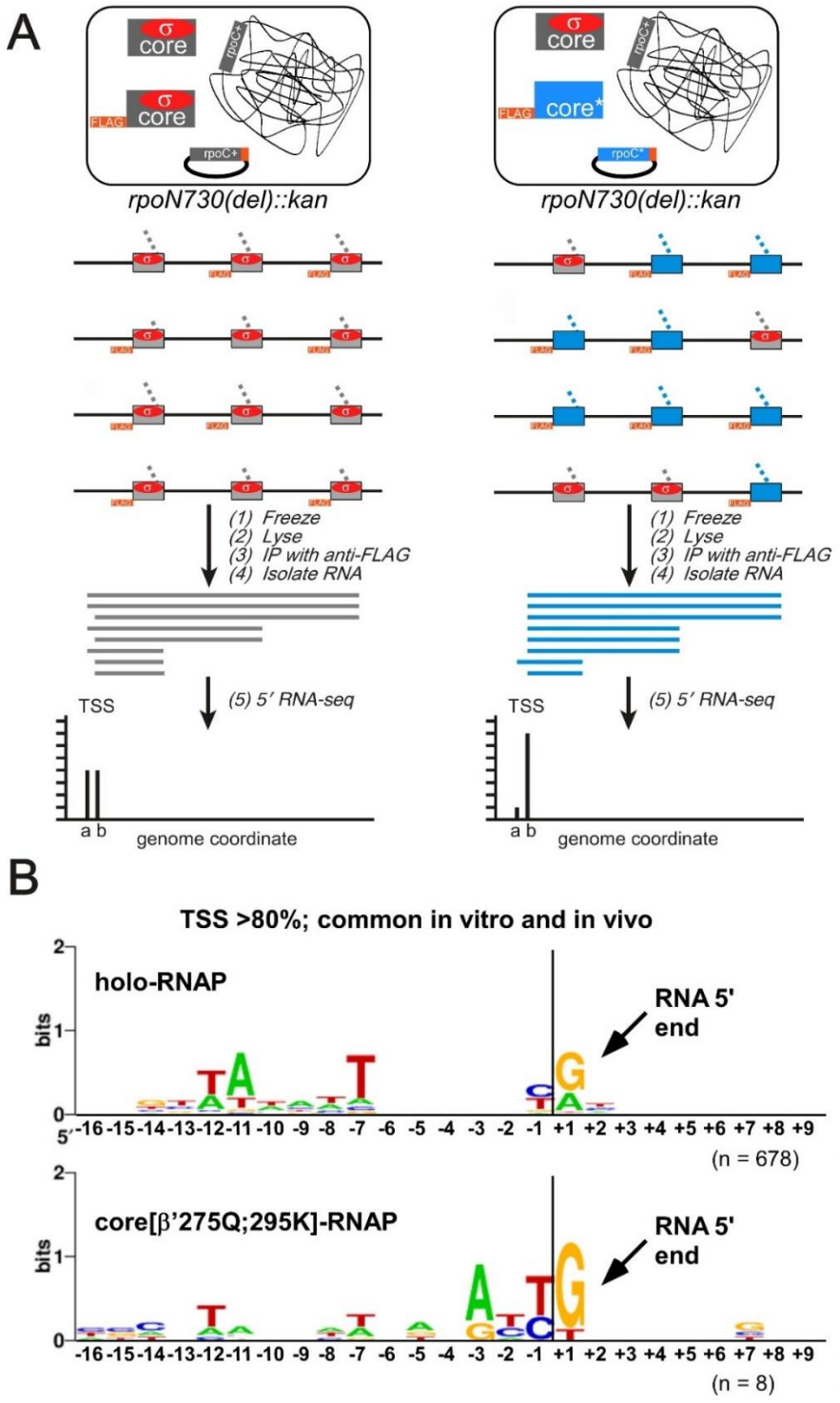


Figure 4. In vivo assessments of sequence-specific transcription by RNAP core.

7.3. Material and methods

Plasmids. Plasmid used for protein purifications were pRL663, pRL706 and pRL706 derivatives constructed using site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit; Agilent, Inc.). Plasmid used for in vivo studies were pRL663-β'WT; 3xFLAG or pRL663-β'[R275Q, E295K]; 3xFLAG pRL663^{WT}-3x-flagged.

