DISSERTATION

FACTOR DEPENDENT ARCHAEAL TRANSCRIPTION TERMINATION

Submitted by

Julie Walker

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Thomas J. Santangelo

Tai Montgomery
Laurie Stargell
Tingting Yao

ABSTRACT

FACTOR DEPENDENT ARCHAEAL TRANSCRIPTION TERMINATION

RNA polymerase activity is regulated by nascent RNA sequences, DNA template sequences and conserved transcription factors. Transcription factors regulate the activities of RNA polymerase (RNAP) at each stage of the transcription cycle: initiation, elongation, and termination. Many basal transcription factors with common ancestry are employed in eukaryotic and archaeal systems that directly bind to RNAP and influence intramolecular movements of RNAP and modulate DNA or RNA interactions. We describe and employ a flexible methodology to directly probe and quantify the binding of transcription factors to the archaeal RNAP *in vivo*. We demonstrate that binding of the conserved and essential archaeal transcription factor TFE to the archaeal RNAP is directed, in part, by interactions with the RpoE subunit of RNAP. As the surfaces involved are conserved in many eukaryotic and archaeal systems, the identified TFE-RNAP interactions are likely conserved in archaeal-eukaryal systems and represent an important point of contact that can influence the efficiency of transcription initiation.

While many studies in archaea have focused on elucidating the mechanism of transcription initiation and elongation, studies on termination were slower to emerge. Transcription factors promoting initiation and elongation have been characterized in each Domain but transcription termination factors have only been identified in bacteria and eukarya. Here we characterize the first archaeal termination factor (termed Eta) capable of disrupting the transcription elongation complex, detail the rate of and requirements for Eta-mediated transcription termination and describe a role for Eta in transcription termination *in vivo*. Eta-mediated transcription termination is energy-dependent, requires upstream DNA sequences and disrupts transcription elongation complexes to release the nascent RNA to solution. Deletion of TK0566 (encoding Eta) is possible, but results in slow growth and renders cells sensitive to

DNA damaging agents. Structure-function studies reveal that the N-terminal domain of Eta is not necessary for Eta-mediated termination *in vitro*, but *Thermococcus kodakarensis* cells lacking the N-terminal domain exhibit slow growth compared to parental strains. We report the first crystal structure of Eta that will undoubtedly lead to further structure-function analyses. The results obtained argue that the mechanisms employed by termination factors in archaea, eukarya, and bacteria to disrupt the transcription elongation complex may be conserved and that Eta stimulates release of stalled or arrested transcription elongation complexes.

ACKNOWLEDGEMENTS

Throughout the past five years there are many people that have played a critical role in helping me become the scientist and person I am today. Firstly, I would like to thank my advisor Tom Santangelo for consistently pushing me to critically think and always challenging me with new ideas. My PhD would not have been nearly as successful without his infectious enthusiasm and motivation for science and consistent advice. Not only did he teach me how to become the scientist I am today he also allowed me to develop my network as a scientist through presenting at conferences and encouraging collaborations with other academic and industrial labs. The connections I was able to make through these interactions will facilitate my career for years to come. For that, and for encouraging independence, enthusiasm, and creativity during my time in his lab, I will always be grateful.

Everyone I have encountered at CSU has always been very eager to help. My labmates were always willing to help me with experiments, listen to my presentations, and have exciting scientific discussions. My time at CSU would not have been nearly as enjoyable without them. I specifically want to thank the undergraduates who I worked closely with: Erin Lynch, Olivia Luyties, and Jocelyn Selena. Erin Lynch's consistent excitement about her project was always refreshing and Olivia Luyties' development into a determined and passionate scientist over the past four years has been enjoyable to watch. I am very proud of them and I know they will continue to be excellent scientists in the future.

I also need to thank a fellow graduate student, Lexi Gehring. My PhD would not have been the same without her friendship. My lunches and coffee breaks would have been very dull without her there. I really appreciate her for always listening to me and always having a good laugh.

I am very grateful for my student advisory committee. The participation of Tai Montgomery, Laurie Stargell, and Tingting Yao in my development as a scientist was very

valuable. They guided me throughout the past five years and always had excellent suggestions to improve my work.

I also need to thank my family and friends. Without the constant support and encouragement of my parents and sister I would not nearly as successful. They always offered an ear to listen and the best advice anyone could give: “to always follow your heart and do what feels right for you.” My parents have always been my biggest cheerleaders and always want the first copy of my published manuscripts.

Finally, I would like to thank Adam Mankin, he has always listened to both my excitement and frustrations over the past five years. He always gets excited when I talk to him about “my science” and more importantly he always put things in perspective when I get frustrated. Throughout my PhD I referred to him as my public speaking coach as he listened to many of my presentations and always gave the best advice on how to calm my nerves and present well.

I was very lucky to have the support of so many people throughout my PhD career and these acknowledgements only begin to explain the gratitude I have towards everyone who has helped me along the way.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION[1]

1.1 Archaeal Transcription Regulation

RNA polymerase (RNAP) is a well-conserved, multi-subunit essential enzyme that transcribes DNA to generate RNA in all cells. Although RNA synthesis is carried out by RNAP, the activities of RNAP during each phase of transcription are subject to basal and regulatory transcription factors. Substantial differences in transcription regulatory strategies exist in the three Domains: *Bacteria, Archaea, and Eukarya*. Only a single transcription factor (NusG or Spt5) is universally conserved (1, 2), and the roles of many archaeal-encoded factors have not been evaluated. *Archaea* are reliant on a transcription apparatus that is homologous to the eukaryotic transcription machinery; similarities include additional RNAP subunits that form a discrete subdomain of RNAP (3, 4), as well as basal transcription factors that direct transcription initiation and elongation (5–10). The shared homology of archaeal-eukaryotic transcription components aligns with the shared ancestry of *Archaea* and *Eukarya*, and this homology often is exclusive of *Bacteria* (Table 1.1).

The archaeal transcription apparatus is most commonly summarized as a simplified version of the eukaryotic machinery. In some respects this is true, as homologs of only a few eukaryotic transcription factors are encoded in archaeal genomes and archaeal transcription *in vitro* can be supported by just a handful of transcription factors (8, 11). However, much regulatory activity in eukaryotes is devoted to post-translational modifications of chromatin, RNAP, and transcription

[1] Chapter One, an introduction to this dissertation was published under: Gehring AM, Walker JE, Santangelo TJ (2016) Transcription Regulation in Archaea. *J Bacteriol*. Parts of the review have been omitted or expanded where appropriate.

Santangelo TJ, Gehring AM, and I conceived the content and co-wrote this manuscript.

Table 1.1. Evolutionary conservation of RNAP subunits and transcription

	Bacter	Archaea	Eukaryotes			Plants	
			RNAPII	RNAPIII	RNAPI	RNAPIV	RNAPV
RNAP subunits	β'	Rpo1 (A)	RPB1	C160	A190	NRPD1	NRPE1
	β	Rpo2 (B)	RPB2	C128	A135	NRPD/E2	NRPD/E2
	α	Rpo3 (D)	RPB3	AC40	AC40	RPB3(+1)	RPB3(+1)
	α	Rpo11 (L)	RPB11	AC19	AC19	RPB11	RPB11
	ω	Rpo6 (K)	RPB6	RPB6	RPB6	RPB6(+1)	RPB6(+1)
		Rpo5 (H)	RPB5	RPB5	RPB5	RPB5(+3)	NRPE5
		Rpo8* (G)	RPB8	RPB8	RPB8	RPB8(+1)	RPB8(+1)
		Rpo10 (N)	RPB10	RPB10	RPB10	RPB10	RPB10
		Rpo12 (P)	RPB12	RPB12	RPB12	RPB12	RPB12
		Rpo4 (F)	RPB4	C17	A14	NRPD/E4	NRPD/E4
		Rpo7 (E)	RPB7	C25	A43	NRPD7(+1)	NRPE7
			RPB9	C11	A12	NRPD9b	RPB9
		Rpo13*					
transcription factors			*TFIIFα*	C53	A49		
			TFIIFβ	C37	A34.5		
		TFEα	*TFIIEα*	C82			
		TFEβ/C34[*]	*TFIIEβ*	C34			
				C31			
		TBP	*TBP*	*TBP*	*TBP*		
		TFB	*TFIIB*	*Brf-1*	*TAF1B*		
			TFIIA				
			TFIIH				
		TFS	*TFIIS*		*TFIIS*		
		Spt4	*Spt4*		*Spt4*		
	NusG	*Spt5*	*Spt5*		*Spt5*		
	NusA	*NusA*					
	rho						
	Sigma						
	Gre						
		Eta					

Adapted from F. Werner. (2012).

factors, and this complexity seemingly does not transfer to the *Archaea* where few post-translational modifications or chromatin-imposed regulation events are currently known. The ostensible simplicity of archaeal transcription is under constant revision as more detailed examinations of archaeal-encoded factors become possible through increasingly sophisticated *in vivo* and *in vitro* techniques.

1.2 The Archaeal Transcription Cycle

Transcription is highly regulated, and the transcription cycle is typically demarcated into three phases: initiation, elongation, and termination (Figure 1.1) (12–16). An abbreviated and overall introduction to this cycle is presented first, with sections below detailing the activities of RNAP and associated factors during each stage of transcription. Briefly, archaeal transcription initiation requires that RNAP be directed to promoter sequences defined by the binding of TATA binding protein (TBP) and transcription factor B (TFB). TBP, TFB, and RNAP are sufficient to generate a single-stranded section of DNA (the transcription bubble) and feed the template strand into the bipartite active center of RNAP (7, 10, 17). RNAP can initiate transcript synthesis *de novo*, and continued synthesis then competes with favorable promoter and initiation factor contacts until promoter escape can be achieved. Release of RNAP from the initiating factors classically defines the end of initiation, although in reality no clear boundary separates the last stages of initiation from the early stages of elongation. Although TFB and TBP are necessary and sufficient to permit promoter-directed transcription initiation, a third conserved factor, transcription factor E (TFE), can also assist in transcription initiation and leaves the promoter with RNAP during the early stages of transcript elongation (18–20). Transition to a stable, long-lived elongation complex is believed to involve internal rearrangements of RNAP. This transition involves the exchange of initiation factors for stably bound elongation factors that monitor RNA synthesis for accuracy, respond to regulatory DNA sequences, react to regulatory inputs of more transiently associated transcription factors, and influence processivity of RNAP (21).

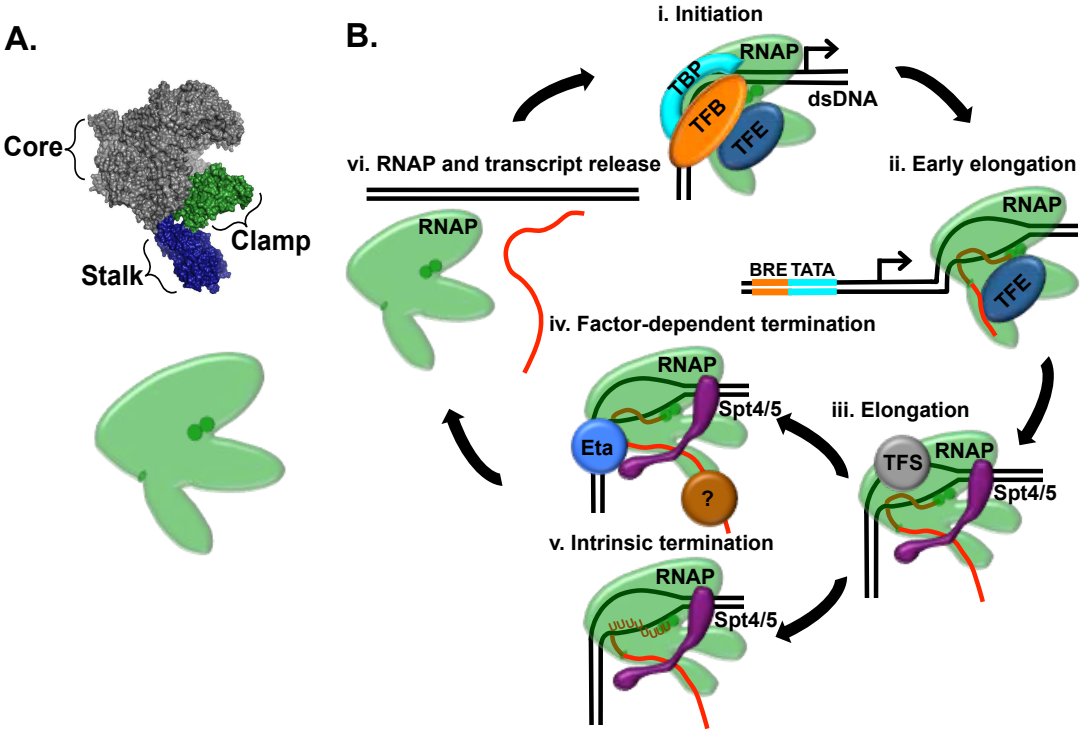


Figure 1.1. The archaeal transcription cycle. A. The euryarchaeal RNA polymerase crystal structure from *Thermococcus kodakarensis* (PDB ID no. 4QIW) is shown in surface representation. The clamp and stalk domains are highlighted. A simplified cartoon structure of RNA polymerase is shown below this in light green; the bipartite active site and RNA exit channel are highlighted in dark green. **B.** Steps in the transcription cycle. (i) RNAP is recruited to the promoter by transcription factors TFB, TFE, and TBP during transcription initiation. (ii) RNAP escapes the promoter, and early elongation begins with TFE bound to RNAP. (iii) TFE is replaced by elongation factor Spt5 during elongation. (iv) Factor-dependent termination occurs in archaea and is mediated by Eta (blue). A second termination factor (brown) is likely encoded in archaeal genomes but remains to be identified. (v) Intrinsic termination sequences are characterized by a run of T's on the nontemplate strand. (vi) The transcript is released, and RNAP is recycled for another round of transcription.

Elongation is, in general, very stable, but specific sequences can lower the overall energy of the transcription elongation complex, permitting either spontaneous intrinsic or factor-assisted termination (22–24). An archaeal transcription termination factor, termed Eta, recognizes arrested RNAPs to mediate transcription termination (16). Transcription termination results in release of both the transcript and RNAP from the DNA template.

1.3 Regulation of Transcription Initiation

TBP and TFB are the only transcription factors required for *in vitro* transcription under optimized conditions, and TFE has been shown to assist promoter opening when conditions are suboptimal (Figure 1.1B; panel i) (19). *In vivo* studies have shown that *Archaea* must retain at least one gene encoding TBP and one gene encoding TFB, although many archaeal species encode multiple TBP and TFB isoforms (6, 11, 25–28). Some differences in promoter sequence preferences and protein pairing have been noted in TBP-TFB isoform pairs (29–33), but these minor differences are not on par with the clear but not always radical promoter sequence differences noted for alternative σ factors in bacterial transcription (34). TFE also appears essential, and it is currently unclear if this essentiality is due to necessary activities during transcription initiation or some other role in the transcription cycle (35, 36).

All three of the aforementioned transcription factors have close eukaryotic homologs: archaeal TBPs are nearly identical to eukaryotic TBPs (37); archaeal TFB proteins are homologous to eukaryotic transcription factor IIB (TFIIB) proteins (38), with homology also seen with the Pol III initiation factor BRF1 (39) and Pol I initiation factor Rrn7/TAF1B (40); and archaeal TFE proteins are homologous to the N-terminal half of the eukaryotic alpha subunit of TFIIE, or TFIIEα, and very recent evidence identified a separate homolog in some lineages to the eukaryotic beta subunit, TFIIEβ (20). TBP is needed to recognize the TATA box, bend the DNA, and recruit TFB; its role had therefore been deemed equivalent to the role of eukaryotic TBPs (10, 38). Total internal-reflection fluorescence–fluorescence resonance energy transfer

measurements now detail differences in the activities of archaeal and eukaryotic TBPs, despite the nearly identical three-dimensional folds of these factors (6, 41). In some cases, archaeal TBPs require the co-binding of TFB to stably bind and bend the promoter DNA (42–44). It is speculated that different promoter sequences may be regulated by different TFB-TBP pairs based on the interdependence, or lack thereof, of cooperative DNA bending for establishing a stable platform for RNAP recruitment. Recent studies suggest that select isoforms of TFB and TBP can result in differences in transcription output, but further studies will be needed to determine if these effects on such preliminary steps of transcription initiation are a direct mode of regulation resulting in phenotypic differences (30, 45).

In contrast to eukaryotic transcription, archaeal promoter opening is not an energy-dependent process (7). Therefore, TBP and TFB alone are capable of assisting RNAP in the formation of the transcription bubble. In all *Archaea*, TFB is responsible for stabilizing the TBP-bound DNA complex and, together, this bipartite protein platform recruits RNAP (46), but how these molecular interactions melt the DNA is still unresolved. Reconstructions and analyses of open complexes using archaeal components reveal an overall architecture of the open promoter complex and provide the first placement of the nontemplate strand within the complex (46). TBP and TFB are located closer to RNAP than would be the case for eukaryotic promoters, and this proximity may provide more intimate contacts that collectively provide the energy to open the promoter DNA. The tight network of interactions in the archaeal open complex may torsionally strain the DNA, and melting is likely to relieve this strain and result in open complex formation.

Several new insights into TFE activity and evolution have been recently described. The archaeal TFE had previously been characterized as a monomer and as a homologue of the alpha subunit of eukaryotic TFIIE, termed TFIIEα (19, 47, 48). Eukaryotic TFIIE is a heterodimeric complex of TFIIEα and TFIIEβ, but archaeal genomes had previously only been shown to encode a homologue of the alpha subunit (49, 50). Eukaryotic RNAPs differ in their requirements for initiation, with RNAP III incorporating homologues of several RNAP II initiation

factors as core components of RNAP III (51–53). Comparisons of the RNAP III subunit hRCP39 revealed a well-conserved archaeal homolog (termed TFEβ) that directly and extensively interacts with TFE (now named TFEα) (20). Although TFEβ is not conserved in all *Archaea*, TFEβ is essential for some *Crenarchaea*. *In vitro*, TFEα-TFEβ complexes are effective in binding RNAP, stabilizing open complex formation, and stimulating total transcriptional output (20).

The mechanism of TFE recruitment to the initiation complex and its activities during initiation has been partially resolved. TFEα simultaneously binds TBP, RNAP, and downstream DNA and remains bound to RNAP during promoter escape and early elongation (Figure 1.1B; panel ii). TFE has been shown to stimulate transcription at noncanonical promoter sequences and at reduced temperatures *in vitro* (18, 19, 48, 54). Several studies have identified critical interactions between TFE and the preinitiation complex that have furthered our understanding of TFE function during initiation (2, 18, 54). TFEα consists of two domains: a winged helix (WH) domain and a zinc ribbon domain (55, 56); TFEβ contains a conserved WH domain and an FeS domain (20). The WH domain of TFEα contacts the upstream, nontemplate strand of DNA and helps form the open promoter complex through an unknown mechanism (18, 46). Several studies have shown that the presence of the RNAP stalk domain—unique to archaeoeukaryotic RNAPs and comprised of two subunits, RpoE and RpoF in archaea and Rpo4 and Rpo7 in eukaryotes—is essential for the full activity of TFEα (10, 54, 57). The predicted interaction between TFEα and the stalk domain was bolstered by copurification of TFEα with intact RNAP and the loss of TFEα from RNAP preparations wherein the stalk domain was missing (36). A structure-function study identified critical interactions between TFEα and RpoE of the stalk domain (35). TFE may have an essential role in modulating intramolecular movements of RNAP during the transcription cycle, most notably movements of the clamp domain. Interaction of TFEα with both the stalk and clamp domains of RNAP during transcription initiation may retain the clamp domain in an open complex conformation necessary for initiation and elongation.