Proteins. Proteins were prepared from E. coli strain XE54 (Tang et al., 1994) transformed with plasmids pRL663 β'WT, pRL663 β'[R275Q, E295K], pRL706 or pRL706-βD446A, using procedures described in ref (Mukhopadhyay et al., 2003b) and in chapter 1. Cations were applied during pooling of fractions to avoid any contamination from sigma.

Analysis of TSS: pGENOME plasmid DNA library was prepared as illustrated in Fig 1. The in vitro and in vivo assays, were analogous to the procedure describe in chapter 3. More information can be obtained upon request.

Chapter 8

Conclusions and additional thoughts

β-subunit of RNAP-core interacts with the downstream-edge of the transcription-bubble nontemplate-strand corresponding to positions -4 to +2, "core recognition element" (CRE).

Our findings establish that RNAP-CRE interactions are sequence-specific at 3-out-of-6 positions and define the RNAP residues and RNAP-DNA interactions responsible for sequence-specific RNAP-CRE interactions. RNAP-CRE sequence-specificity impacts thermodynamics, kinetics, and conformation. As RNAP core and transcription bubbles are present during all stages of transcription, we proposed that RNAP-CRE interactions could occur and impact all stages of transcription. In subsequent works, we confirmed that RNAP-CRE interactions play functional roles in three key stages of transcription initiation (open complex formation, transcription start-site selection, and abortive transcription) and key stages of transcription elongation (translocation bias and sequence-dependent pausing). The results further indicate that RNAP-CRE interactions occur in, and play functional roles in, all three domains of life: bacteria, archaea, and eukaryotes.

The observations that RNAP-core makes sequence-specific interaction with DNA led us to investigate if RNAP-core is capable of performing sequence-specific transcription in the absence of sigma factor. Our results indicate that RNAP-core can indeed perform sequence-specific transcription at sequences that seems to include two features: a 4 bp AT-rich segment followed immediately by a TG motif, which looks similar to the consensus sequence of CRE. We propose that direct sequence-specific interactions between the

RNAP-core and the CRE, given a specific sequence context for the latter (AT-richness), are accountable for the sequence-specific sigma-independent transcription events.

One of the implications of this thesis could be in the understanding of the architecture of primordial promoter sequences. The core subunits of multi-subunit RNAPs show high degree of homology across three domains of life. However, the initiation factors are not conserved across three domains of life. Based on these observations, it has been proposed that in the last universal common ancestor (LUCA), RNAP initiated transcription non-sequence-specifically without the requirement of any basal transcription factor and possibly at AT-rich DNA region (Werner & Grohmann, 2011a). However, to date, there has not been any experimental evidence for this hypothesis. Our data presented in this thesis seems to agree with this hypothesis. The observations are that the RNAP core makes sequence-specific interactions with the CRE, which is located in the core promoter regions, and that these interactions facilitate promoter melting, that these interactions can act as one of the determinants of the TSS, and that these interactions are conserved across three domains of life. Additionally, it was illustrated in this thesis that the RNAP core is capable of performing sequence-specific transcription in the absence of sigma factor in a manner that correlates with the presence of an AT-rich region followed by a TG-motif (which resemble a TG_{CRE} consensus sequence). Based on this observation, we propose that in LUCA, proto-promoters were composed of an AT-rich region followed by a consensus CRE element, the TG-motif. RNAP-CRE interactions enabled sequence-specific recognition of proto-promoters and assisted the RNAP in DNA unwinding in the absence of any initiation factor. A major drawback from this hypothesis is that the position of the CRE seems to be off by one nucleotide. One explanation for this might be that the DNA

conformation in sigma-independent initiation complex is different from that during sigma-dependent transcription (probably forming L-shaped DNA conformation compared to the situation where sigma is present and additional interactions are established at -35 element). Therefore, it would be essential for future work to study the effect of CRE on sigma-independent transcription. In a second note, it might be possible that in a sigma-independent transcription complex, the most preferred bases for loading to the active sites are a T at position -1 and a G at position +1. Therefore, it would be reasonable to assume that sigma-independent transcription originate, at least partly, from active site preference, independent of the effect of RNAP-CRE interaction. In a third note, it is worth mentioning that the TG dinucleotide demonstrate a high propensity to bend DNA and act as flexible hinge compared to other dinucleotide (Olson, Gorin et al., 1998). One might postulate that a local distortion at DNA duplex with a TG sequence adjacent to a low melting AT-rich region could cooperatively assist in DNA unwinding. Therefore, it would be reasonable to assume that sigma-independent transcription originate, at least partly, from an intrinsic property of the DNA helix opening, independent of the effect of RNAP-CRE interaction. These three possibilities are not mutually exclusive and could occur together.

A second implication of this thesis could be in the understanding of the persistence state. Persister cells are a fraction of cells in a population that although are genetically identical, nonetheless, are resistant to antibiotic by entering into a non-growing state. This phenotypic switching between growing and non-growing states, has been linked to stochasticity in gene expression which give rise to a preexisting heterogeneous population (Balaban, Merrin et al., 2004). The origin of this stochasticity is not very clear. Some evidence shows that the persistence state is linked to the stringent response, which are

activated during stress (Lebeaux, Ghigo et al., 2014). During the stringent response, the rRNA synthesis is shutdown. Since about 75% of transcription is directed toward rrn operons, upon the suppression of rRNA transcription, a pool of RNAP core becomes available (Mauri & Klumpp, 2014). It is possible that under this condition sigma-independent transcription could become available in vivo. Given the stochastic and low-fidelity nature of the sigma-independent transcription, and large number of cells in a population undergoing stringent response, it would be reasonable to assume that some cells might by chance make a repertoire of new products. Some of these new products could provide phenotypic resistance to antibiotic, or give rise to the stochastic state switching to overcome susceptibility to antibiotic.

References

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

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography 66: 213-21

Aivazashvili VA, Bibilashvili R, Vartikian RM, Kutateladze TA (1981) [Effect of the primary structure of RNA on the pulse character of RNA elongation in vitro by Escherichia coli RNA polymerase: a model]. Molekuliarnaia biologiia 15: 915-29

Aoyama T, Takanami M (1985) Essential structure of E. coli promoter II. Effect of the sequences around the RNA start point on promoter function. Nucleic acids research 13: 4085-96

Arthur TM, Anthony LC, Burgess RR (2000) Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase. The Journal of biological chemistry 275: 23113-9

Artsimovitch I, Landick R (2000a) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci USA 97 SRC - GoogleScholar: 7090-7095

Artsimovitch I, Landick R (2000b) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci USA 97: 7090-5

Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. eLife 4

Bai L, Shundrovsky A, Wang MD (2004) Sequence-dependent kinetic model for transcription elongation by RNA polymerase. Journal of molecular biology 344: 335-49

Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. Science 305: 1622-5

Buc H, Strick T (2009) RNA polymerases as molecular motors. RSC Pub., Cambridge

Burgess RR (1969) The subunit structure of RNA polymerase. In

Burgess RR, Travers AA, Dunn JJ, Bautz EK (1969) Factor stimulating transcription by RNA polymerase. Nature 221: 43-6

Busby S, Ebright R (1994a) Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79 SRC - GoogleScholar: 743-746

Busby S, Ebright RH (1994b) Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79: 743-6

Camacho A, Salas M (1999) Effect of mutations in the “extended-10” motif of three Bacillus subtilis σ A-RNA polymerase-dependent promoters. Journal of molecular biology 286 SRC - GoogleScholar: 683-693

Chamberlin M, Berg P (1962) DEOXYRIBONUCLEIC ACID-DIRECTED SYNTHESIS OF RIBONUCLEIC ACID BY AN ENZYME FROM ESCHERICHIA COLI. Proceedings of the National Academy of Sciences of the United States of America 48: 81-94

Chan CL, Landick R (1993) Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. Journal of molecular biology 233: 25-42

Chessin H, Summers WC (1970) Initiation by RNA polymerase on UV or x-ray damaged T7 DNA. Biochem Biophys Res Commun 38: 40-5