Replacement of TFE by the elongation factors Spt4/5 during early elongation may alter clamp positioning and further stabilize the elongation complex (58).

1.4 Regulation of Transcription Elongation

As transcription transitions from initiation to elongation, RNAP undergoes a conformational change accompanied by the replacement of initiation factors with elongation factors (2, 58–61). It is plausible that the emerging nascent transcript stimulates the swap of regulatory factors and initiates the intramolecular movements that result in stable elongation complex formation (10, 62, 63). Very few transcription elongation factors have been bioinformatically identified within archaeal genomes, and it is probable that archaeon-specific factors await discovery. It is worth noting what is seemingly not encoded in archaeal genomes, given that so much of archaeal and eukaryotic transcription machinery is shared. Archaeal genomes do not appear to encode any coactivator complexes or megacomplexes for chromatin modification or rearrangements. There does not appear to be machinery for regulated posttranslational modifications of the archaeal transcription apparatus nor of chromatin, with the exception of acetylation/deacetylation of the small chromatin-associated protein Alba (64, 65). Furthermore, archaeal transcripts are not capped, do not require nuclear export, and, with the exception of self-splicing introns, lack introns; thus, factors responsible for these activities are similarly lacking from archaeal genomes (66–68).

Transcription elongation factors have various roles, including increasing processivity and fidelity of RNAP and/or increasing genome stability. Only three archaeal elongation factors have been experimentally studied: transcription factor S (TFS) (69, 70) and the aforementioned universally conserved elongation factor Spt5, often with a conserved binding partner Spt4 (Spt4/5) (Figure 1.1B; panel iii) (1, 71, 72). TFS, with homology to the C-terminal domain of eukaryotic TFIIS and functionally analogous to GreA/GreB in *Bacteria* (8, 73–75), can stimulate

endonucleolytic cleavage of the RNA from backtracked RNAP complexes (69, 70, 76). Several recent studies have shed light on the roles of Spt5 during elongation (58, 77, 78).

Archaeal Spt5, homologous to bacterially encoded NusG, consists of two domains: the NusG N-terminal (NGN) domain and a single C-terminal Kyrpides-Ouzounis-Woese (KOW) domain with affinity for single-stranded RNA (71, 72, 77); eukaryotic Spt5 typically contain three to six repeats of the C-terminal KOW domain (9, 79, 80). Critical, direct molecular interactions between Spt5 and RNAP have been identified in both *Bacteria* and *Archaea*, and the conservation of RNAP and Spt5 infers that these same interactions are used in *Eukarya*. Briefly, a hydrophobic depression on the NGN domain interacts with the mobile clamp domain of RNAP, with additional interactions between the NGN domain and RNAP jaw domain likely fixing the location of the clamp domain in a closed configuration (81). Spt5 interaction with RNAP is not necessary for productive and processive elongation *in vitro*, but the interaction does increase the total output of transcription systems (1). It is plausible that Spt5 increases elongation rates and processivity, as NusG in *Escherichia coli* does, and it is further possible that the increased efficiency of transcription results from the stabilization of the clamp domain that in turn stabilizes the DNA-RNA hybrid in place during transcription elongation (82–84). The NGN domain also contacts the upstream strands of DNA, offering protection from backtracking, and, by inference, may reduce pausing of the transcription elongation complex (78, 85, 86). It is of importance to note that NusG/Spt5 can have a positive and/or negative effect on elongation rates and pause events of RNAP. In *Thermus thermophilus*, NusG slows down RNA elongation rather than increases elongation rates (87). In *Bacillus subtilis*, sequence-specific interactions of the NGN and nontemplate DNA strand within the paused transcription bubble stabilize the pause event in the *trp* operon (88). Furthermore, evidence has shown that Spt4/5 induces pauses during early elongation of Pol I but promotes elongation downstream (89). Although NusG can elicit opposite roles on transcription elongation, the NusG-RNAP binding sites remain well conserved across various species. Archaeal and eukaryotic genomes often encode an additional elongation factor,

Spt4 (annotated as RpoE″/RpoE2 in *Archaea*). Spt4 forms a complex with Spt5 and stabilizes the Spt5-RNAP interaction (9, 72). Spt4 does not appear to be essential; however, the affinity of Spt5 for RNAP decreases in the absence of Spt4 *in vitro* (1).

The primary interacting partners (e.g., RNAP and Spt4) of the Spt5-NGN domain have been established in molecular detail; however, no specific interacting partners of the KOW domain have been identified in archaea. It is possible that the affinity of the KOW domain for RNA leads to nonspecific interactions with the emerging transcript; however, it is tempting to speculate greater involvement of the KOW domain based on the known activities of the C-terminus of bacterial NusG (Figure 1.2) (90). Bacterial NusG facilitates elongation or termination depending on its binding partner (82, 83, 91–94). Studies suggest, the bacterial NusG KOW domain interacts with the S10 ribosomal subunit (NusE) during elongation, thereby coupling the leading ribosome with the transcription apparatus (93, 94). In contrast, a recent study suggested the leading ribosome interacts directly with RNAP thus speculating that Spt5 is not essential for the coupling of transcription and translation (95). Spt5 may transiently interact with the S10 ribosomal subunit before the ribosome contacts RNAP. Nonetheless, when Spt5 is not bound to the leading ribosome or the leading ribosome lags substantially behind RNAP, the bacterial NusG-KOW domain can be bound by and stimulate the activity of the transcription termination factor Rho (92, 96, 97). Archaeal transcription and translation are similarly coupled (98, 99), and it is reasonable to venture that archaeal Spt5 may have a role in linking the archaeal transcription and translation apparatuses and also potentially interact with termination factor(s). The archaeal transcription termination factor Eta is not responsible for polarity and by extension is not responsible for global 3’ end transcript formation (16). Archaeal genomes likely encode a second termination factor that is responsible for polarity that remains to be identified.

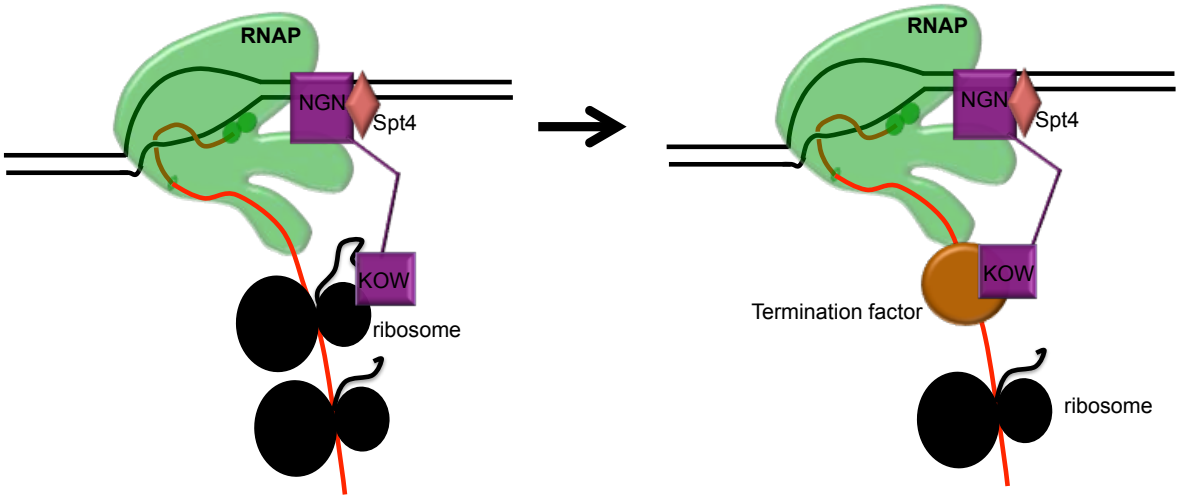


Figure 1.2. Coupling and uncoupling of transcription and translation. Transcription and translation are coupled in bacteria and archaea and it is hypothesized that Spt5 (purple) is the coupling factor that binds to both RNAP (green) and the leading ribosome (black). When transcription and translation become uncoupled a termination factor, termed Rho in bacteria, binds to the Spt5-KOW domain to mediate transcription termination.

1.5 Intramolecular Rearrangements of RNA Polymerase May Increase Processivity

The archaeal and three eukaryotic RNAPs can be reduced in complexity to three large domains: the core, the mobile clamp, and the stalk (Fig 1.1A) (60, 100). The archaeoeukaryotic stalk, absent from bacterial RNAP, is used by a host of archaeal and eukaryotic transcription factors to bind and regulate the activities of RNAP. Increasing evidence from biochemical, biophysical, and *in vivo* approaches indicate that transcription factor binding often stimulates intramolecular movements of RNAP that appear necessary for transitions between phases of the transcription cycle (2, 78, 101, 102).

Hinge-like movement of the mobile clamp domain has been demonstrated for the bacterial RNAP (59). The movements of the mobile clamp are sufficiently large enough to open the main channel of RNAP, such that double-stranded DNA can easily enter and exit when the clamp is open, whereas double-stranded DNA—or the RNA-DNA hybrid—would be trapped inside RNAP when the clamp is closed. The bacterial RNAP clamp is open during initiation but remains closed during processive elongation (59), leading to a simple model of encapsulation of the nucleic acids to explain the dramatic stability of the elongation complex. It is logical to propose mechanistic actions of transcription factors that may modulate the clamp positioning with respect to the core and stalk domains of RNAP and thus alter the stability and transitions of RNAP throughout the transcription cycle. TFE is predicted to make contacts with both the clamp and stalk domain of RNAP, thereby fixing the clamp into the open conformation critical for initiation (41, 54, 102, 103). In Chapter 2, we identify the critical residues in the stalk domain of RNAP responsible for the interaction between TFE and the stalk domain of RNAP during initiation (35). As transcription transitions into the elongation phase, RNA emerges from the enzyme and interacts with the stalk domain (63), where a predicted steric clash occurs between the RNA and the TFE, likely driving TFE to disengage from RNAP. The disengagement of TFE allows for Spt5 to bind to the clamp and core domains of RNAP and lock the clamp in the closed position, thus ensuring processivity during elongation (77).

RNAP clamp movement is predicted to be universal; however, both the archaeal and the eukaryotic RNAP contain additional subunits, including the stalk domain (100, 104), and previous structural data predicted that the stalk domain would sterically limit or abolish major movements of the clamp domain. Recent crystallographic evidence of the complete euryarchaeal RNAP demonstrated that the clamp is able to open without a steric clash with the stalk domain through a coordinated swing and rotation movement of both the clamp and stalk domains (60). This evidence supports the bacterial mechanism of the clamp opening and closing during initiation/termination or elongation, respectively, thus supporting a universal model of clamp movement.

1.6 Regulation of Transcription Termination

Transcription termination occurs when the transcription elongation complex becomes sufficiently unstable and fails to maintain contact between RNAP and the encapsulated nucleic acids. The stability of the transcription elongation complex is derived from (i) contacts between RNAP and the RNA-DNA hybrid, (ii) contacts between RNAP and single-stranded RNA in the exit channel, (iii) contacts between RNAP and the downstream DNA, and (iv) the base pairing of the RNA-DNA hybrid (105–112). The first and last of these contacts are most likely to be altered during the termination process. Transcription through specific DNA sequences can result in stronger or weaker base pairing within the RNA-DNA hybrid, and contacts between RNAP and the nucleic acids are most easily modified by movements of the clamp domain that relieve movements of the hybrid with respect to the core of RNAP (113–115). Release of the nascent RNA may be possible through continued translocation in the absence of RNA synthesis, or the RNA-DNA hybrid could be released in bulk if the clamp domain transitions from a closed to an open position. The gene-dense nature of many archaeal genomes necessitates timely termination of transcription to prevent aberrant transcription of neighboring genes. There are

two mechanisms of termination across all domains: intrinsic termination and factor-dependent termination (Figure 1.1B; panel iv, v, & vi and Figure 1.3).

Intrinsic transcription termination is driven primarily by weak base pairing within the RNA-DNA hybrid and occurs independent of the activity of transcription factors (Figure 1.3A) (105, 116). Intrinsic transcription termination has been established in all three domains, with some differences in sequence and structural requirements (22, 117–121). The archaeal RNAP, like eukaryotic RNAP III, is sensitive to intrinsic termination (23, 122, 123). Eukaryotic RNAP I and RNAP II do respond to DNA sequence context in the form of pauses and arrests but rarely release the transcript at such positions (124–126). Archaeal intrinsic termination is characterized by a run of 5 to 10 thymidine residues in the nontemplate strand, encoding a poly(U) run at the 3′ end of the nascent RNA (22, 23). The weak rU:dA RNA-DNA hybrid at or near the positions of termination is seemingly insufficiently energy rich to maintain the stability of the elongation complex; RNAP III similarly spontaneously dissociates upon transcription of poly(T) nontemplate tracts.

Transcription factors involved in initiation and elongation have been characterized in all domains, while until my thesis work, transcription termination factor(s) had only been characterized in *Bacteria* and *Eukarya* (Figure 1.3B) (16, 127–130). By inference, from known termination factors that are employed in bacterial and eukaryotic systems, it was easily argued that protein factors are encoded in archaeal genomes that have the capacity to direct transcription termination. Bioinformatic analyses revealed some potential targets that our lab has partially evaluated, but there are no easily identified homologues of known eukaryotic or bacterial termination factors. Two well-studied transcription bacterial termination factors, Rho and Mfd, lack clear homologues in archaeal genomes, but there are hints that analogous activities may be present in archaeal species (99, 131–137).

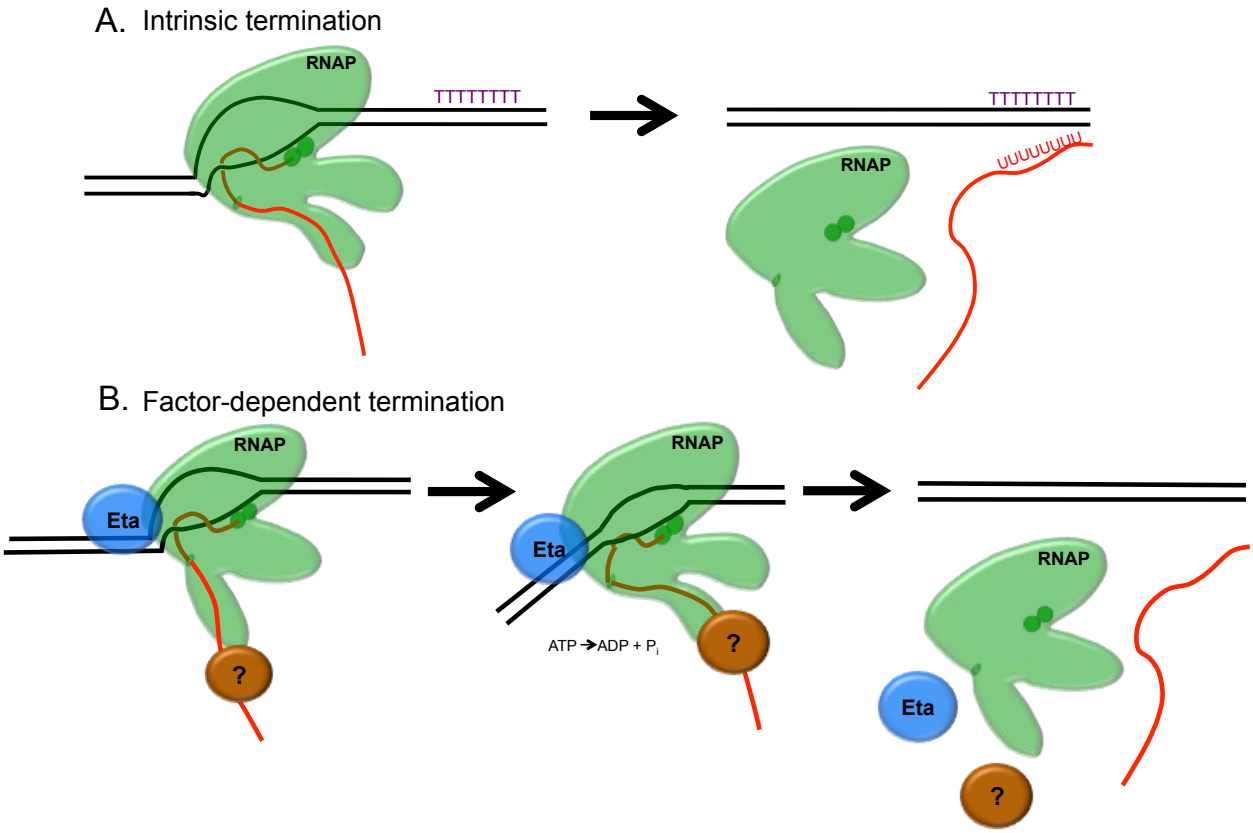


Figure 1.3. Archaeal intrinsic and factor-dependent termination. A. Intrinsic termination is characterized by a run of 5-10 thymidines in the non-template DNA strand. When RNAP transcribes through the intrinsic termination sequence the weakened RNA:DNA hybrid causes transcription termination. **B.** Eta (blue) is a novel, euryarchaeal transcription termination factor that binds upstream of the transcription elongation complex to promote termination. It is predicted that archaeal genomes encode a second termination factor (brown) that is responsible for polarity but this factor remains to be identified. In general, factor dependent termination is dependent on ATP-hydrolysis to apply force to the transcription elongation complex and mediate transcription termination.

The bacterial transcription termination factor Mfd removes RNAP from sites of DNA damage and initiates transcription-coupled DNA repair (133, 135, 137). Recently, transcription coupled DNA repair (TCR) was described in euryarchaea (131), and there is evidence that the archaeal RNAP halts synthesis and forms long-lived complexes at the site of lesions in DNA *in vitro* (132). This data suggests that mechanisms and presumably a transcription termination factor is encoded in euryarchaeal genomes to remove RNAP from the site of damage. The second bacterial termination factor Rho is a homohexamer helicase that represses phage transcription and mediates polar repression of downstream genes when transcription and translation become uncoupled (i.e., polarity) (24, 121, 138). *Archaea* demonstrate polar repression of downstream genes in the absence of continued translation, and it is likely that a factor or factors mediate polarity in archaea (99).

To investigate analogous activities to Rho or Mfd a biochemical grind and find was employed and a novel archaeal transcription termination factor, termed Eta, was characterized (16). Briefly, Eta binds to the upstream DNA of the TEC to release the nascent transcript and RNAP into solution. Eta recognizes paused/arrested RNAP and Eta-mediated termination is non-competitive with transcription elongation rates. Our studies provide strong evidence that Eta is analogous in function to the bacterial termination factor Mfd and is likely responsible for transcription-coupled DNA repair. Eta is not responsible for polar repression (i.e., polarity) and by extension is not responsible for global 3' end transcript formation (16). Archaeal genomes likely encode a second termination factor that is responsible for polarity that remains to be identified. In Chapter 3, we characterize Eta-mediated transcription termination, detail the rate of and requirements for Eta-mediated transcription termination and describe a role for Eta in transcription termination *in vivo* (16). In Chapter 4, we report the first crystal structure of Eta and employ Eta structure-function analyses. We continue to probe the mechanism of Eta as well as further investigate the role of Eta in transcription coupled DNA repair.

In the future, it is critical to identify and characterize the termination factor(s) responsible for polarity. It is tempting to use the bacterial model of NusG-Rho interactions to conjure a similar picture for Spt5 (homologous to NusG) interactions with an archaeal transcription termination factor (96, 134, 138, 139). In chapter 5, we identify several potential candidates that we predict are responsible for archaeal polarity. Future studies will continue to provide insight into archaeal polarity and mechanistic insights into transcription termination.

1.7 Chromatin architecture affects the transcription cycle

Archaea employ two seemingly distinct mechanisms to compact, wrap, and condense their genomes to fit within the cell (140). Most euryarchaeal species are polyploid and encode histone proteins that dominate chromatin architecture; archaeal histones mimic the core eukaryotic histone-fold (141–145). In contrast, most crenarchaeal species are diploid and are reliant on small, basic nucleoid proteins to organize their genomes (140, 146). Condensation demands organization of the genome and offers regulatory opportunities by controlling the accessibility of promoter sequences, the introduction of local superhelicities that may promote or inhibit promoter opening, and the potential for the introduction of chromatin-based obstacles to transcription elongation. The overall role of genome architecture with respect to archaeal transcription is an emerging area with several recent studies highlighting the breadth of influences genome architecture can have on transcription output at the organismal level.

Archaeal histone-based chromatin is composed of nucleosome particles that wrap and condense the genome. The best-described complexes are homo- or hetero-histone tetramers, homologous to the H3/H4 tetramer in eukaryotes, that associate with ~60 bp of double-stranded DNA. Archaeal histones share similar biases with eukaryotic nucleosomes for flexible DNA sequences and are, in general, absent from the core promoters of archaeal genes (147, 148). Archaeal histone proteins share the same core-fold as eukaryotic histones, but most lack the extensions from this fold (i.e. tails) that are highly modified and essential for proper nucleosome

dynamics in eukaryotes (149). Higher order structure has been demonstrated in *T. kodakarensis* in the form of dynamic histone polymers that have the ability to wrap up to 180 bp (150). Recently, the archaeal histone based chromatin crystal structure from *Methanothermus fervidus* was solved (151). The structure showed DNA wraps around an extended polymer, formed by archaeal histone homodimers, in a quasi-continuous superhelix with the same geometry as DNA in the eukaryotic nucleosome.

Archaeal nucleosomes present a surmountable barrier to the progression of the transcription elongation complex, although traversion does slow the elongation complex (152). The lack of known modifications to archaeal histones, and the lack of known machinery for the repositioning or movement of archaeal nucleosomes suggests that transcription elongation complexes simply traverse the nucleosomes and that chromatin organization spontaneously reforms when the histones gain access to preferred binding positions following RNAPs departure. This mechanism of elongation through the histones is similar to the mechanism of Pol III in eukaryotes (153–155).

The activities or stimulatory effects of archaeal elongation factors on transcription through archaeal histone-based chromatin remain to be explored; the substantial pausing and delayed progress of RNAP on chromatinized-templates suggests that elongation factors will accelerate progress of the transcription elongation complex. Any role of chromatin architecture in transcription termination is similarly unexplored. Topology of naked DNA templates does influence the positions and efficiencies of intrinsic terminators, suggesting that chromatin-templates may also influence termination patterns. Nucleosomes are not only depleted from promoter regions, but also from predicted termination regions, suggesting a potential regulatory role for chromatin architecture on termination of transcription (147, 148).

1.8 Thesis Rationale

Exploration of archaeal transcription and regulation continues to yield a bounty of evolutionary, biophysical, and mechanistic details of transcription mechanisms that are often applicable to all extant life. The ability to reconstitute the complete archaeal transcription apparatus permits biophysical studies not possible with eukaryotic components and the simplicity and explicit homology of many factors provides meaningful insight into the mechanistic roles of individual factors and even specific domains and residues of archaeal transcription components. The development and recent advances in genetic techniques for more archaeal species is now offering complementary *in vivo* studies to probe regulatory strategies and rationally manipulate protein interfaces and activities in the cell. My thesis work took advantage of both *in vitro* and *in vivo* studies to detail the TFE-RNAP molecular interactions (see Chapter 2) (35), to mechanistically characterize a novel archaeal transcription termination factor (see Chapter 3 and Chapter 4) (16), and to identify additional archaeal transcription termination factors (see Chapter 5).

REFERENCES

1. Hirtreiter A, et al. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* 38(12):4040–51.
2. Grohmann D, et al. (2011) The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* 43(2):263–274.
3. Armache K-J, Mitterweger S, Meinhart A, Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280(8):7131–4.
4. Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451(7180):851–4.
5. Rowlands T, Baumann P, Jackson SP (1994) The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* 264(5163):1326–9.
6. Qureshi SA, Bell SD, Jackson SP (1997) Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J* 16(10):2927–36.
7. Hausner W, Thomm M (2001) Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. *J Bacteriol* 183(10):3025–31.
8. Langer D, Hain J, Thuriaux P, Zillig W (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92(13):5768–72.
9. Guo M, et al. (2008) Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure* 16(11):1649–58.
10. Blombach F, Smollett KL, Grohmann D, Werner F (2016) Molecular Mechanisms of Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol* 428(12):2592–606.
11. Santangelo TJ, Cubonova L, James CL, Reeve JN (2007) TFB1 or TFB2 is sufficient for Thermococcus kodakaraensis viability and for basal transcription in vitro. *J Mol Biol*

367(2):344–357.

12. Decker KB, Hinton DM (2013) Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* 67:113–39.
13. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I (2009) Functional specialization of transcription elongation factors. *EMBO J* 28(2):112–22.
14. Fouqueau T, Zeller ME, Cheung AC, Cramer P, Thomm M (2013) The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res* 41(14):7048–7059.
15. D'Heygère F, Schwartz A, Coste F, Castaing B, Boudvillain M (2015) ATP-dependent motor activity of the transcription termination factor Rho from Mycobacterium tuberculosis. *Nucleic Acids Res* 43(12):6099–111.
16. Walker JE, Luyties O, Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci*:201704028.
17. Bartlett MS, Thomm M, Geiduschek EP (2000) The orientation of DNA in an archaeal transcription initiation complex. *Nat Struct Biol* 7(9):782–5.
18. Grunberg S, Bartlett MS, Naji S, Thomm M (2007) Transcription factor E is a part of transcription elongation complexes. *J Biol Chem* 282(49):35482–35490.
19. Bell SD, Brinkman AB, van der Oost J, Jackson SP (2001) The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep* 2(2):133–8.
20. Blombach F, et al. (2015) Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife* 4. doi:10.7554/eLife.08378.
21. Smollett K, Blombach F, Reichelt R, Thomm M, Werner F (2017) A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat Microbiol* 2:17021.
22. Santangelo TJ, Cubonová L, Skinner KM, Reeve JN (2009) Archaeal intrinsic

transcription termination in vivo. *J Bacteriol* 191(22):7102–8.

23. Santangelo TJ, Reeve JN (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol* 355(2):196–210.
24. Ciampi MS (2006) Rho-dependent terminators and transcription termination. *Microbiology* 152(Pt 9):2515–28.
25. Micorescu M, et al. (2007) Archaeal Transcription: Function of an Alternative Transcription Factor B from Pyrococcus furiosus. *J Bacteriol* 190(1):157–167.
26. Reichlen MJ, Murakami KS, Ferry JG (2010) Functional analysis of the three TATA binding protein homologs in Methanosarcina acetivorans. *J Bacteriol* 192(6):1511–7.
27. Lee HS, et al. (2008) The Complete Genome Sequence of Thermococcus onnurineus NA1 Reveals a Mixed Heterotrophic and Carboxydotrophic Metabolism. *J Bacteriol* 190(22):7491–7499.
28. Ng W V., et al. (2000) Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci* 97(22):12176–12181.
29. Coker JA, DasSarma S (2007) Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet* 8(1):61.
30. Hidese R, et al. (2014) Different roles of two transcription factor B proteins in the hyperthermophilic archaeon Thermococcus kodakarensis. *Extremophiles* 18(3):573–588.
31. Bonneau R, et al. (2007) A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell* 131(7):1354–1365.
32. Baliga NS, et al. (2000) Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol* 36(5):1184–5.
33. Paytubi S, White MF (2009) The crenarchaeal DNA damage-inducible transcription factor B paralogue TFB3 is a general activator of transcription. *Mol Microbiol* 72(6):1487–1499.
34. Österberg S, Peso-Santos T del, Shingler V (2011) Regulation of Alternative Sigma

Factor Use. *Annu Rev Microbiol* 65(1):37–55.

35. Walker JE, Santangelo TJ (2015) Analyses of in vivo interactions between transcription factors and the archaeal RNA polymerase. *Methods*. doi:10.1016/j.ymeth.2015.05.023.
36. Hirata A, et al. (2008) Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a Thermococcus kodakarensis mutant lacking subunit F is temperature-sensitive. *Mol Microbiol* 70(3):623–633.
37. DeDecker BS, et al. (1996) The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J Mol Biol* 264(5):1072–84.
38. Littlefield O, Korkhin Y, Sigler PB (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* 96(24):13668–73.
39. Kassavetis GA, et al. (1992) The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell* 71(6):1055–64.
40. Knutson BA, Hahn S (2011) Yeast Rrn7 and Human TAF1B Are TFIIB-Related RNA Polymerase I General Transcription Factors. *Science (80-)* 333(6049):1637–1640.
41. Grohmann D, Klose D, Fielden D, Werner F (2011) FRET (fluorescence resonance energy transfer) sheds light on transcription. *Biochem Soc Trans* 39(1):122–127.
42. Gietl A, et al. (2014) Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res* 42(10):6219–6231.
43. Adachi N, Senda M, Natsume R, Senda T, Horikoshi M (2008) Crystal structure of *Methanococcus jannaschii* TATA box-binding protein. *Genes to Cells*. doi:10.1111/j.1365-2443.2008.01233.x.
44. Qureshi SA, Jackson SP (1998) Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell* 1(3):389–400.
45. Bleiholder A, Frommherz R, Teufel K, Pfeifer F (2012) Expression of multiple tfb genes in different Halobacterium salinarum strains and interaction of TFB with transcriptional

activator GvpE. *Arch Microbiol* 194(4):269–79.

46. Nagy J, et al. (2015) Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat Commun* 6:6161.
47. Werner F, Weinzierl RO (2005) Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol* 25(18):8344–8355.
48. Hanzelka BL, Darcy TJ, Reeve JN (2001) TFE, an archaeal transcription factor in Methanobacterium thermoautotrophicum related to eucaryal transcription factor TFIIEalpha. *J Bacteriol* 183(5):1813–8.
49. Ohkuma Y, Sumimoto H, Horikoshi M, Roeder RG (1990) Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. *Proc Natl Acad Sci* 87(23):9163–9167.
50. Gregory Peterson M, et al. (1991) Structure and functional properties of human general transcription factorIIE. *Nature* 354(6352):369–373.
51. Vannini A, Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45(4):439–46.
52. Lefèvre S, et al. (2011) Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nat Struct Mol Biol* 18(3):352–358.
53. Carter R, Drouin G (2010) The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol* 27(5):1035–43.
54. Naji S, Grünberg S, Thomm M (2007) The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem* 282(15):11047–57.
55. Meinhart A, Blobel J, Cramer P (2003) An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem* 278(48):48267–74.
56. Okuda M, et al. (2004) A novel zinc finger structure in the large subunit of human general

transcription factor TFIIE. *J Biol Chem* 279(49):51395–403.

57. Ouhammouch M, Werner F, Weinzierl ROJ, Geiduschek EP (2004) A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* 279(50):51719–21.

58. Blombach F, et al. (2013) Archaeology of RNA polymerase: factor swapping during the transcription cycle. *Biochem Soc Trans* 41(1):362–7.

59. Chakraborty A, et al. (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science* 337(6094):591–5.

60. Jun S-H, et al. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat Commun* 5:5132.

61. Hartzog G a, Kaplan CD (2011) Competing for the clamp: promoting RNA polymerase processivity and managing the transition from initiation to elongation. *Mol Cell* 43(2):161–3.

62. Hirtreiter A, Grohmann D, Werner F (2010) Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Res* 38(2):585–596.

63. Grohmann D, et al. (2010) RNA-binding to archaeal RNA polymerase subunits F/E: a DEER and FRET study. *J Am Chem Soc* 132(17):5954–5955.

64. Bell SD, Botting CH, Wardleworth BN, Jackson SP, White MF (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296(5565):148–51.

65. Ma C, et al. (2015) Alba from Thermoplasma volcanium belongs to α-NAT's: An insight into the structural aspects of Tv Alba and its acetylation by Tv Ard1. *Arch Biochem Biophys*. doi:10.1016/j.abb.2015.11.039.

66. Tang TH, et al. (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* 30(4):921–30.

67. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) Coevolution of tRNA

intron motifs and tRNA endonuclease architecture in Archaea. *Proc Natl Acad Sci U S A* 102(43):15418–22.

68. Kjems J, Jensen J, Olesen T, Garrett RA (1989) Comparison of transfer RNA and ribosomal RNA intron splicing in the extreme thermophile and archaebacterium Desulfurococcus mobilis. *Can J Microbiol* 35(1):210–4.

69. Hausner W, Lange U, Musfeldt M (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* 275(17):12393–9.

70. Lange U, Hausner W (2004) Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* 52(4):1133–43.

71. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 30(7):1302–10.

72. Zhou H, Liu Q, Gao Y, Teng M, Niu L (2009) Crystal structure of NusG N-terminal (NGN) domain from Methanocaldococcus jannaschii and its interaction with rpoE''. *Proteins* 76(4):787–93.

73. Marr MT, Roberts JW (2000) Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol Cell* 6(6):1275–85.

74. Erie DA, Hajiseyedjavadi O, Young MC, von Hippel PH (1993) Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* 262(5135):867–73.

75. Grünberg S, Reich C, Zeller ME, Bartlett MS, Thomm M (2010) Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes. *Nucleic Acids Res* 38(6):1950–63.

76. Schweikhard V, et al. (2014) Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc Natl*

Acad Sci U S A 111(18):6642–7.

77. Klein BJ, et al. (2011) RNA polymerase and transcription elongation factor Spt4 / 5 complex structure. doi:10.1073/pnas.1013828108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1013828108.

78. Guo G, et al. (2015) Structural and biochemical insights into the DNA-binding mode of MjSpt4p:Spt5 complex at the exit tunnel of RNAPII. *J Struct Biol*. doi:10.1016/j.jsb.2015.09.023.

79. Meyer PA, et al. (2015) Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Mol Cell Biol* 35(19):3354–3369.

80. Yakhnin A V, Babitzke P (2014) NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? *Curr Opin Microbiol* 18:68–71.

81. Sevostyanova A, Belogurov G a, Mooney R a, Landick R, Artsimovitch I (2011) The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol Cell* 43(2):253–62.

82. Burova E, Hung SC, Sagitov V, Stitt BL, Gottesman ME (1995) Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J Bacteriol* 177(5):1388–92.

83. Herbert KM, et al. (2010) E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J Mol Biol* 399(1):17–30.

84. Belogurov GA, Artsimovitch I (2015) Regulation of Transcript Elongation. *Annu Rev Microbiol* 69:49–69.

85. Yakhnin A V, Murakami KS, Babitzke P (2016) NusG is a Sequence-specific RNA Polymerase Pause Factor that Binds to the Non-template DNA Within the Paused Transcription Bubble. *J Biol Chem*. doi:10.1074/jbc.M115.704189.

86. Crickard JB, Fu J, Reese JC (2016) Biochemical Analysis of Yeast Suppressor of Ty 4/5

(Spt4/5) Reveals the importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. *J Biol Chem*. doi:10.1074/jbc.M116.716001.

87. Sevostyanova A, Artsimovitch I (2010) Functional analysis of Thermus thermophilus transcription factor NusG. *Nucleic Acids Res* 38(21):7432–45.

88. Yakhnin A V, Yakhnin H, Babitzke P (2008) Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. *Proc Natl Acad Sci U S A* 105(42):16131–6.

89. Anderson SJ, et al. (2011) The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. *J Biol Chem* 286(21):18816–24.

90. Tomar SK, Artsimovitch I (2013) NusG-Spt5 proteins-Universal tools for transcription modification and communication. *Chem Rev* 113(11):8604–19.

91. Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R (2009) Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol* 391(2):341–58.

92. Chalissery J, et al. (2011) Interaction Surface of the Transcription Terminator Rho Required to Form a Complex with the C-Terminal Domain of the Antiterminator NusG. *J Mol Biol* 405(1):49–64.

93. Gottesman ME, Rösch P (2010) A NusE : NusG Complex Links Transcription and Translation. 283(APRIL):501–505.

94. Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328(5977):504–8.

95. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P (2017) Architecture of a transcribing-translating expressome. *Science (80-)* 356(6334):194–197.

96. Peters JM, et al. (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev* 26(23):2621–33.

97. Burns CM, Richardson JP (1995) NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. *Proc Natl Acad Sci U S A* 92(11):4738–42.

98. French SL, Santangelo TJ, Beyer AL, Reeve JN (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol* 24(4):893–895.

99. Santangelo TJ, et al. (2008) Polarity in archaeal operon transcription in Thermococcus kodakaraensis. *J Bacteriol* 190(6):2244–8.

100. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* 292(5523):1876–82.

101. Drögemüller J, et al. (2015) Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci Rep* 5:16428.

102. Schulz S, et al. (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci* 113(13):E1816–E1825.

103. Tanaka A, et al. (2015) Association of the winged helix motif of the TFIIEα subunit of TFIIE with either the TFIIEβ subunit or TFIIB distinguishes its functions in transcription. *Genes Cells* 20(3):203–16.

104. Hirata A, Murakami KS (2009) Archaeal RNA polymerase. *Curr Opin Struct Biol* 19(6):724–31.

105. Gusarov I, Nudler E (1999) The mechanism of intrinsic transcription termination. *Mol Cell* 3(4):495–504.

106. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448(7150):157–62.

107. Kettenberger H, Armache K-J, Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* 16(6):955–65.

108. Kireeva ML, Komissarova N, Waugh DS, Kashlev M (2000) The 8-nucleotide-long

RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* 275(9):6530–6.

109. Andrecka J, et al. (2009) Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res* 37(17):5803–9.

110. Reeder TC, Hawley DK (1996) Promoter Proximal Sequences Modulate RNA Polymerase II Elongation by a Novel Mechanism. *Cell* 87(4):767–777.

111. Nudler E, Avetissova E, Markovtsov V, Goldfarb A (1996) Transcription Processivity: Protein-DNA Interactions Holding Together the Elongation Complex. *Science (80-)* 273(5272):211–217.

112. Sidorenkov I, Komissarova N, Kashlev M (1998) Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell* 2(1):55–64.

113. Komissarova N, Becker J, Solter S, Kireeva M, Kashlev M (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell* 10(5):1151–62.