Cheung A, Cramer P (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature 471: 249-253

Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469: 368-73

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome research 14: 1188-90

Dalal R, Larson M, Neuman K, Gelles J, Landick R, Block S (2006) Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. Molecular cell 23 SRC - GoogleScholar: 231-239

Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C (2014a) Molecular mechanisms of transcription through single-molecule experiments. Chemical reviews 114: 3203-23

Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C (2014b) Molecular Mechanisms of Transcription through Single-Molecule Experiments. Chemical reviews

Dangkulwanich M, Ishibashi T, Liu S, Kireeva ML, Lubkowska L, Kashlev M, Bustamante CJ (2013) Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. eLife 2: e00971

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640-5

Dausse JP, Sentenac A, Fromageot P (1972) Interaction of RNA polymerase from Escherichia coli with DNA. Influence of DNA scissions on RNA-polymerase binding and chain initiation. Eur J Biochem 31: 394-404

Drennan A, Kraemer M, Capp M, Gries T, Ruff E, Sheppard C, Wigneshweraraj S, Artsimovitch I, Record MT, Jr. (2012) Key roles of the downstream mobile jaw of Escherichia coli RNA polymerase in transcription initiation. Biochemistry 51: 9447-59

Ebright RH (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. Journal of molecular biology 304: 687-98

Ederth J, Artsimovitch I, Isaksson LA, Landick R (2002) The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. The Journal of biological chemistry 277: 37456-63

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta crystallographica Section D, Biological crystallography 66: 486-501

Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, Ebright RH, Gourse RL (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes & development 13: 2134-47

Feklistov A, Darst SA (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. Cell. 147 SRC - GoogleScholar: 1257-1269

Feklistov A, Sharon BD, Darst SA, Gross CA (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. Annual review of microbiology 68: 357-76

Geiduschek EP, Tocchini-Valentini GP, Sarnat MT (1964) ASYMMETRIC SYNTHESIS OF RNA IN VITRO: DEPENDENCE OF DNA CONTINUITY AND CONFORMATION. Proceedings of the National Academy of Sciences of the United States of America 52: 486-93

Goldman SR, Nair NU, Wells CD, Nickels BE, Hochschild A (2015) The primary sigma factor in Escherichia coli can access the transcription elongation complex from solution in vivo. eLife 4

Goldman SR, Sharp JS, Vvedenskaya IO, Livny J, Dove SL, Nickels BE (2011) NanoRNAs prime transcription initiation in vivo. Molecular cell 42: 817-25

Haas BJ, Chin M, Nusbaum C, Birren BW, Livny J (2012) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC genomics 13: 734

Haugen SP, Ross W, Gourse RL (2008a) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol 7: 507-519

Haugen SP, Ross W, Gourse RL (2008b) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol 6: 507-19

Hawley DK, McClure WR (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic acids research 11: 2237-55

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

Herbert K, Porta A, Wong B, Mooney R, Neuman K, Landick R, Block S (2006a) Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell 125: 1083-1094

Herbert KM, Greenleaf WJ, Block SM (2008) Single-molecule studies of RNA polymerase: motoring along. Annual review of biochemistry 77: 149-76

Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, Block SM (2006b) Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell 125: 1083-94

Hinkle DC, Chamberlin MJ (1972a) Studies of the binding of Escherichia coli RNA polymerase to DNA. I. The role of sigma subunit in site selection. Journal of molecular biology 70: 157-85

Hinkle DC, Chamberlin MJ (1972b) Studies of the binding of Escherichia coli RNA polymerase to DNA. II. The kinetics of the binding reaction. Journal of molecular biology 70: 187-95

Hsu LM (1996) Quantitative parameters for promoter clearance. Methods in enzymology 273: 59-71

Hsu LM (2009) Monitoring abortive initiation. Methods 47: 25-36

Hurwitz J (2005) The discovery of RNA polymerase. The Journal of biological chemistry 280: 42477-85

Hurwitz J, Furth JJ, Anders M, Ortiz PJ, August JT (1961) The enzymatic incorporation of ribonucleotides into RNA and the role of DNA. Cold Spring Harb Symp Quant Biol 26: 91-100