114. Hein PP, et al. (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* 21(9):794–802.

115. Bochkareva A, Yuzenkova Y, Tadigotla VR, Zenkin N (2012) Factor-independent transcription pausing caused by recognition of the RNA-DNA hybrid sequence. *EMBO J* 31(3):630–9.

116. Martin FH, Tinoco I (1980) DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res* 8(10):2295–2300.

117. Czyz A, Mooney RA, Iaconi A, Landick R (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* 5(2):e00931.

118. Arimbasseri AG, Rijal K, Maraia RJ Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta* 1829(3–4):318–30.

119. Reiter WD, Palm P, Zillig W (1988) Transcription termination in the archaebacterium Sulfolobus: signal structures and linkage to transcription initiation. *Nucleic Acids Res* 16(6):2445–59.

120. Santangelo TJ, Roberts JW (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell* 14(1):117–26.

121. Reynolds R, Chamberlin MJ (1992) Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators. *J Mol Biol* 224(1):53–63.

122. Arimbasseri AG, Maraia RJ (2015) Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. *Mol Cell*. doi:10.1016/j.molcel.2015.04.002.

123. Dieci G, Sentenac A (1996) Facilitated recycling pathway for RNA polymerase III. *Cell* 84(2):245–52.

124. Dedrick RL, Kane CM, Chamberlin MJ (1987) Purified RNA polymerase II recognizes specific termination sites during transcription in vitro. *J Biol Chem* 262(19):9098–9108.

125. Palangat M, Hittinger CT, Landick R (2004) Downstream DNA selectively affects a paused conformation of human RNA polymerase II. *J Mol Biol* 341(2):429–42.

126. Kerppola TK, Kane CM (1988) Intrinsic sites of transcription termination and pausing in the c-myc gene. *Mol Cell Biol* 8(10):4389–4394.

127. ROBERTS JW (1969) Termination Factor for RNA Synthesis. *Nature* 224(5225):1168–1174.

128. Park J-S, Marr MT, Roberts JW (2002) E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109(6):757–67.

129. Fong N, et al. (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* 60(2):256–267.

130. Chen X, et al. (2017) Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the Saccharomyces cerevisiae genome. *PLOS Genet* 13(6):e1006863.

131. Stantial N, Dumpe J, Pietrosimone K, Baltazar F, Crowley DJ (2016) Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst)* 41:63–8.

132. Gehring AM, Santangelo TJ (2017) Archaeal RNA Polymerase Arrests Transcription at DNA Lesions. *Transcription* In press.

133. Howan K, et al. (2012) Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* 490(7420):431–4.

134. Leela JK, Syeda AH, Anupama K, Gowrishankar J (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proc Natl Acad Sci U S A* 110(1):258–63.

135. Smith AJ, Pernstich C, Savery NJ (2012) Multipartite control of the DNA translocase, Mfd. *Nucleic Acids Res* 40(20):10408–16.

136. Park J-S, Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc Natl Acad Sci U S A* 103(13):4870–5.

137. Haines NM, Kim Y-IT, Smith AJ, Savery NJ (2014) Stalled transcription complexes promote DNA repair at a distance. *Proc Natl Acad Sci U S A* 111(11):4037–42.

138. Cardinale CJ, et al. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320(5878):935–8.

139. Valabhoju V, Agrawal S, Sen R (2016) Molecular Basis of NusG-mediated Regulation of Rho-dependent Transcription Termination in Bacteria. *J Biol Chem* 291(43):22386–22403.

140. Peeters E, Driessen RPC, Werner F, Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol* 13(6):333–341.

141. Jaakkola ST, et al. (2014) Halophilic Archaea Cultivated from Surface Sterilized Middle-Late Eocene Rock Salt Are Polyploid. *PLoS One* 9(10):e110533.

142. Zerulla K, Soppa J (2014) Polyploidy in haloarchaea: advantages for growth and survival. *Front Microbiol* 5:274.

143. Spaans SK, van der Oost J, Kengen SWM (2015) The chromosome copy number of the hyperthermophilic archaeon Thermococcus kodakarensis KOD1. *Extremophiles*. doi:10.1007/s00792-015-0750-5.

144. Breuert S, Allers T, Spohn G, Soppa J (2006) Regulated polyploidy in halophilic archaea. *PLoS One* 1(1):e92.

145. Pereira SL, Grayling RA, Lurz R, Reeve JN (1997) Archaeal nucleosomes. *Proc Natl Acad Sci U S A* 94(23):12633–7.

146. Driessen RPC, Dame RT (2013) Structure and dynamics of the crenarchaeal nucleoid: Figure 1. *Biochem Soc Trans* 41(1):321–325.

147. Ammar R, et al. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* 1:e00078.

148. Nalabothula N, et al. (2013) Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* 14:391.

149. Forbes AJ, et al. (2004) Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proc Natl Acad Sci* 101(9):2678–2683.

150. Maruyama H, et al. (2013) An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Rep* 14(8):711–7.

151. Mattiroli F, et al. (2017) Structure of histone-based chromatin in Archaea. *Science (80-)* 357(6351):609–612.

152. Xie Y, Reeve JN (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol* 186(11):3492–8.

153. Xie Y, Reeve JN (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol* 186(11):3492–3498.

154. Studitsky VM (1997) Mechanism of Transcription Through the Nucleosome by Eukaryotic RNA Polymerase. *Science* 278(5345):1960–1963.

155. Clark DJ, Felsenfeld G (1992) A nucleosome core is transferred out of the path of a transcribing polymerase. *Cell* 71(1):11–22.

CHAPTER 2

ANALYSES OF IN VIVO INTERACTIONS BETWEEN TRANSCRIPTION FACTORS AND THE ARCHAEAL RNA POLYMERASE[2]

2.1 Introduction

Transcription is a highly regulated and orchestrated process in each Domain. Chemical and structural transitions of RNA polymerase are often directed by conserved and essential transcription factors (1–4). The transcription systems employed in Archaea mimic those employed in eukaryotes, and although component complexity is reduced, the archaeal RNA polymerase is highly conserved both in overall structure and regulatory capacity to eukaryotic RNA polymerase II (Pol II) (5–8). All currently available archaeal genomes encode a single but multi-subunit RNA polymerase (RNAP) (9) that is regulated by conserved transcription factors to ensure proper transcription initiation, elongation, and termination (2, 4, 10–14). Details of transcription mechanisms and regulation in the archaeal domain have lagged behind the knowledge of eukaryotic and bacterial systems, at least in part, due to the previous lack of genetic techniques. However, in recent years, several recombinant archaeal systems have been established, and genetic techniques for archaeal systems have evolved to permit rational manipulation of transcription systems *in vivo*. Tractable transcription systems from both of the dominantly studied and heavily populated archaeal phylums (the crenarcheota and euryarcheota) are available (15–25). The hyperthermophilic marine euryarchaeon *Thermococcus kodakarensis* has both a robust genetic system and a highly purified *in vitro*

[2] The work in this chapter was previously published under: Walker JE, Santangelo TJ (2015) Analyses of in vivo interactions between transcription factors and the archaeal RNA polymerase. *Methods*. Here I have written the contents of that work to focus on my contributions and findings important to our following studies.

Santangelo TJ, and I conceived the content, co-wrote the manuscript and contributed in experimental design, data collection, analysis, and interpretation.

transcription system (17, 22, 25, 26). The atomic resolution structure of 11-subunit RNAP from *T. kodakarensis* was recently solved at 3.5 Å (Figure 2.1) (5). This is the first structural analysis of the euryarchaeal RNAP and the first structure wherein the RNAP clamp domain is in an open configuration. The newly solved crystal structure provides a wealth of information that can be used to probe RNAP activities and structural-based hypotheses.

Three transcription initiation factors are seemingly essential in archaeal organisms: TATA-binding protein (TBP; homologous to the eukaryotic TBP), transcription factor B (TFB; homologous to eukaryotic transcription factor IIB), and TFE (homologous to eukaryotic TFIIEα) (6, 12, 27, 28). TFE is dispensable during *in vitro* transcription assays; however, all attempts to genetically remove TFE from the genome have, to date, been unsuccessful arguing for essential *in vivo* activities (22, 23, 29). TFE binds to the non-template strand of DNA and promotes DNA melting to assist formation of the open-promoter complex (30, 31). Previous studies have also revealed that TFE interacts directly with both RNAP and TBP (32).

The mechanistic activities of the conserved transcription factors are only partially known. Identification of binding sites and defined protein-protein contact surfaces is critical to further establish the mechanistic details of transcription factors. There are several clues to sites of important molecular interactions between TFE and RNAP. Biochemical assays suggest that TFE binds both the RpoE subunit (also termed Rpo7, although conventional nomenclature for archaeal RNAP subunits will be used hereafter) and the RpoA subunit coiled-coil domain of RNAP (Figure 2.1) (29, 33, 34). More specifically, the N-terminal winged-helix (WH) domain of TFE binds the RpoA subunit coiled-coil domain of RNAP while the zinc-ribbon (ZR) C-terminal domain of TFE is predicted to bind the stalk domain comprised of RpoE and RpoF (33, 35).

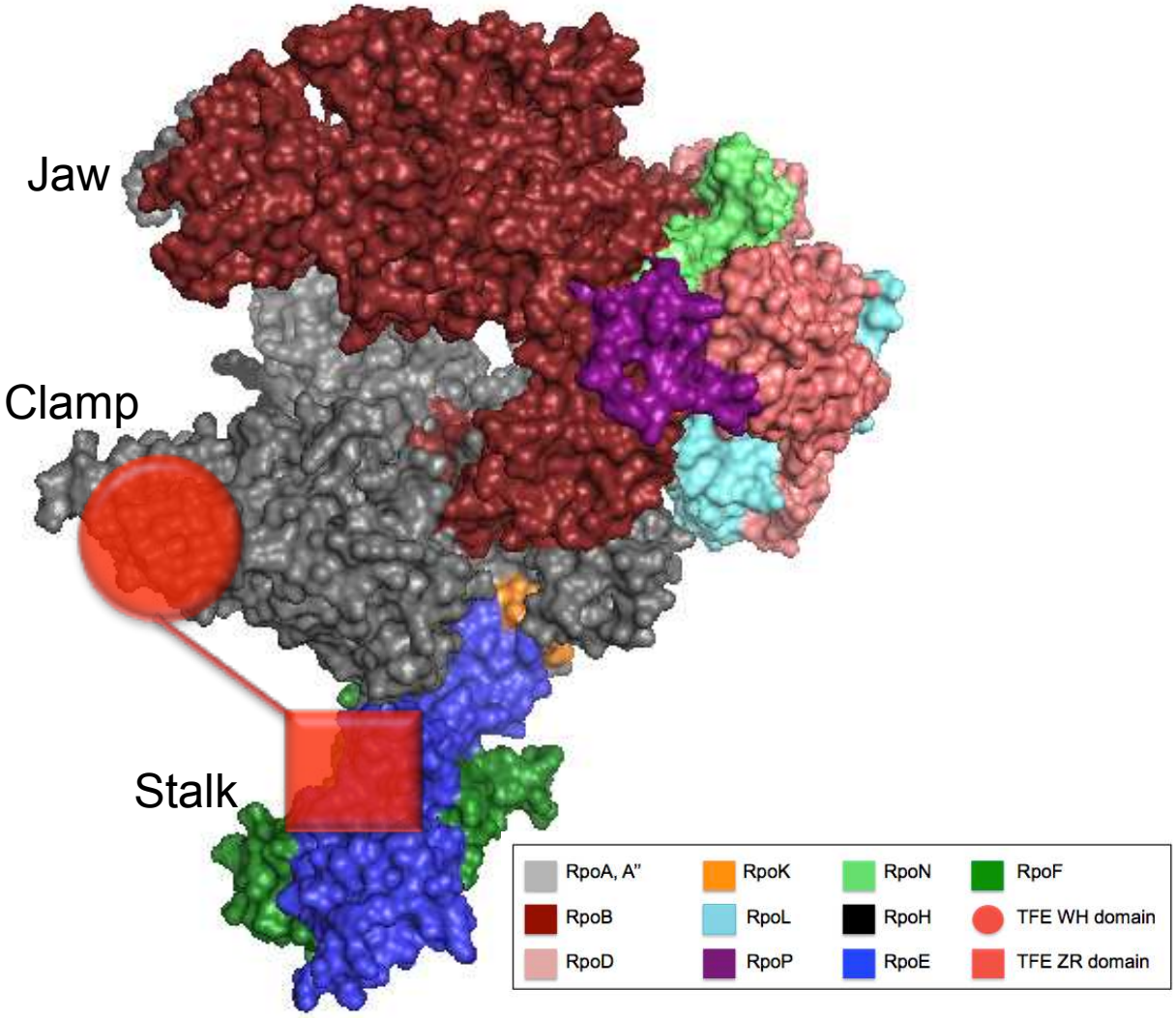


Figure 2.1. Atomic structure of the archaeal RNA polymerase. The *T. kodakarensis* RNA polymerase structure (PDB ID: 4QIW) in a surface representation. Note the clamp domain is in the open conformation and each RNAP subunit is shown in a different color. TFE is shown in rose bound to the clamp (grey) and stalk (blue and green) domains of RNAP.

While several studies have suggested that TFE binds directly to RPoE the molecular surfaces involved have not been identified. It was of interest to further investigate these molecular interactions, and this article provides details of a structure-function analysis of the role of RpoE in retention of TFE. The principles outlined within can easily be adapted to permit *in vivo* analysis of any predicted or known molecular interactions with the archaeal RNAP.

2.2 Results

Genetic techniques in *T. kodakarensis*.

Genetic techniques available to probe and explore the physiology, metabolism, and regulation of several archaeal lineages have exploded over the past ten years (17–21, 25, 26, 36–38). The natural competence and facile genetic system available for *T. kodakarensis* is consistently employed to probe the composition, activities, and interplay of the information processing systems (17, 26). Any gene in the *T. kodakarensis* genome can be markerlessly affinity tagged, deleted (only non-essential genes), and/or modified to encode mutational protein variants permitting detailed dissections of protein activities *in vivo*. Furthermore, autonomously replicating expression plasmids can be retained in *T. kodakarensis* allowing for ectopic expression of a WT or a mutational-variant protein (25, 39).

Genomic manipulations to RNAP subunits.

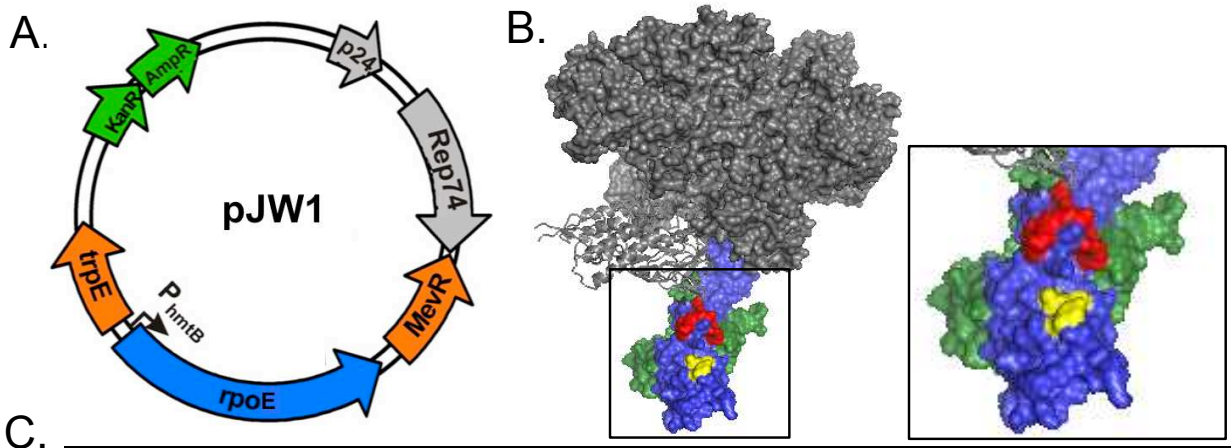
T. kodakarensis strains have been constructed wherein the gene encoding for a single subunit of RNAP (*rpoL*) was modified to encode a protein with a His_6 or His_6-HA affinity tag (6, 29). Through tagging a single subunit of RNAP, the entire, *in vivo* assembled RNAP complex can be purified from cells using a single column purification. Interestingly, when purifying *T. kodakarensis* RNAP TFE co-purifies in a sub-stoichiometric manner (29).

Co-purification is independent of the affinity-tag employed, and co-purification of TFE continues after several chromatography steps.

Plasmid expressed RNAP subunits.

T. kodakarensis does not contain endogenous plasmids. This limitation hindered progress in many arenas, and thus an expression vector was constructed by combining sequences from several origins to generate a shuttle vector that could autonomously replicate in both *T. kodakarensis* and *E. coli*. Briefly, the full sequence of a naturally occurring plasmid (pTN1) from a related Thermococcale species, *Thermococcus nautilus* (*Tn*), was fused with a common *E. coli* cloning vector (pCR2.1-Topo). The resultant fusion vector was further modified by inclusion of sequences encoding selectable markers that could be employed to force retention of the plasmid in select *T. kodakarensis* strains (Figure 2.2A) (25, 40).

The flexible movement of this shuttle vector between species permits researchers to clone and modify plasmid-encoded sequences in *E. coli*, then immediately transfer the desired plasmid to *T. kodakarensis*. The transformed plasmid often encodes a modified *T. kodakarensis* gene resulting in the establishment of a *T. kodakarensis* strain that is merodiploid; that is, the cells will contain both an endogenous genomic and ectopically-expressed copy of the gene of choice. Construction of merodiploid strains is a convenient methodology that allows for manipulation of essential genes in *T. kodakarensis* without the need to alter the genomic copy of the essential factor. Each of the 11-subunits of *T. kodakarensis* RNAP has been individually cloned into a pLC70-based vector under a constitutive promoter, P_{hmtB}, that drives high level expression of the subunit-encoding gene.



C.

Plasmid	Description	RpoE variant
pLC70	Empty vector	--
pJW1	Vector with WT *rpoE*	C-terminal his_6 tagged RpoE
pJW2	Vector with mutated *rpoE*	C-terminal his_6 tagged RpoE; S150-R157 deleted
pJW3	Vector with mutated *rpoE*	C-terminal his_6 tagged RpoE; Y95E, F98E, L107E

Figure 2.2. Plasmid expressed RNAP variants. A. A shuttle vector that can autonomously replicate in both *E. coli* and *T. kodakarensis* has been constructed with *rpoE* under control of a constitutive promoter, P_{hmtB}. **B.** *T. kodakarensis* RNAP as a surface model with the exception of the coiled-coil domain shown in a ribbon model. The core of RNAP is shown in gray, RpoF is shown in green, and RpoE is shown in blue. The hydrophobic patch of RpoE is highlighted in yellow (Y95, F98, L107). Residues S150-R157 of RpoE are highlighted in red. PDB ID: 4QIW. **C.** Description of the plasmids employed and the mutational variants encoded.

Plasmids encoding modified RpoE subunits.