James E, Liu M, Sheppard C, Mekler V, Camara B, Liu B, Simpson P, Cota E, Severinov K, Matthews S, Wigneshweraraj S (2012) Structural and mechanistic basis for the inhibition of Escherichia coli RNA polymerase by T7 Gp2. Molecular cell 47: 755-66

Jeong W, Kang C (1994) Start site selection at lacUV5 promoter affected by the sequence context around the initiation sites. Nucleic acids research 22: 4667-72

Jorgensen SE, Buch LB, Nierlich DP (1969) Nucleoside triphosphate termini from RNA synthesized in vivo by Escherichia coli. Science 164: 1067-70

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

Kireeva ML, Domecq C, Coulombe B, Burton ZF, Kashlev M (2011) Interaction of RNA polymerase II fork loop 2 with downstream non-template DNA regulates transcription elongation. The Journal of biological chemistry 286 SRC - GoogleScholar: 30898-30910

Kireeva ML, Kashlev M (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America 106: 8900-5

Knaus R, Bujard H (1990) Principles Governing the Activity of E. coli Promoters Nucleic Acids and Molecular Biology pp 4 SRC - GoogleScholar: 110-122

Ko J, Heyduk T (2014) Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. Biochem J 463: 135-44

Kyzer S, Ha K, Landick R, Palangat M (2007a) Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. The Journal of biological chemistry 282: 19020-19028

Kyzer S, Ha KS, Landick R, Palangat M (2007b) Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. The Journal of biological chemistry 282: 19020-8

Landick R (2009) Transcriptional pausing without backtracking. Proceedings of the National Academy of Sciences of the United States of America 106: 8797-8

Landick R, Wang D, Chan CL (1996) Quantitative analysis of transcriptional pausing by Escherichia coli RNA polymerase: his leader pause site as paradigm. Methods in enzymology 274: 334-53

Larson MH, Landick R, Block SM (2011) Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. Molecular cell 41: 249-62

Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science 344: 1042-7

Lebeaux D, Ghigo JM, Beloin C (2014) Biofilm-related infections: bridging the gap between clinical management and fundamental aspects of recalcitrance toward antibiotics. Microbiology and molecular biology reviews : MMBR 78: 510-43

Lee DN, Phung L, Stewart J, Landick R (1990) Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. The Journal of biological chemistry 265: 15145-53

Lewis DE, Adhya S (2004) Axiom of determining transcription start points by RNA polymerase in Escherichia coli. Molecular microbiology 54: 692-701

Linn TG, Greenleaf AL, Shorenstein RG, Losick R (1973) Loss of the sigma activity of RNA polymerase of Bacillus subtilis during sporulation. Proceedings of the National Academy of Sciences of the United States of America 70: 1865-9

Liu J, Turnbough CL, Jr. (1994) Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in Escherichia coli. Journal of bacteriology 176: 2938-45

Losick R, Chamberlin M, Cold Spring Harbor Laboratory of Quantitative Biology. (1976) RNA polymerase. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Maitra U, Hurwitz H (1965) The role of DNA in RNA synthesis, IX. Nucleoside triphosphate termini in RNA polymerase products. Proceedings of the National Academy of Sciences of the United States of America 54: 815-22

Mauri M, Klumpp S (2014) A model for sigma factor competition in bacterial cells. PLoS computational biology 10: e1003845

McClure WR, Hawley DK, Youderian P, Susskind MM (1983) DNA determinants of promoter selectivity in Escherichia coli. Cold Spring Harb SympQuant Biol 47: 477-481

SRC - GoogleScholar

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

Mekler V, Minakhin L, Severinov K (2011a) A critical role of downstream RNA polymerase-promoter interactions in the formation of initiation complex. The Journal of biological chemistry 286: 22600-8

Mekler V, Pavlova O, Severinov K (2011b) Interaction of Escherichia coli RNA polymerase sigma70 subunit with promoter elements in the context of free sigma70, RNA polymerase holoenzyme, and the beta'-sigma70 complex. The Journal of biological chemistry 286: 270-9

Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH (2001) Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell 106: 453-63

Mukhopadhyay J, Mekler V, Kortkhonjia E, Kapanidis AN, Ebright YW, Ebright RH (2003a) Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. Methods in enzymology 371: 144-59

Mukhopadhyay J, Mekler V, Kortkhonjia E, Kapanidis AN, Ebright YW, Ebright RH (2003b) Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. Method Enzymo 371: 144-159

Mulligan ME, Hawley DK, Entriken R, McClure WR (1984) Escherichia colipromoter sequences predict in vitro RNA polymerase selectivity. Nucleic Acids Research 12: 789-800

SRC - GoogleScholar

Murakami KS, Darst SA (2003) Bacterial RNA polymerases: the wholo story. Current opinion in structural biology 13: 31-9

Mustaev A, Roberts J, Gottesman M (2017) Transcription elongation. Transcription: e1289294

Naryshkin N, Revyakin A, Kim Y, Mekler V, Ebright RH (2000) Structural organization of the RNA polymerase-promoter open complex. Cell 101: 601-11

Neuman K, Abbondanzieri E, Landick R, Gelles J, Block S (2003a) Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell 115 SRC - GoogleScholar: 437-447

Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM (2003b) Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell 115: 437-47

Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A (2004) The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. Nature structural & molecular biology 11: 544-50

Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proceedings of the National Academy of Sciences of the United States of America 95: 11163-8

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol 276: 307-326

Pribnow D (1975a) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proceedings of the National Academy of Sciences of the United States of America 72: 784-8

Pribnow D (1975b) Nucleotide sequence of and RNA polymerase binding site at an early T7 promoter. Proc Nat Acad Sci USA 72: 784-788

SRC - GoogleScholar

Qi F, Turnbough CL, Jr. (1995) Regulation of codBA operon expression in Escherichia coli by UTP-dependent reiterative transcription and UTP-sensitive transcriptional start site switching. Journal of molecular biology 254: 552-65

Revyakin A, Ebright R, Strick T (2004) Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. Proc Natl Acad Sci USA 101 SRC - GoogleScholar: 4776-4780

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

Robb NC, Cordes T, Hwang LC, Gryte K, Duchi D, Craggs TD, Santoso Y, Weiss S, Ebright RH, Kapanidis AN (2013) The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. Journal of molecular biology 425: 875-85

Roberts JW (1969a) Promoter mutation in vitro. Nature 223: 480-2

Roberts JW (1969b) Termination factor for RNA synthesis. Nature 224: 1168-74

Roberts JW (1970) Initiation and termination of RNA synthesis in vitro. In

Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, Severinov K, Gourse RL (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science 262: 1407-13

Ruff EF, Record MT, Jr., Artsimovitch I (2015) Initial events in bacterial transcription initiation. Biomolecules 5: 1035-62

Saecker RM, Record MT, Jr., Dehaseth PL (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. Journal of molecular biology 412: 754-71

Severinov K, Mooney R, Darst SA, Landick R (1997) Tethering of the large subunits of Escherichia coli RNA polymerase. The Journal of biological chemistry 272: 24137-40

Shorenstein RG, Losick R (1973) Purification and properties of the sigma subunit of ribonucleic acid polymerase from vegetative Bacillus subtilis. The Journal of biological chemistry 248: 6163-9

Shultzaberger RK, Chen Z, Lewis KA, Schneider TD (2007) Anatomy of Escherichia coli sigma70 promoters. Nucleic acids research 35: 771-88

Skancke J, Bar N, Kuiper M, Hsu LM (2015) Sequence-Dependent Promoter Escape Efficiency Is Strongly Influenced by Bias for the Pretranslocated State during Initial Transcription. Biochemistry 54: 4267-75

Sorensen KI, Baker KE, Kelln RA, Neuhard J (1993) Nucleotide pool-sensitive selection of the transcriptional start site in vivo at the Salmonella typhimurium pyrC and pyrD promoters. Journal of bacteriology 175: 4137-44

Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. Proceedings of the National Academy of Sciences of the United States of America 90: 6498-502

Strobel EJ, Roberts JW (2015) Two transcription pause elements underlie a sigma70-dependent pause cycle. Proceedings of the National Academy of Sciences of the United States of America 112: E4374-80