Our immediate goal was to identify and mutationally alter the proposed surface(s) of RpoE that directly interact with TFE. We predicted that any TFE-RpoE interactions would be conserved across many species, and thus we modeled conserved and surface exposed residues of many RNAPs focusing on the solvent exposed surfaces of RpoE. This modeling revealed two immediate targets that were conserved and solvent exposed. A hydrophobic solvent exposed patch on RpoE (residues Y95, F98, and L107) was identified, and we predicted that this surface might represent an important point of contact between RNAP and TFE. A second conserved loop region was identified on RpoE and these residues (S150-R157) were similarly predicted to contribute to the TFE-RpoE interaction surface (Figure 2.2B). We rationalized that mutation of these surfaces may affect the stability of the TFE-RpoE interaction, and thus the overall affinity and co-purification of TFE with RNAP. We therefore generated a plasmid (pJW1) expressing full-length *rpoE* modified to encode a protein with a C-terminal His_6-tag (Figure 2.2C). Site-directed mutagenesis was employed to delete the nucleic acid sequences encoding for S150-R157 of RpoE, resulting in pJW2. A third vector, pJW3, was constructed wherein the sequences encoding for the solvent exposed hydrophobic patch of RpoE were altered to encode for glutamic acids (Y95E, F98E, and L107E).

Analysis of co-purification of TFE with the variant RpoE-RNAP variants.

TS517 strains carrying pJW1, pJW2, and pJW3 all yield obvious large quantities of RNAP holoenzyme from clarified cell lysates and the resultant RNAP complex remained intact through many chromatographic separations. Excess and non-RNAP incorporated RpoE was not visualized under any circumstances and this was anticipated given that RpoE is unlikely to stably fold and remain soluble or intact in the absence of its binding partner, RpoF. TFE was visible as a co-purification partner in RNAP preparations from TS517 carrying pJW1-3 and was

easily identified based on the migration of the protein during SDS-PAGE (Figure 2.3A). In contrast, TS517 carrying the empty vector pLC70 did not yield identifiable quantities of RNAP following any chromatographic purification; TFE could also not be identified in any bound fractions from such purifications. The relative stoichiometry of TFE and RNAP within each preparation was quantified following a combination of Ni^{2+}-affinity and MonoQ purification (see below; steps 1-11 of protein purification protocol). To analyze the amount of TFE co-purifying with RNAP, 7.5 µg of each protein preparation was resolved via SDS-PAGE and SYPRO Ruby stained according to the manufacturer's instructions. The stained gel was digitized using a Typhoon scanner and the relative staining of protein bands was quantified using ImageJ software. Specifically, the staining intensity of TFE was quantified within each preparation and normalized to the staining intensity of the easily identified largest subunit of *T. kodakarensis* RNAP, RpoB. The amount of TFE co-purifying with WT RpoE-RNAP in *T. kodakarensis* TS517 with pJW1 was set as a value of 100%.

A two-fold decrease in co-purification of TFE with RNAP was observed for TS517 pJW2 (ΔS150-R157) as well as TS517 pJW3 (Y95E, F98E, and L107E) when compared to TS517 pJW1 (WT) (Figure 2.3B). This decrease was repeatedly observed in multiple RNAP preparations from each strain, and the results obtained argue that both mutational variants of RpoE hinder stable interactions of TFE and RpoE in the context of the fully formed RNAP complex. Further purification of RNAP (steps 12-17 of protein purification protocol) was employed to determine if TFE would continue to co-purify with WT RNAP and the variant RNAPs as conditions became more selective. Size exclusion chromatography followed by heparin affinity chromatography separated the remaining non-specifically bound proteins from RNAP. To analyze the amount of TFE co-purifying with this more highly purified form of RNAP, 7.5 µg of each RNAP variant was resolved and quantified as before. A two-fold decrease was observed in co-purification of TFE in TS517 pJW3. An unanticipated loss of the stalk domain of

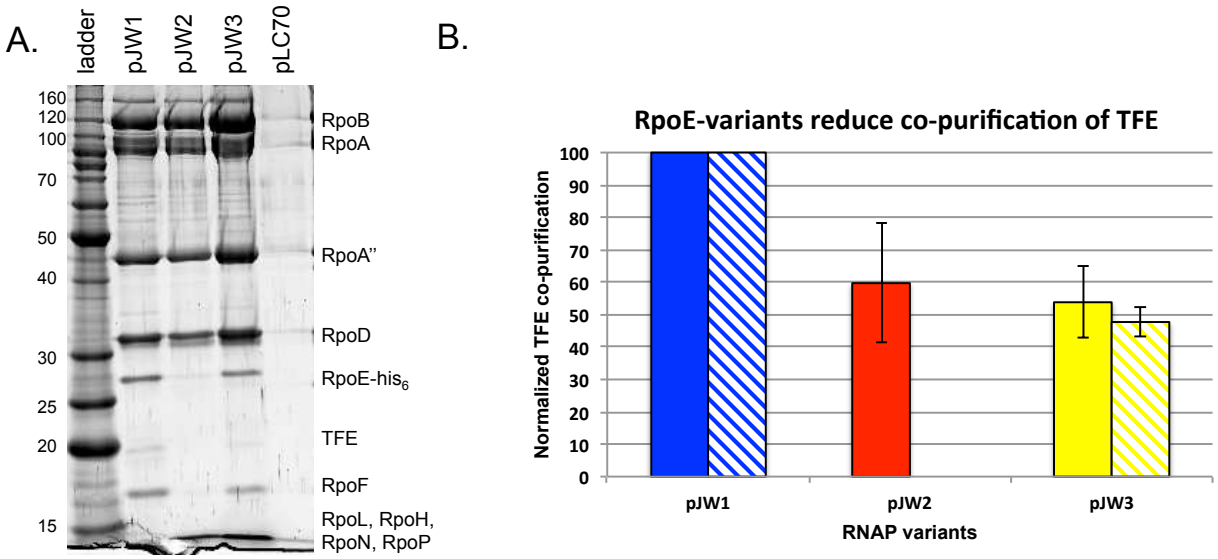


Figure 2.3. RpoE variants reduce co-purification of TFE with RNAP. A. SYPRO Ruby staining after SDS-PAGE of aliquots of RNAP (7.5 µg) purified from *T. kodakarensis* strain TS517 containing the vectors shown above the gel. Lane 1 contains size standards in kDa. Lanes 2-4 contain final purifications of RNAP from strains containing pJW1-3 respectively. An identical purification scheme from TS517 with the empty vector pLC70 is shown in lane 5 to demonstrate that RNAP purification is only possible based on the His_6-tag on RpoE. RNAP subunits are labeled to the right of the gel. **B.** Quantification of the normalized amount of TFE co-purifying with RNAP in the initial purification, denoted as a solid bar to the left, and final purification, denoted as a lined bar to the right. TFE was normalized to 100% in the *T. kodakarensis* strain encoding for the WT RNAP variant (pJW1), the amount of TFE co-purifying with RNAP in the mutant RNAP variants was normalized to WT RNAP. pJW2 denotes the RpoE variant wherein S150-R157 were deleted. pJW3 denotes the RpoE variant wherein the hydrophobic patch was disrupted.

RNAP (along with TFE) from the bulk of the RNAP purified from TS517 pJW2 following size exclusion and heparin affinity chromatography hindered accurate measurements of TFE-RNAP ratios (Figure 2.3B). A ten-fold decrease was observed in co- purification of TFE with RNAP; however, this loss reflects both a weakened co- purification of TFE with RNAP, as well as an unexpected decrease in the abundance of the stalk domain with respect to the remaining RNAP subunits. RpoE residues S150-R157 were not anticipated to directly interact with the core subunits of RNAP, but it is possible that misfolding or altered interactions between RpoE and RpoF in this variant could lower the affinity of the stalk domain in general for the core of RNAP. This result is supported by other studies wherein the RNAP stalk domain is required for association and/or activity of TFE (29, 33, 41).

2.3 Discussion

In this article, we describe a method to directly probe and quantify the binding of a conserved and essential transcription factor, TFE, to RNAP *in vivo*. *In vivo* structure-function analyses of transcription factor binding provide insight into mechanistic details that underlie regulation of RNAP. Structure-function analyses are often limited to *in vitro* assays; however, we describe a protocol that facilitates the investigation of transcription factor binding to RNAP *in vivo*. Furthermore, considerations specific to obtaining *T. kodakarensis* biomass, purifying RNAP, and final analysis of data were addressed.

2.4 Methods

Protocol for transformation and selection of plasmids into T. kodakarensis.

Transformation of *T. kodakarensis* differs significantly from more routine transformations of aerobic bacteria and yeasts. As such, care must be taken to ensure that anaerobic conditions are maintained at all times. We typically perform all manipulations within an anaerobic chamber maintained with an atmosphere of 95% nitrogen and 5%

hydrogen. The metabolism of *T. kodakarensis* also produces large quantities of pressured gases, typically H_2 and H_2S, and these explosive and toxic gases should always be properly and safely vented.

1. Anaerobically and sterilely inoculate 100 ml of anaerobic medium, typically an artificial sea water base supplemented with tryptone, yeast extract, and sulfur (ASW-YT-S) with 1 ml of a saturated overnight culture of TS517. Incubate the inoculated culture at 85°C for ~12 hours until cells have reached and remained in stationary phase for approximately 2 hours.

2. Anaerobically harvest the 100 ml culture and resuspend the cells in 3 ml of 0.8X ASW. Allow cells to rest, anaerobically, at 4˚C for ~30 minutes

3. Split the cell suspension into ~200 µl aliquots and add ~3 µg of plasmid DNA for each transformation. Chill the cells at 4˚C for ~60 minutes. Heat shock the cell- DNA mixture for 45 seconds at 85°C, then rapidly reduce the temperature of the transformation by immediately chilling the tube to 4˚C for minimally 5 minutes.

4. Carefully spread the resultant cell mixture onto gel-rite solidified solid media containing only a 19 amino acid mixture (-trp) as a nutrient source. The lack of tryptophan in the media inhibits the growth of all tryptophan auxotrophic cells. However, the reintroduction of a complete tryptophan biosynthetic pathway by means of plasmid-encoded proteins capable of catalyzing the terminal steps in tryptophan biosynthesis restores prototrophic cells. Place the dry, inverted plates in a sealed anaerobic vessel fitted with a GasPak (BD; 260678) at 85°C for ~4 days. The anaerobic vessel should not be artificially pressured with gases before incubation, and care should be taken to safely vent temperature- and microbiological-induced pressurized gases when opening the anaerobic vessel.

5. Identify initial colonies resultant from transformation. *T. kodakarensis* colonies are nearly colorless, rise only slightly above the plane of the surface of the medium, and typically reach a

maximum diameter of only ~2 mm, although colony morphology is often irregular. Proper identification of colonies is non-trivial given the propensity of defects in the solid-medium to mimic small colonies. *T. kodakarensis* colonies are often identifiable by halos of reduced opacity in the solid medium, and occasionally by dark centers. Presumptive colonies are lifted from the medium (typically with a sterile pipette tip), resuspended, and serial

diluted in 0.8X ASW. Several dilutions are necessary to yield single colonies in step 6 (below), as a typical *T. kodakarensis* colony contains ~10^6 cells.

6. Aliquots of serial diluted colonies are applied as small spots to gel-rite solidified media containing 13 µM mevinolin. This second round of selective pressure eliminates all cells not carrying the desired plasmid construct and ensures clonal populations of cells for subsequent studies. Colonies should be allowed to form (as above) on ASW-YT-S based medium containing 13 µM mevinolin for ~48 hours at 85˚C in a sealed anaerobic vessel fitted with a GasPak.

7. Discrete single colonies resultant after two days growth on gel-rite solidified ASW-YT-S + 13µM mevinolin media can be lifted and used to inoculate ~10 ml liquid cultures of ASW-YT-S + 13µM mevinolin medium. It is critical to ensure that the desired plasmid is retained in the clonal culture by physically purifying the plasmid to verify its presence as a plasmid. At a frequency of approximately just 1%, plasmid sequences can be randomly incorporated into the *T. kodakarensis* genome. After anaerobically inoculating and sealing the culture, incubate the culture at 85˚C for ~24 hours to permit growth to stationary phase.

T. kodakarensis plasmid prep protocol.

1. Anaerobically and sterilely remove ~5 ml of culture medium from the serum vial, and aerobically harvest the cell mass at room temperature via centrifugation. Discard supernatant and resuspend the cell pellet in 500 µl 100 mM Tris-HCl pH7.5, 50 mM EDTA, and 4 M NaCl.

2. Add 25 µL 20% SDS and mix well by vortexing to solubilize the cellular membranes.

4. Add 20 µL freshly prepared 5 M NaOH and mix well by vortexing to fully lyse the cells. Incubate ~3 minutes at room temperature (RT).

5. Add 400 µL 3 M potassium acetate pH 4.8, mix gently by inverting tube, and allow lysate to coagulate at RT for 1-2 minutes.

6. Centrifuge at 15,000 rpm for ~10 minutes at RT to clear the cellular debris from the solution.

7. Carefully remove the resultant clarified cellular lysate to new tube. Add one- tenth volume 3 M sodium-acetate pH 5.2, then seven-tenths volume 100% isopropanol to the solution. Mix thoroughly, then spin at ~15,000 rpm in a bench top centrifuge to pellet the nucleic acids. Discard the supernatant and air dry the resulting pellet. Resuspend the pellet of nucleic acids in ~20 µl 10 mM Tris-HCl pH 8.0.

8. Transform ~3 µl of recovered plasmid DNA into a 50 µL aliquot of XL1-blue competent cells. Plate on LB-AMP and incubate at 37°C overnight. Recovery of Amp^R-colonies from the transformation is used to verify plasmid recovery and maintenance in the original *T. kodakarensis* culture. Plasmid recovery from *E. coli* using established techniques followed by sequencing is used to verify retention of the desired mutational variants in the original *T. kodakarensis* cultures.

Preparation of gel-rite solidified media.

The optimal growth temperature of *T. kodakarensis* is 85˚C. These temperatures demand use of glass petri-plates and specialized agars that retain solid-form at elevated temperatures. The use of gel-rite, which activates and solidifies within ~30 seconds upon exposure to divalent cations, in an artificial sea water based medium necessitates an accurate and rapidly deployed mechanism to yield anaerobic, solid media for genetic manipulations. Preparation of solid-medium with only a pure amino-acid based nutrient source:

Combine 50 ml 2X ASW and 500 µl of 200X trace minerals in a 100 ml glass serum bottle and seal bottle (bottle #1) with a temperature resistant septum and crimped aluminum ring. Also combine 50 ml 18 mega-ohm purified water and 1.0 g gelrite in a 100 ml glass serum bottle and seal bottle (bottle #2) with a temperature resistant septum and crimped aluminum ring. Autoclave both bottles at 121˚C for 20 minutes and slowly exhaust the chamber using a liquid cycle. When internal autoclave temperature reaches ~95˚C, remove the contents and rapidly transfer the heated materials inside an anaerobic chamber. Carefully open bottle #1, add 5.0 ml 20X 19-AA mixture, 100 µl 1000X vitamin mixture, and 200 µL 500X polysulfides solution. Carefully open bottle #2, add the entirety of bottle #1 into bottle #2, gently swirl, and quickly dispense the contents into four glass plates (~25 ml/plate). The media will solidify fully in ~30 seconds following the mixture of bottles #1 and #2.

Preparation of solid-medium with a rich nutrient source: Combine 50 ml 2X ASW, 500µL 2000X trace minerals, 0.5 g yeast extract, and 0.5 g tryptone in a 100 ml glass serum bottle and seal bottle (bottle 1) with a temperature resistant septum and crimped aluminum ring. Also combine 50 ml 18 mega-ohm purified water and 1.0 g gelrite in a 100 ml glass serum bottle and seal bottle (bottle #2) with a temperature resistant septum and crimped aluminum ring. Autoclave both bottles at 121˚C for 20 minutes and slowly exhaust the chamber using a liquid cycle. When internal autoclave temperature reaches ~95˚C, remove the contents and rapidly transfer the heated materials inside an anaerobic chamber. Carefully open bottle #1, and add 100 µl 1000X vitamin mixture and 200 µL 500X polysulfides solution. When desired, mevinolin should be added here. Carefully open bottle #2, add the entirety of bottle #1 into bottle #2, gently swirl, and quickly dispense the contents into four glass plates (~25 ml/plate).

Biomass of T. kodakarensis for protein purification of RNAP variants.

T. kodakarensis is able to use elemental sulfur (S˚) and/or protons as a terminal electron acceptor, thereby generating H_2S or H_2 respectively during growth. A recent transcriptomics study revealed that the abundance of transcripts encoding *T. kodakarensis* TFE (TK2024) was most abundant when the cultures were supplied with both S˚ and grown glycolytically in the presence of pyruvate. Liquid cultures of TS517 carrying pLC70, pJW1, pJW2, and pJW3 were therefore grown in ASW-YT-S˚-pyruvate supplemented with 13 µM mevinolin and harvested at mid-exponential growth phase at an OD_{600} of ~0.4.

Protocol for growth of liquid *T. kodakarensis* cultures and collection of biomass.

1. Liquid medium is composed of 1X artificial sea water (ASW) supplemented with 0.5% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) pyruvate and 1X trace mineral solution. Media is prepared using boiled, 18 mega-ohm water to reduce overall O_2 content. 1.2 L of media is autoclaved (20 minutes, 121˚C, with a liquid exhaust cycle) in septum sealed 2 L pyrex vessels. Care should be taken not to exceed this ratio to ensure vessel integrity under high internal pressures. Always autoclave media in a secondary explosion proof container and employ necessary safety equipment.
2. Transfer vessels containing sterilized medium within an anaerobic chamber, sterilely add 2.0 g/L of finely pulverized elemental sulfur (flowers of sulfur). Mevinolin is added to 13 µM [final]. Add 1 mL/L of 1000X vitamin mixture (recipe available in *3.1.3.* below). Inoculate media with an appropriate amount of *T. kodakarensis* culture (1:100).
3. Incubate sealed vessel at 85˚C until an OD_{600} ~ 0.4 is achieved.
4. Harvest biomass (aerobically, 5000 x *g*) and discard supernatant.
5. Store pellets at -80˚C.

Protein purification and co-isolation of TFE with RNAP.

Purification of *in vivo* assembled RNAP containing ectopically expressed and affinity tagged RpoE, in WT and mutationally varied forms from pJW1, pJW2, and pJW3, should permit co-purification of TFE. If the introduced variants disrupt TFE interaction with RpoE in the context of RNAP holoenzyme, a discrepancy in the yields of TFE co-purifying with RNAP is anticipated. Ni^{2+}-affinity chromatography is employed for an initial purification of RNAP away from crude cell lysate. *T. kodakarensis* encodes several protein complexes with a natural affinity for the Ni2+-charged affinity matrix, only through more exhaustive chromatographic purifications is highly purified RNAP recovered.

Protocol for RNAP purification.

1. Resuspend biomass (3 mL per gram) in 25 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol and lyse by repeated freeze/thawing (-80˚C to 85˚C). Sonicate until all viscosity is lost signaling complete shearing of the nucleic acids.
2. Clarify lysate by centrifugation (15,000 x *g*). Discard pelleted debris.
3. Slowly add sufficient 30% 8K PEG in 2 M KCl to achieve a final concentration of 6% 8K PEG. Stir for ~1 hour at 4°C to precipitate nucleic acids. Clarify lysate by centrifugation (15,000 x *g*). Discard pelleted debris.
4. Load the clarified supernatant over a 5 mL Ni^{2+}-charged Hi-Trap chelating column (pre-equilibrated with 25 mM Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol) attached to a chromatography system. Discard flowthrough.
5. Wash the column with minimally 20 column volumes (CV) 25 mM Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol. Continue to flush the column with 25 mM Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol if substantial protein elution can be detected by monitoring the UV absorbance of the eluant. Elute the bound proteins, including RNAP, with a

~30 CV linear gradient from 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol to 25 ml Tris-HCl pH 8.0 containing 0.1M NaCl, 10% (v/v) glycerol, and 100 mM imidazole.