Sugimoto N, Nakano S, Katoh M, Matsumura A, Nakamuta H, Ohmichi T, Yoneyama M, Sasaki M (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry 34: 11211-6

Sugiura M, Okamoto T, Takanami M (1969) Starting nucleotide sequences of RNA synthesized on the replicative form DNA of coliphage fd. Journal of molecular biology 43: 299-315

Tadigotla VR, D OM, Sengupta AM, Epshtein V, Ebright RH, Nudler E, Ruckenstein AE (2006) Thermodynamic and kinetic modeling of transcriptional pausing. Proceedings of the National Academy of Sciences of the United States of America 103: 4439-44

Tang H, Severinov K, Goldfarb A, Fenyo D, Chait B, Ebright RH (1994) Location, structure, and function of the target of a transcriptional activator protein. Genes & development 8: 3058-67

Travers AA (1980) Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. Journal of bacteriology 141: 973-976

Travers AA, Burgessrr (1969) Cyclic re-use of the RNA polymerase sigma factor. Nature 222: 537-40

Tu AH, Turnbough CL, Jr. (1997) Regulation of upp expression in Escherichia coli by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription. Journal of bacteriology 179: 6665-73

Turnbough CL, Jr. (2008) Regulation of bacterial gene expression by the NTP substrates of transcription initiation. Molecular microbiology 69: 10-4

Turnbough CL, Jr., Switzer RL (2008) Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. Microbiology and molecular biology reviews : MMBR 72: 266-300, table of contents

Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography 30: 1022-1025

Vogt V (1969) Breaks in DNA stimulate transcription by core RNA polymerase. Nature 223: 854-5

Voskuil MI, Chambliss GH (2002) The TRTGn motif stabilizes the transcription initiation open complex. Journal of molecular biology 322: 521-532

Vvedenskaya I (2015) Massively Systematic Transcript Readout (MASTER): Transcription Initiaion. Molecular Cell in press

Vvedenskaya IO, Goldman SR, Nickels BE (2015) Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends. Methods in molecular biology 1276: 211-28

Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE (2014) Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. Science 344: 1285-9

Vvedenskaya IO, Zhang Y, Goldman SR, Valenti A, Visone V, Taylor DM, Ebright RH, Nickels BE (2015) Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. Molecular cell 60: 953-65

Walker KA, Mallik P, Pratt TS, Osuna R (2004) The Escherichia coli Fis promoter is regulated by changes in the levels of its transcription initiation nucleotide CTP. The Journal of biological chemistry 279: 50818-28

Walker KA, Osuna R (2002) Factors affecting start site selection at the Escherichia coli fis promoter. Journal of bacteriology 184: 4783-91

Weixlbaumer A, Leon K, Landick R, Darst SA (2013) Structural basis of transcriptional pausing in bacteria. Cell 152: 431-41

Werner F (2008) Structural evolution of multisubunit RNA polymerases. Trends in microbiology 16: 247-50

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

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

Wheeler AR, Woody AY, Woody RW (1987) Salt-dependent binding of Escherichia coli RNA polymerase to DNA and specific transcription by the core enzyme and holoenzyme. Biochemistry 26: 3322-30

Wilson HR, Archer CD, Liu JK, Turnbough CL, Jr. (1992) Translational control of pyrC expression mediated by nucleotide-sensitive selection of transcriptional start sites in Escherichia coli. Journal of bacteriology 174: 514-24

Winkelman JT, Chandrangsu P, Ross W, Gourse RL (2016a) Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of E. coli ribosomal RNA promoters. Proceedings of the National Academy of Sciences of the United States of America 113: E1787-95

Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, Gourse RL, Ebright RH, Nickels BE (2016b) Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science 351: 1090-3

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton

EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography 67: 235-42

Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N (2011) A new basal promoter element recognized by RNA polymerase core enzyme. EMBO J 30: 3766-75

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell 98: 811-24

Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, Ebright RH (2012) Structural basis of transcription initiation. Science 338: 1076-80

Zhou J, Ha KS, La Porta A, Landick R, Block SM (2011) Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. Molecular cell 44: 635-46