6. Collect 1-2 ml fractions during the elution. 10 µl of peak fractions (identified via UV absorbance) are combined with 2 µl 6X SDS-loading buffer and boiled for 3 minutes then separated using 10% discontinuous SDS-PAGE.

7. Fractions containing RNAP (molecular weight ~380,000 Kd; 11 subunits) should be identified by SDS-PAGE and coomassie staining.

8. Fractions containing RNAP should be pooled, diluted with 25 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM $MgCl_2$, 10% (v/v) glycerol until the conductivity of the sample is below the conductivity of 25 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM $MgCl_2$, 200mM KCl, 10% (v/v) glycerol. The diluted sample should be loaded onto and resolved through a 1 ml Mono Q column. Discard flowthrough. The column should be washed with minimally 20 CV 25 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM $MgCl_2$, 10% (v/v) glycerol containing 200 mM KCl, then bound proteins should be eluted with a linear gradient of 200 mM to 400 mM KCl in 25 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10% (v/v) glycerol.

9. Collect 1-2 ml fractions during the elution. 10 µl of peak fractions (identified via UV absorbance) are combined with 2 µl 6X SDS-loading buffer and boiled for 3 minutes.

11. Fractions containing RNAP (molecular weight ~380000 Kd; 11 subunits) should be identified by SDS-PAGE and coomassie staining.

12. Fractions containing RNAP should be pooled and concentrated to < 1 ml using 100 kDa molecular-weight cut-off centrifugal concentrators. Concentrated material should be loaded onto and resolved through a Superdex 200 16/60 column equilibrated with 20 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 100 mM NaCl at 0.2 ml/min.

13. Collect 1-2 ml fractions during the elution. 10 µl of peak fractions (identified via UV absorbance) are combined with 2 µl 6X SDS-loading buffer and boiled for 3 minutes.

14. Fractions containing RNAP (molecular weight ~380000 Kd; 11 subunits) should be identified by SDS-PAGE and coomassie staining. Pool appropriate fractions.

15. Fractions containing RNAP should be pooled and loaded onto a 1 ml heparin column equilibrated with 200 mM Tris-HCl pH 8.0, 100 mM NaCl. Bound proteins should be isocratically eluted with 200 mM Tris-HCl pH 8.0 and 1M NaCl.

16. Collect 1 mL fractions during the elution. 10 µl of peak fractions (identified via UV absorbance) are combined with 2 µl 6X SDS-loading buffer and boiled for 3 minutes. Fractions containing RNAP (molecular weight ~380,000 Kd; 11 subunits) should be identified by SDS-PAGE and coomassie staining. Fractions containing RNAP should be identified by SDS-PAGE and coomassie staining. These fractions should be pooled. Secure samples in 100kD dialysis tubing and dialyze samples (twice, for at least 6 hours) against minimally 1000-volumes protein storage buffer. Recover purified protein from the dialysis tube and store at -20˚C.

REFERENCES

1. Hausner W, Lange U, Musfeldt M (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* 275(17):12393–9.
2. Werner F (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J Mol Biol* 417(1–2):13–27.
3. Blombach F, et al. (2013) Archaeology of RNA polymerase: factor swapping during the transcription cycle. *Biochem Soc Trans* 41(1):362–7.
4. Walker JE, Luyties O, Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci*:201704028.
5. Jun S-H, et al. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat Commun* 5:5132.
6. Santangelo TJ, Cubonova L, James CL, Reeve JN (2007) TFB1 or TFB2 is sufficient for Thermococcus kodakaraensis viability and for basal transcription in vitro. *J Mol Biol* 367(2):344–357.
7. Werner F (2008) Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* 16(6):247–250.
8. Langer D, Hain J, Thuriaux P, Zillig W (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92(13):5768–72.
9. Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451(7180):851–4.
10. Blombach F, et al. (2015) Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife* 4. doi:10.7554/eLife.08378.
11. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 30(7):1302–10.

12. Gietl A, et al. (2014) Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res* 42(10):6219–6231.
13. Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9(2):85–98.
14. Brindefalk B, et al. (2013) Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res* 41(5):2832–2845.
15. Werner F, Weinzierl RO (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* 10(3):635–646.
16. Grohmann D, Hirtreiter A, Werner F (2009) RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro. *Biochem J* 421(3):339–343.
17. Hileman TH, Santangelo TJ (2012) Genetics Techniques for Thermococcus kodakarensis. *Front Microbiol* 3:195.
18. Wagner M, et al. (2012) Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. *Front Microbiol* 3:214.
19. Kohler PRA, Metcalf WW (2012) Genetic manipulation of Methanosarcina spp. *Front Microbiol* 3:259.
20. Santangelo TJ, Cubonová L, Reeve JN (2010) Thermococcus kodakarensis genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol* 76(4):1044–52.
21. Soppa J (2006) From genomes to function: haloarchaea as model organisms. *Microbiology* 152(Pt 3):585–90.
22. Gehring AM, Santangelo TJ (2015) Manipulating archaeal systems to permit analyses of transcription elongation-termination decisions in vitro. *Methods Mol Biol* 1276:263–79.
23. Smollett K, Blombach F, Werner F (2015) Transcription in Archaea: In Vitro Transcription Assays for mjRNAP. *Methods Mol Biol* 1276:305–14.

24. Sato T, Fukui T, Atomi H, Imanaka T (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *J Bacteriol* 185(1):210–20.

25. Santangelo TJ, Cubonová L, Reeve JN (2008) Shuttle vector expression in Thermococcus kodakaraensis: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl Environ Microbiol* 74(10):3099–104.

26. Farkas JA, Picking JW, Santangelo TJ (2013) Genetic techniques for the archaea. *Annu Rev Genet* 47:539–61.

27. Qureshi SA, Bell SD, Jackson SP (1997) Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J* 16(10):2927–36.

28. Bell SD, Magill CP, Jackson SP (2001) Basal and regulated transcription in Archaea. *Biochem Soc Trans* 29(4):392.

29. Hirata A, et al. (2008) Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a Thermococcus kodakarensis mutant lacking subunit F is temperature-sensitive. *Mol Microbiol* 70(3):623–633.

30. Schulz S, et al. (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci* 113(13):E1816–E1825.

31. Hanzelka BL, Darcy TJ, Reeve JN (2001) TFE, an archaeal transcription factor in Methanobacterium thermoautotrophicum related to eucaryal transcription factor TFIIEalpha. *J Bacteriol* 183(5):1813–8.

32. Bell SD, Brinkman AB, van der Oost J, Jackson SP (2001) The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep* 2(2):133–8.

33. Blombach F, Smollett KL, Grohmann D, Werner F (2016) Molecular Mechanisms of Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol* 428(12):2592–606.

34. Ouhammouch M, Werner F, Weinzierl ROJ, Geiduschek EP (2004) A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* 279(50):51719–21.

35. Grünberg S, Warfield L, Hahn S (2012) Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* 19(8):788–96.

36. Waege I, Schmid G, Thumann S, Thomm M, Hausner W (2010) Shuttle vector-based transformation system for Pyrococcus furiosus. *Appl Environ Microbiol* 76(10):3308–13.

37. Reich C, et al. (2009) The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable in vivo and in vitro. *Mol Microbiol* 71(4):989–1002.

38. Sato T, Fukui T, Atomi H, Imanaka T (2005) Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon Thermococcus kodakaraensis. *Appl Environ Microbiol* 71(7):3889–99.

39. Santangelo TJ, Reeve JN (2010) Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J Biol Chem* 285(31):23908–23915.

40. Soler N, et al. (2007) The rolling-circle plasmid pTN1 from the hyperthermophilic archaeon Thermococcus nautilus. *Mol Microbiol* 66(2):357–370.

41. Naji S, Grünberg S, Thomm M (2007) The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem* 282(15):11047–57.

CHAPTER 3

FACTOR-DEPENDENT ARCHAEAL TRANSCRIPTION TERMINATION[3]

3.1 Introduction

Each stage of transcription offers regulatory potential and increasing evidence supports that post-initiation transcription regulation may dominate in many instances (1–6). Although processive, transcription elongation is not uniform and regulated pausing through interactions with DNA- or nascent RNA-sequences, or through the action of conserved global and gene-specific regulators influences the elongation of RNA polymerase (RNAP) (6–20). Archaeal transcription is reliant on initiation factors with eukaryotic homology, but their access to promoter sequences is often limited or facilitated by bacterial-like transcription factors. In contrast, archaeal transcription elongation is seemingly regulated by universal or archaea-eukaryotic specific homologous factors (2, 21–33).

The ultimate control of transcription elongation is provided by factors and sequences that can disrupt the normally extremely-stable transcription elongation complex (TEC) to terminate transcription and release the nascent transcript and RNAP from the DNA template (3, 34–47). The archaeal RNAP is sensitive to intrinsic transcription termination. DNA sequences encoding poly-U rich sequences are sufficient to disrupt the archaeal TEC both *in vivo* and *in vitro*, and although surrounding sequence context can influence intrinsic transcription termination efficiency, there is no requirement for RNA structure for intrinsic termination (38, 41, 46, 48). Bioinformatic analyses of archaeal genomes reveals that many genes are organized into

---

[3] The work in this chapter was published under: Walker JE, Luyties O, Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci*.

Santangelo TJ and I conceived the content, co-wrote the manuscript, contributed in experimental design and data interpretation. Santangelo TJ, Luyties O, and I contributed in data collection and analysis.

operons and further that approximately one-half of genes and operons have sequences near their 3'-end that are consistent with intrinsic termination signals (46, 49–51). The absence of consensus intrinsic termination signals for many genes is concerning given the gene dense nature of many archaeal genomes. The genome of *Thermococcus kodakarensis* is >92% coding and the average intergenic space (after accounting for genes in operons) is only ~50 bp (52). In the absence of an intrinsic termination sequence, the stability of the TEC would be predicted to easily permit continued elongation from one gene to the next and thus remove the normal regulation imposed on expression of downstream sequences.

Intrinsic transcription termination has been demonstrated in all life, but protein factors that can disrupt the TEC have only been characterized for bacteria and eukarya. Insufficient intrinsic termination sequences for each gene and operon, combined with the observation that polar repression of gene expression occurs in the absence of coupled transcription and translation (i.e. polarity), as well as the recent description of transcription coupled DNA repair (TCR) in euryarchaea argued strongly that factor(s) were encoded in archaeal genomes that would be capable of disrupting TECs (52, 53). Bioinformatic analyses of archaeal genomes have identified some genes with homology to eukaryotic factors involved in RNA 3'-end formation (including cleavage and polyadenylation specificity factor subunits), but, to date, no biochemical activities have been described from archaeal cells that can disrupt the archaeal TEC. Importantly, these analyses have not identified any obvious homologues of the well-characterized bacterial termination factors Rho (54) or Mfd (55) although our studies predict that analogous activities are present in archaea cells.

We established a biochemical assay, using a robust *in vitro* transcription system dependent on purified RNAP and basal initiation factors from the model hyperthermophilic archaea *Thermococcus kodakarensis* to purify the first archaeal-encoded activity that can disrupt the TEC (38, 56). Our assay is dependent on the disruption of stalled archaeal TECs that are normally extremely stable and remain intact even when challenged with the strong

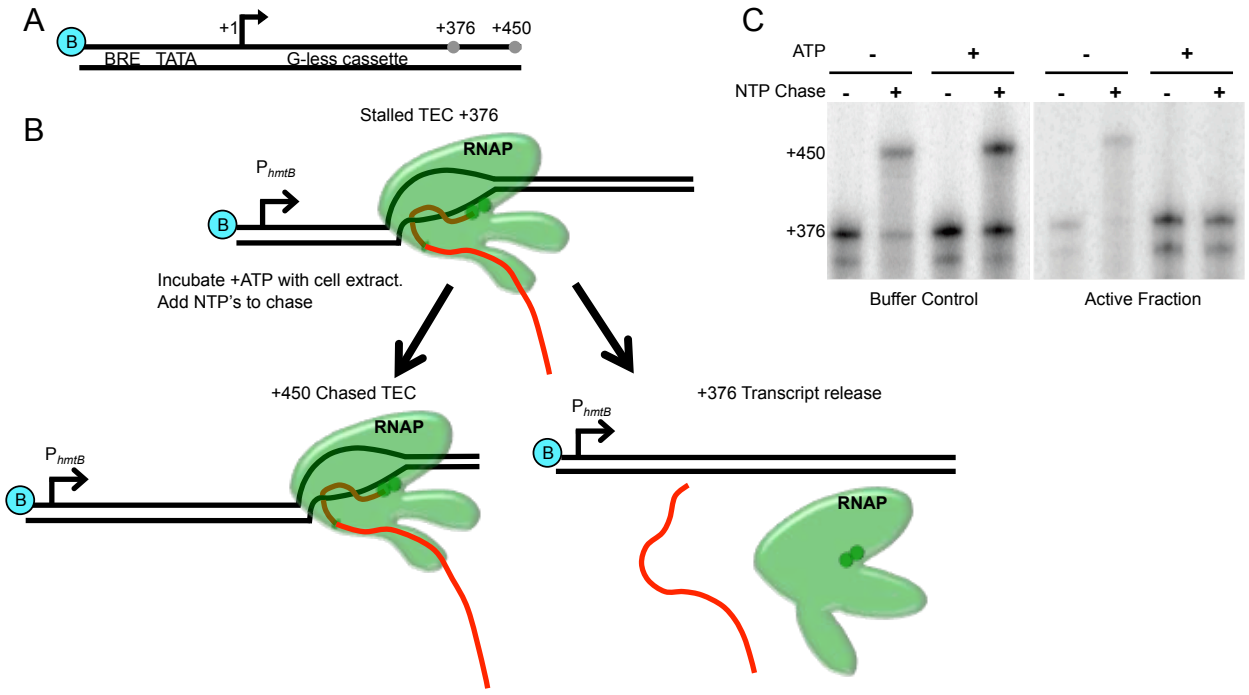
replicative minichromosome maintenance (MCM) helicase (57). The factor so purified, in native or recombinant form, requires ATP-hydrolysis to disrupt the stalled TEC and release both RNAP and the nascent transcript to solution. The encoding gene, TK0566, is universally conserved in known euryarchaeal genomes, and thus the termination factor was named Eta for Euryarchaeal Termination Activity (58, 59).

Eta is annotated as a DEAD-box RNA helicase (58, 60), but our analysis demonstrates that Eta does not require access to the nascent transcript but does require upstream DNA sequences to disrupt the archaeal TEC. Eta-mediated termination is not competitive with standard elongation rates, arguing that Eta targets stalled or arrested elongation complexes. Although conserved, deletion of TK0566 (encoding Eta) from the genome of *T. kodakarensis* was possible. Deletion of Eta does not influence polarity, suggestive of at least one additional termination factor in *T. kodakarensis*, but does render the cells sensitive to DNA damaging agents. The nucleic acid requirements, slow rate of termination and the sensitivity of strains lacking Eta to mutagens suggests that Eta may function analogously to the bacterial transcription-repair coupling factor Mfd. The combined *in vitro* and *in vivo* characterization of Eta demonstrates that factor-dependent transcription termination is employed in all extant life and reveals similarities in TEC stability and susceptibility to factor-dependent termination.

3.2 Results

Identification of a Euryarchaeal Transcription Termination Activity (Eta).

Our purified *in vitro* transcription system from *T. kodakarensis* permits promoter-directed transcription initiation on DNA templates attached to a solid-support, allowing stable TECs to be generated at defined template positions by nucleotide deprivation (Figure 3.1A) (56, 61). TECs are resistant to repeated washing and ^{32}P-labeled nascent transcripts are detected in the bound or pellet fraction. Any RNAs released by dissociation of the TEC can be recovered from the supernatant fraction, and thus the integrity of TECs can be monitored by the distribution of



D **Proteins identified via Mass Spectrometry in purified active fractions.**

Gene	Annotation	MASCOT	Molecular Weight, kDa
TK0566	DEAH box RNA helicase	645	96
TK0752	Acylamino acid releasing enzyme	1,120	72
TK1626	Hypothetical protein	726	32
TK2151	Archaeal transcription regulator	725	17

Figure 3.1. Identification of an archaeal termination factor. A. DNA templates contain a biotin moiety (blue B), a strong promoter, P_{hmtB}, a 376bp G-less cassette, and permit elongation to produce a full-length transcript of +450. **B.** Stalled TECs at the end of a G-less cassette (TEC_{+376}) were incubated with cell lysate to identify termination factors. **C.** Active fractions were identified as those that did not produce +450 transcripts when supplemented with lysate and ATP. **D.** Proteins identified via Mass Spectrometry in purified active fractions.

transcripts between the pellet and supernatant fractions. Washed, NTP-starved TECs can also be stimulated to resume elongation upon NTP addition, and the ability to resume elongation can be used to assess the stability of the TEC (Figure 3.1B).

Washed TECs stalled at the end of a 376 nt G-less cassette (TEC_{+376}) (Figure 3.1A) were incubated with partially purified *T. kodakarensis* lysates to permit any termination factors present an opportunity to act on the stalled TEC. Previous studies demonstrated the resistance of stalled TECs to the activities of purified strong helicases (57), suggesting that only legitimate transcription termination factors would act to disrupt the stalled TECs. Reactions were then supplemented with all four NTPs to determine if the TEC could resume transcription and generate a run-off transcript of 450 nt (Figure 3.1B). Fractionated lysates were defined as 'active' if the TECs incubated with these fractions were unable to resume elongation upon NTP addition (Figure 3.1C). Given that all known transcription termination factors are energy-dependent enzymes, an additional constraint, namely ATP-dependence of presumptive termination activity, was added to define 'active' fractions.

Despite complicating RNase contamination, several activities were chromatographically identified in *T. kodakarensis* lysates that were designated "active" fractions, implying the presence of multiple termination factors or the association of a single termination factor in differentially separate complexes. The complexity of the active fraction was refined by repeated chromatographic separations until only a few proteins were present. Separation by SDS-PAGE, followed by excision and mass spectrometry of the dominant bands identified the major proteins present in this purified active fraction. Of the four abundant factors, the ~96 kDa product of TK0566 was the dominant factor deemed responsible for the apparent ATP-dependent transcription termination activity (Figure 3.1D).

Eta is an energy-dependent transcription termination factor.

The failure of TECs to resume elongation upon NTP addition (Figure 3.1C) was suggestive but not definitive evidence of a transcription termination factor. To eliminate concerns of any additional or unidentified factor being responsible for the presumed termination activity, the protein product of TK0566 (832 aa; termed Eta; Figure 3.2A) was recombinantly expressed and purified (Figure 3.2B). To demonstrate that Eta is a *bona fide* termination factor recombinant Eta was added to *in vitro* transcription assays containing stalled, washed $TECs_{+58}$. TECs were stalled by NTP-deprivation on biotin-labeled DNA templates attached to streptavidin-coated paramagnetic particles (Figure 3.2C). This solid-support permits washing and separation of pellet and supernatant fractions to monitor dissociation of the TEC. In the absence of Eta, or in the presence of Eta but in the absence of an energy source, just 4-8% of $TECs_{+58}$ dissociate and release transcripts to solution (Figure 3.2D). In contrast, addition of Eta resulted in ATP-dependent release of RNA transcripts to solution (Figure 3.2D) resulting from dissociation of nearly half of all TECs. ATP and dATP both support Eta-mediated transcription termination, and dATP is used throughout to avoid supplying RNAP with ATP.

Eta-mediated transcription termination is dependent on ATP hydrolysis.

Eta is a superfamily II helicase family member and is annotated as a DEAD/DEAH box RNA helicase (59). Conserved Walker A and B motifs that are responsible for the binding and hydrolysis of ATP, respectively are easily identified in the central P-loop NTPase domain of Eta (Figure 3.2A). The N-terminus of Eta (residues 1-193) is less conserved and appears to contain a Zn-finger motif. The Walker B motif consensus sequence ‘hhhhDE’, contains an aspartate (D) residue that coordinates Mg^{2+} and a glutamate (E) that is essential for the NTP hydrolysis (62, 63).

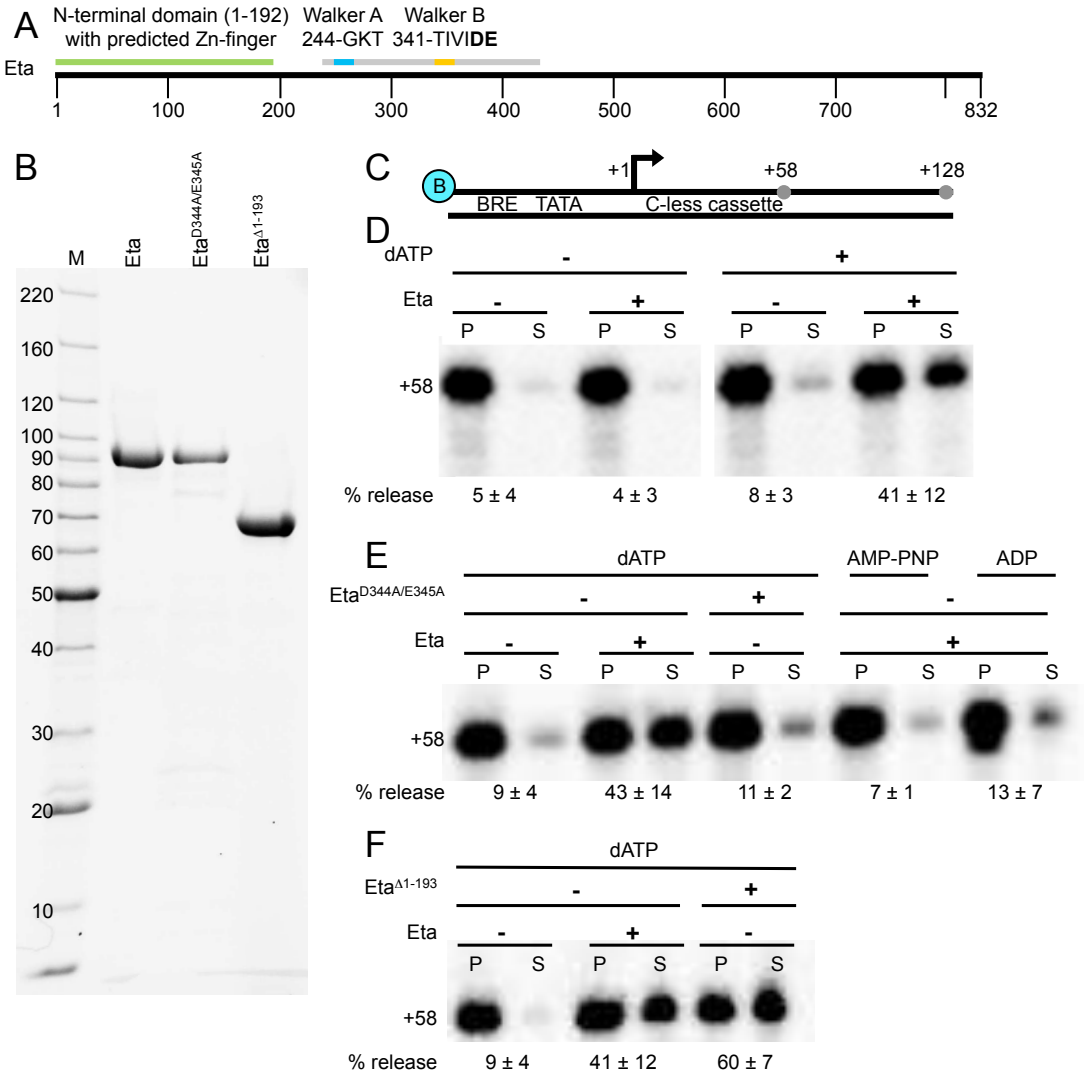


Figure 3.2. Eta is an energy-dependent transcription termination factor. A. Representation of Eta highlighting the N-terminal domain (green), the P-loop NTPase domain (gray), and the Walker A and B motifs (blue and orange, respectively). **B.** Recombinant Eta and Eta variants ($Eta^{D344A/E345A}$ and $Eta^{\Delta 1\text{-}193}$) were purified for use during in vitro transcription. **C.** DNA templates contain a biotin moiety, $P_{hmtB,}$ a 58 bp C-less cassette, and permit elongation to produce full-length +128 transcripts. **D.** Eta requires ATP or dATP to mediate transcription termination. **E.** Eta requires dATP hydrolysis to mediate transcription termination. **F.** $Eta^{\Delta 1\text{-}193}$ retains termination activity in vitro.

To determine if Eta-mediated transcript release is dependent on ATP-binding and hydrolysis ATP was replaced with ADP and the non-hydrolysable analog AMP-PNP but Eta could not stimulate RNA release to the supernatant in the presence of either co-factor (Figure 3.2E). An Eta variant ($Eta^{D344A+E345A}$; Figure 3.2B), wherein two critical Walker B residues were replaced by alanine (D344A + E345A), cannot stimulate transcription termination above background levels (Figure 3.2E). Near homogenous preparations of full-length Eta and $Eta^{D344A+E345A}$ were possible, but laborious. Deletion of the less conserved N-terminus (aa 1-193) resulted in a protein that chromatographed more uniformly and retained full termination activity in vitro (Figure 3.2F).

Eta-mediated transcription termination is relatively slow and is not competitive with transcription elongation at physiological NTP concentrations.

Addition of Eta to stalled TECs results in a near linear but slow rate of transcript release to solution (Figure 3.3A and 3.3B) when dATP is provided. Transcription elongation and known mechanisms of factor-dependent transcription termination are in competition *in vivo* and *in vitro*. Eta was added to stalled or slowly elongating TECs to monitor the ability of Eta to release transcripts from active TECs (Figure 3.3C). When TECs were generated, washed, and incubated in the absence of NTPs, most TECs remained intact but backtracked and shortened their nascent transcripts, presumably by a combination of endonucleolytic cleavage and reverse catalysis (Figure 3.3C, lanes 3&4). TECs could be maintained in the forward position by supplementing wash- and incubation-buffers with 10 μM ATP, GTP and UTP (Figure 3.3C, lanes 1&2). Addition of all four NTPs, at increasing concentrations, lead to release of the TECs from the C-less cassette and continued elongation to generate full-length +128 nt transcripts. Elongation to +128 was limited by sequences that direct pausing near ~+70 at very low NTP concentrations. With the exception of full-length +128 transcript, minimal transcript release to solution was observed in the absence of Eta addition.

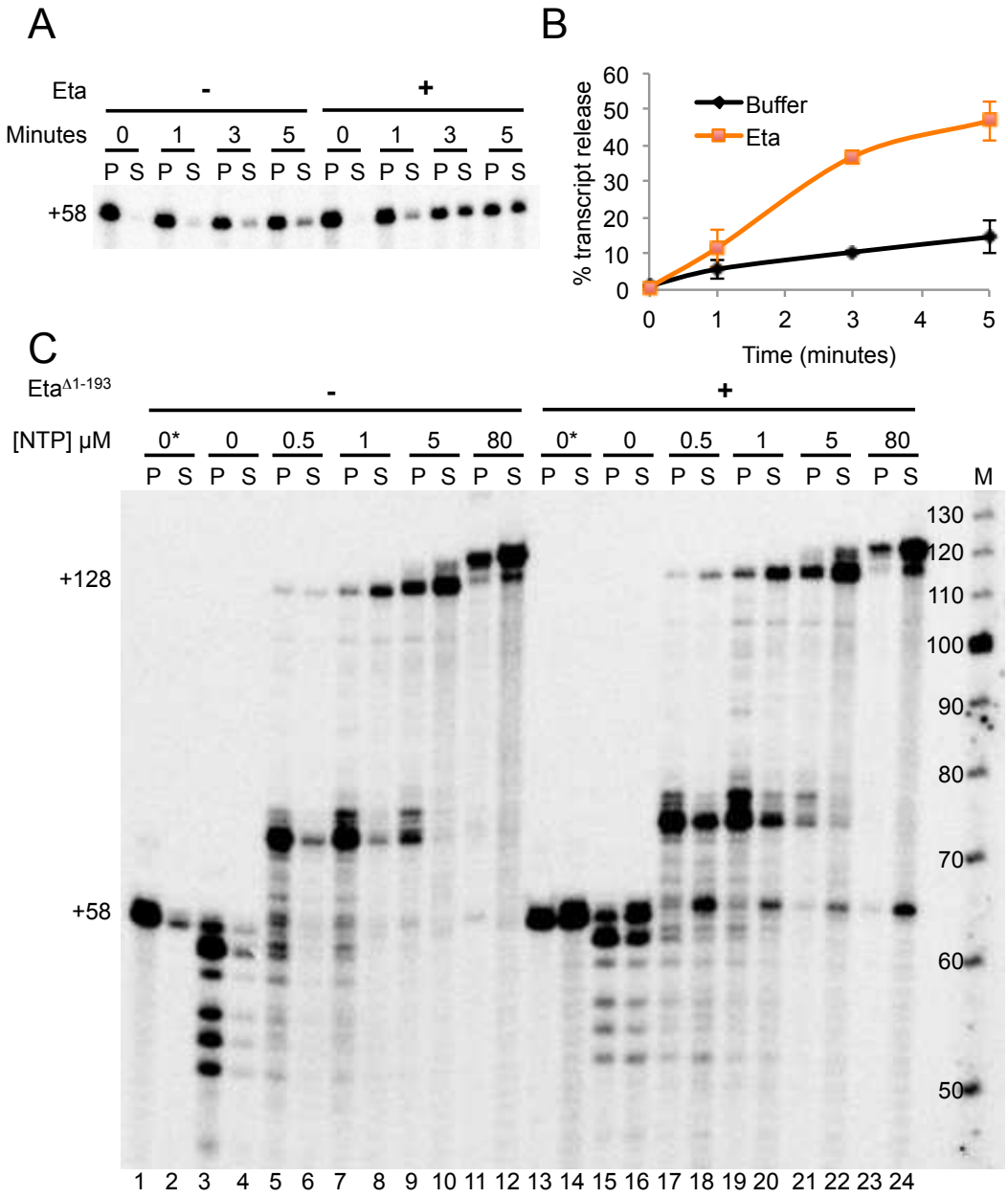


Figure 3.3. Eta mediates termination of stalled or slowly elongating TECs. A. Eta mediates the slow release of transcripts from stalled TECs$_{+58}$. **B.** Quantification from A. **C.** Eta-mediated termination limits backtracking and is only competitive with transcription elongation at low NTP concentrations. M, labled ssDNA marker to provide an approximation of RNA lengths.

In contrast addition of Eta to stalled or very slowly elongating TECs resulted in substantial transcript release to solution (Figure 3.3C, lanes 13-20). Eta released most stalled TECs to solution (Figure 3.3C, lanes 13 & 14), and addition of Eta limited the percentage of TECs that backtracked and shortened their nascent transcripts (Figure. 3.3C, lanes 15 & 16) suggestive that Eta keeps RNAP in the forward configuration and/or pushes forward backtracked complexes. Eta-mediated termination was most efficient for TECs that retained +58 transcripts and presumably reflect fully forward isomerization variants of the TEC (Figure. 3.3C, lanes 15 & 16). Eta was also capable of disrupting TECs stalled for long intervals at the low concentration NTP-dependent pause at ~+70 (Figure 3.3C, lanes 17-22), but as the interval of the pause was shortened by increasing NTP concentrations, the ability of Eta to direct transcription termination was compromised. At just 80 μM NTPs, Eta-mediated termination was non-competitive with elongation, with the exception of a small percentage of TEC_{+58} that likely failed to elongate quickly. Eta-mediated termination therefore appears too slow to result in the governance of coupled-transcription translation that is necessary for polarity.

Eta interacts with RNAP *in vivo*.

Genetic techniques were employed to generate a strain of *T. kodakarensis* wherein sequences encoding affinity- and epitope-tags were appended to the N-terminus of TK0566 (64). The resultant strain produced full-length, N-terminally tagged Eta from the natural TK0566 locus under normal regulatory control. Cell lysis, followed by gentle nickel-affinity purifications of complexes containing His6-tagged Eta were performed, and the identity of co-purifying partners revealed by mass spectrometry (Table 3.1). Several subunits of RNA polymerase were identified, supporting Eta-RNAP interactions *in vivo*. Several large helicases, an ATPase and NusA were also identified as Eta partners, suggesting that Eta may mediate interactions that facilitate events in addition to termination of TECs.

Table 3.1. Proteins copurifying with His_6-Eta from cellular lysates.

Gene	Annotation	MASCOT	Molecular Weight, kDa
TK0566	DEAH box RNA helicase	47,546	96
TK1314	ATPase	6,035	50
TK1015	Large-helicase related protein	1,093	106
TK1332	RNA helicase Ski2-like protein	5,483	129
TK1083	RNAP subunit beta	1,268	127
TK1081	RNAP subunit A''	216	44
TK1079	NusA	192	16
TK1503	RNAP subunit D	144	29
TK1499	RNAP subunit N	123	8
TK1082	RNAP subunit alpha	111	103

Eta-mediated termination requires access to upstream DNA sequences.

Despite the ability to identify RNAP in purifications of His6-Eta from *T. kodakarensis* cell lysates, Eta does not retain long-lived associations with RNAP in solution (Figure 3.4). Eta likely targets TECs through association with upstream or downstream DNA sequences or through the nascent transcript. To address the requirements for upstream DNA sequences in Eta-mediated transcription termination, stalled TECs were generated on templates that contained recognition sequences for the Ssp1 restriction endonuclease within the C-less cassette (Figure 3.5A and 3.5B). TECs were stable during the Ssp1 digestions, and monitoring radiolabeled DNA before and after cleavage demonstrated digestion of each template and release of the TEC to solution (Figure 3.5B). Stalling TECs by NTP-deprivation at +58 on template #1 allowed Ssp1 digestion at +18, thus retaining at least a full turn of accessible upstream DNA on the TECs that were released to solution. In contrast, Ssp1 digestion of template #2 at +37 resulted in release of TECs with minimal upstream sequences, as the footprint of TECs likely extends to approximately ~+40. Digestion of templates containing the +37 Ssp1-site were less efficient at releasing TECs to solution than the digestion at +18 in the presence of a TEC_{+58} (Figure 3.5B; lanes 7, 8, 15 & 16). The decrease in digestion efficiency is most likely due to RNAP obstructing access of Ssp1 to this RNAP-proximal site.

Radiolabeling the nascent transcripts in Ssp1-released TECs allowed the ability of Eta to disrupt the TECs to be inferred from the ability of the TECs to elongate upon addition of NTPs (Figure 3.4C). Ssp1 digestion was necessary to release any significant TECs to solution (Figure 3.5C; lane 2 vs 4 & 6 vs 8), and in the absence of Eta, resumed elongation upon NTP addition was readily evident by extension of the RNA to generate +128 nt transcripts (Figure 3.5 lanes 3 & 8). On the template with Ssp1-site 1, digestion produced TECs in solution (Figure 3.5; lane 4) that were largely not extended when Eta was present (Figure 3.5; lane 5), suggestive of Eta-mediated transcription termination being active under such conditions.

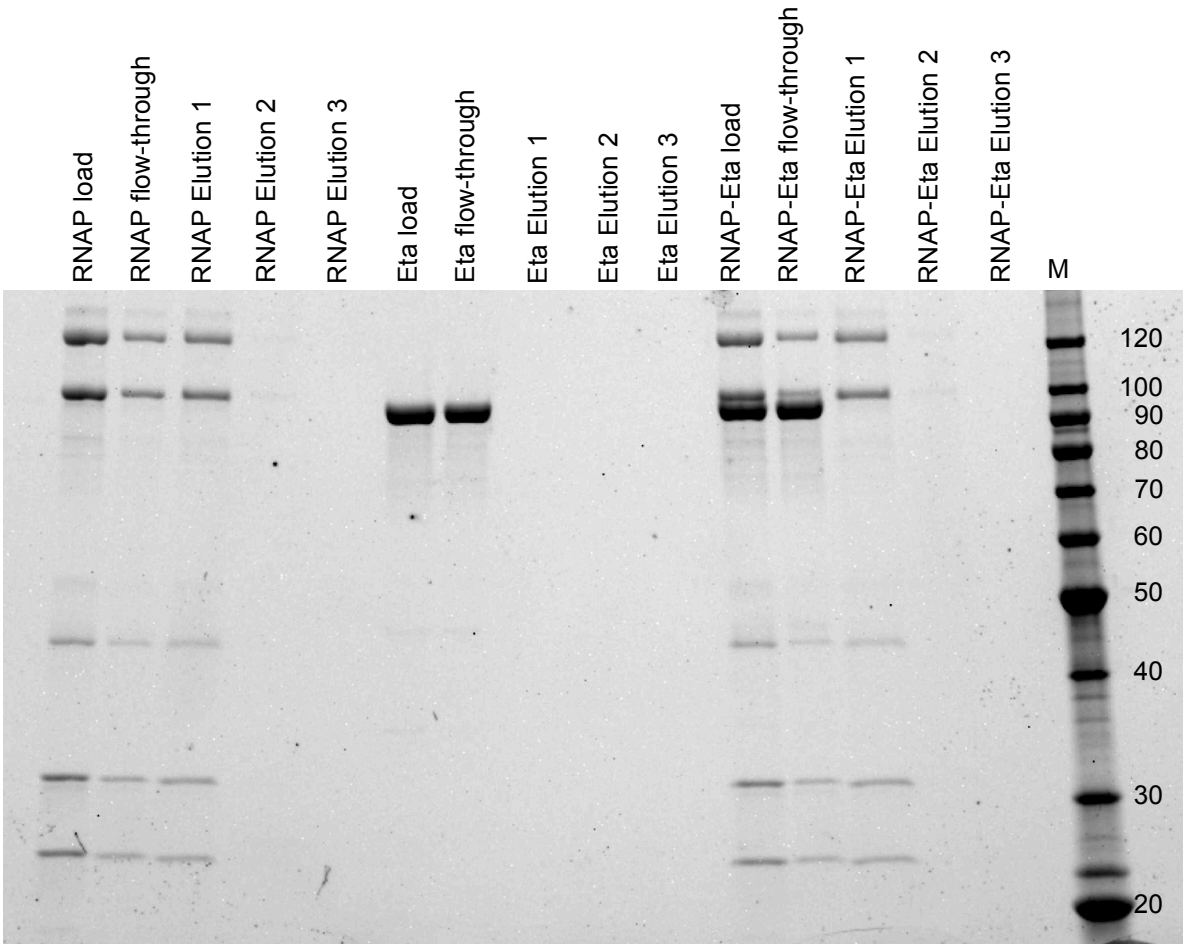


Figure 3.4. Eta and RNAP did not form long-lived interactions. A pull-down experiment was employed using a Nickel resin and only RNAP-his6 (positive control), only Eta (negative control) and RNAP-his6 + Eta. The portion of the sample that did not stick to the resin is indicated as 'flow-through' and the portion of the sample that eluted when a high imidazole buffer was added is indicated as 'elution'. Samples were resolved on an SDS-PAGE with a marker shown in kDa.

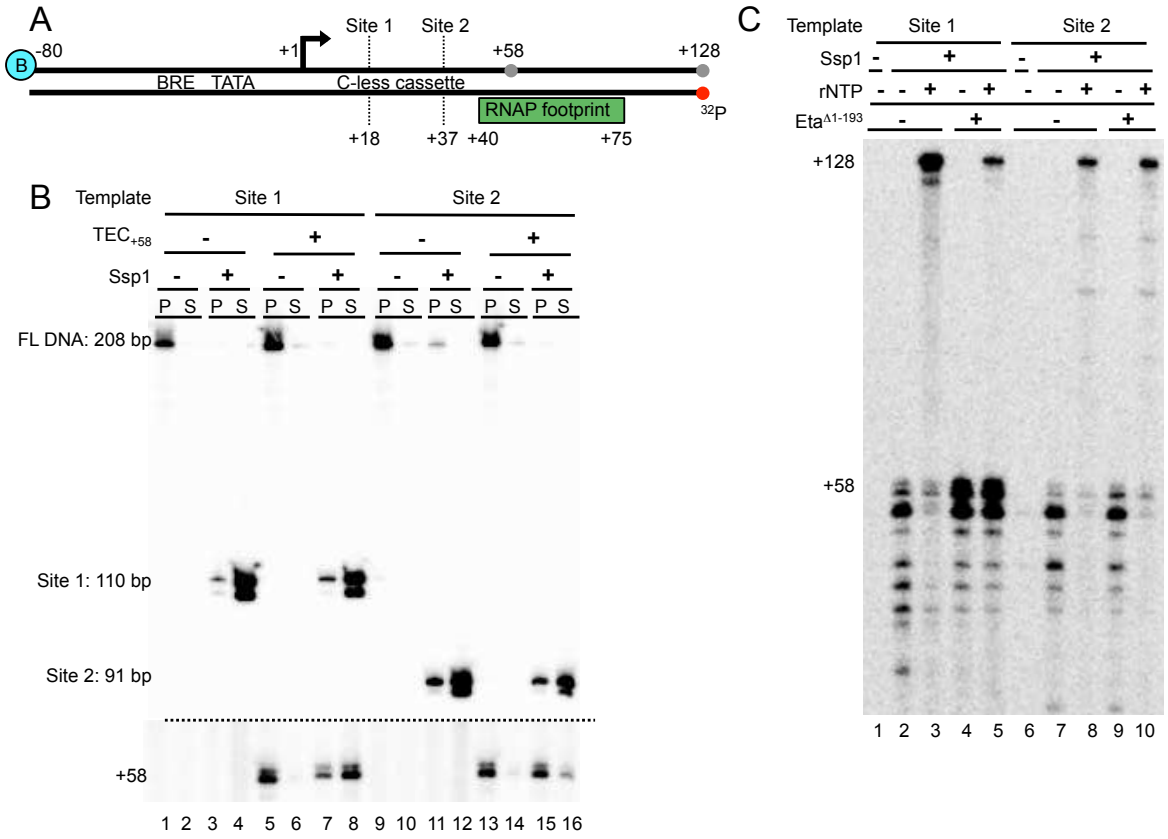


Figure 3.5. Eta-mediated termination requires upstream DNA sequences. A. DNA templates are identical except for the Ssp1-recognition sequences at positions +18 (Site 1) and +37 (Site 2) respectively. **B.** Ssp1 digestion releases stalled $TECs_{+58}$ to solution. DNA templates were radiolabeled at the 5' position of the template strand (red dot). **C.** Upstream DNA is required for Eta-mediated transcription termination *in vitro*.

In contrast, Ssp1-digestion of DNA templates containing site 2 produced TECs in solution (Figure 3.5; lane 9) largely were elongated to +128 (Figure 3.5; lane 10). Suggesting that Eta-mediated termination was not active when upstream DNA sequences were removed from the TECs.

To address requirements for downstream DNA for Eta-mediated transcription termination, $TECs_{+58}$ were assembled on templates that contained minimal accessible downstream DNA (Figure 3.6A; upper panel). By use of short templates that would only permit elongation to generate TEC_{+80}, essentially all downstream DNA sequences would be predicted to be enveloped by the TEC and not solvent accessible. Eta-mediated termination was efficient on such templates (Figure 3.6A; lower panel) suggestive that Eta-mediated transcription termination does not require downstream DNA sequences.

Any requirements for or use of RNA transcripts to target TECs for Eta-mediated transcription termination were tested in two complementary formats. First, (Figure 3.6B), $TECs_{+20}$ were generated by NTP deprivation to limit the extent of enzyme-accessible transcript sequences. Approximately half of such TECs were unstable at high temperature, but Eta-mediated transcription termination was still active, releasing nearly all transcripts to solution under such conditions (Figure 3.6B).

Given the spontaneous instability of TECs with short transcripts, we were concerned that such complexes may not represent TECs that have fully transitioned from initiation to elongation. To more accurately determine the requirements for any enzyme-accessible transcript sequences for Eta-mediated transcription termination, we generated $TECs_{+58}$, then added $RNaseI_f$ (an RNA endonuclease that cleaves at all RNA dinucleotide bonds) to cleave enzyme-accessible RNA to a minimum (Figure 3.6C). Generating stable TEC_{+58} (lanes 1 & 2) permitted 37˚C $RNaseI_f$ digestions (Figure 3.6C; lanes 3 & 4) that yielded stable TECs that contained a range of shortened, but RNAP-associated transcripts of just ~18-25 nucleotides on average.

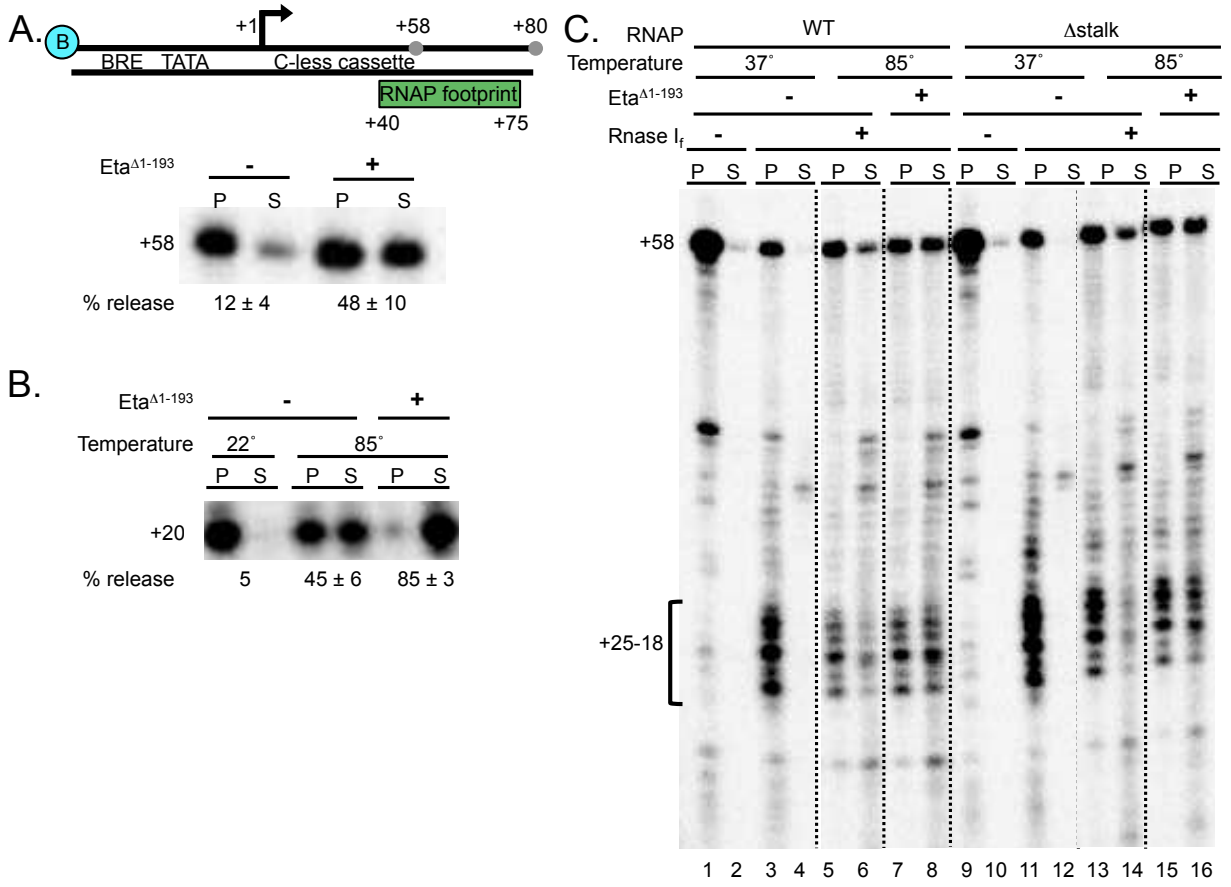


Figure 3.6. Enzyme-accessible downstream DNA nor nascent RNA is required for effective Eta-mediated termination. A. Downstream DNA is dispensable for Eta-mediated termination. **B.** Enzyme-accessible RNA sequences are not required for effective Eta-mediated termination. **C.** RNase I_f digestion of stable TEC_{+58} eliminates enzyme accessible RNA but does not block Eta-mediated termination.

Challenging such complexes at physiological temperatures (85˚C, Figure 3.6C; lanes 5 & 6) resulted in minimal disassociation of TECs containing ~18-25 nt transcripts. Addition of Eta, however, resulted in release of the bulk of these complexes to solution

Previous evidence shows that RNAP subunits E and F (i.e., the stalk domain) interact with the nascent RNA (65–69). Furthermore, it has been established that the deletion of the stalk domain has no effect on promoter-dependent transcription initiation, abortive transcript synthesis, transcript elongation or termination (69). We hypothesized that RNase treatment using a RNAP variant deleted for the stalk domain would allow for more RNA to be exposed to solution and thus cleaved by RNase I_f. The experiment described above was repeated using a $RNAP^{\Delta stalk}$ variant (Figure 3.6C; lanes 9-16). Surprisingly, RNase I_f left an identical footprint in the presence and absence of the RNAP stalk domain. Challenging TEC_{+58} complexes at physiological temperatures (85˚C, Figure 3.6C; lanes 13 & 14) resulted in minimal disassociation of TECs containing ~18-25 nt transcripts. Addition of Eta, however, resulted in release of the bulk of these complexes to solution. The ability of Eta to drive release of transcripts from TECs with minimal enzyme-accessible RNA in the presence and absence of the RNAP stalk domain suggests that Eta-mediated termination is not reliant on RNA transcript sequences or the stalk domain of RNAP to mediate termination.

Eta is non-essential but deletion impacts growth of *T. kodakarensis*.

T. kodakarensis readily incorporates exogenous DNA into the genome permitting rapid and facile strain construction (70). Non-replicative plasmid sequences can be targeted via sequence homology and temporarily inserted, then excised from the genome to delete or modify a specific locus. Established genetic techniques (71, 72) were employed to markerlessly delete the entirety of the TK0566 coding sequence from the *T. kodakarensis* genome (Figure 3.7A), and to introduce allelic changes to TK0566 such that a D344A + E345A variant Eta was encoded. Diagnostic PCRs (Figure 3.7B) using genomic DNA isolated from parental and

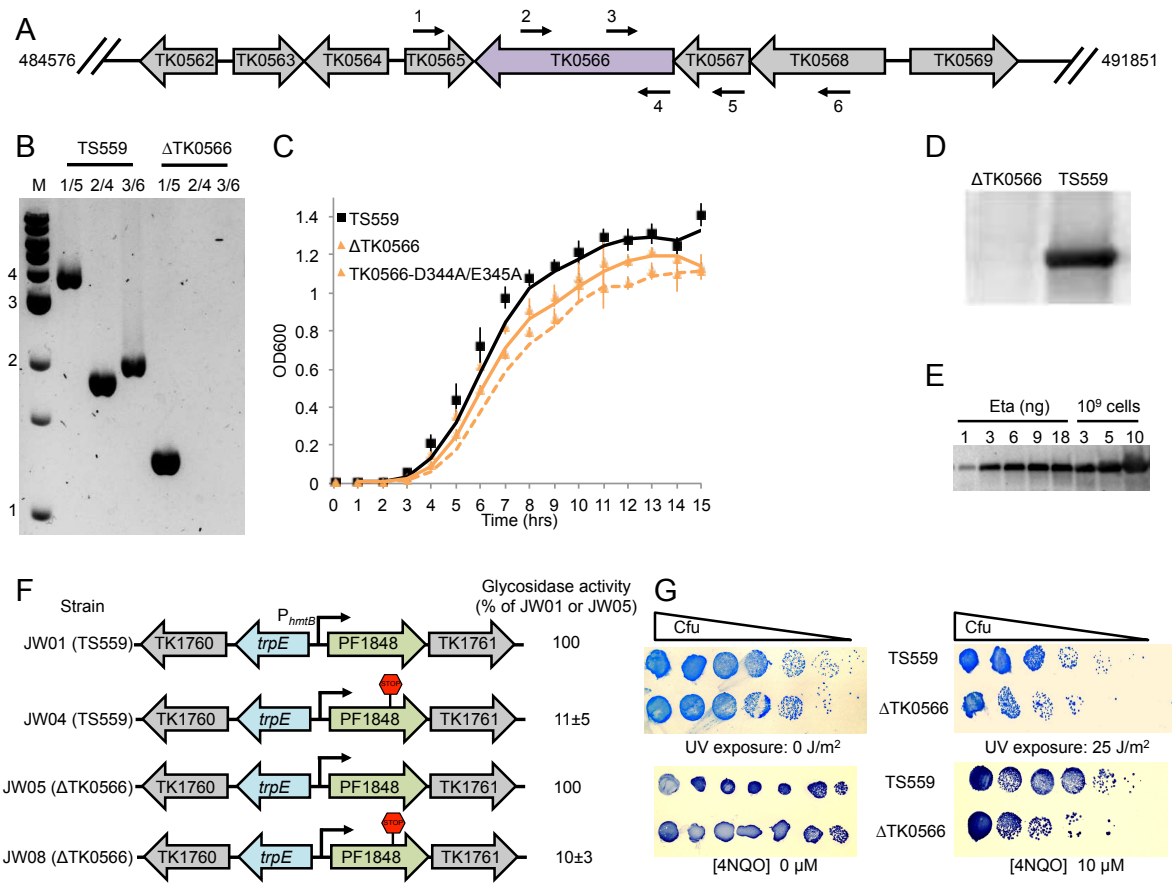


Figure 3.7. Eta is non-essential, is not responsible for polarity *in vivo*, but loss of Eta limits growth and increases DNA damage susceptibility. A. A map of the *T. kodakarensis* genome highlighting TK0566 and surrounding sequences. Primer binding positions 1-6 are designated by arrows. **B.** Diagnostic PCRs confirm deletion of TK0566 sequences. **C.** Deletion or inactivation of Eta hinders growth rate and final cell densities. The curves and errors shown represent means and standard averages of triplicate technical repeats of triplicate biological samples. **D.** Western blots using anti-Eta antibodies confirm that deletion of TK0566 eliminates production of Eta *in vivo*. **E.** Semi-quantitative western blots using anti-Eta antibodies suggest that Eta is present in low concentrations *in vivo*. **F.** Eta is not responsible for polarity *in vivo*. The presence or absence of the nonsense codon in PF1848 is noted and the percent β-glycosidase activity is reported as the mean and standard average of triplicate technical repeats of triplicate biological samples. **G.** Deletion of Eta increases sensitivity to DNA damaging agents.

deletion stains confirmed excision of the TK0566 sequence from the natural locus and the absence of an amplicon resultant from primers complementary to TK0566 sequences eliminated concerns of genomic rearrangements or deletion of TK0566 from only some of the many genomes retained in *T. kodakarensis* cells (73). Amplification with primer pairs adjacent to TK0566 sequences (#1 and #5) result in ~3.8 Kbp and just ~1.2 Kbp amplicons from parental (TS559) and ΔTK0566 genomic DNA preparations, respectively. Amplifications with primer pairs complementary to TK0566 sequences (#2 and #4), or with one primer complementary to TK0566 and one adjacent primer (#3 and #6) were successful only from TS559 genomic DNA preparations (Figure 3.7B).

Deletion or inactivation of Eta slows and limits growth of *T. kodakarensis* (Figure 3.7C). Western blots using anti-Eta antibodies confirm loss of Eta in ΔTK0566 strains (Figure 3.7D). Semi-quantitative Western blots (Figure 3.7E), comparing the signal from purified Eta to the Eta-signal present in the lysates of total *T. kodakarensis* cells, suggests that Eta is normally present at low (~50 copies per cell) levels compared to RNAP levels (~2,000-3,000 copies per cell).

Eta is not responsible for polarity.

Transcription and translation are tightly coupled in archaea and bacteria (74–76). Disrupting the normally tight association of transcription and translation results in polar repression of downstream expression due to premature factor-driven transcription termination (52, 77, 78). The bacterial termination factor Rho is responsible for polarity, but there are no Rho homolgous encoded in archaeal genomes. To determine if Eta is the factor responsible for polarity in *T. kodakarensis* we compared repression of expression of a downstream gene in an operon with non-sense codons introduced into an upstream gene in the same operon in strains with and without TK0566 (Eta) (Figure 3.7F). Activity from the downstream reporter gene was substantially reduced by introduction of nonsense codons into the upstream coding but the

degree of reduction was not changed by the presence or absence of Eta. These results indicate that Eta-mediated termination is unlikely to direct polarity and that another factor is encoded in *T. kodakarensis* that can disrupt the TEC.

Deletion of TK0566 (Eta) increases sensitivity to DNA damaging agents.

The requirements for Eta-mediated transcription termination – energy-dependence, a static or slowly elongating TEC, access to upstream DNA sequences and no requirements for downstream DNA or RNA sequences – are reminiscent of the only other prokaryotic factor capable of disrupting the TEC, namely Mfd. Mfd initiates RNAP removal and transcription-coupled DNA repair (TCR) in Bacteria and cells deleted for *mfd* exhibit a mild phenotype to some DNA damaging agents (79, 80). Although TCR has not yet been investigated in *T. kodakarensis*, recent support for TCR in related euryarchaea has been provided (53), and thus we accessed the potential for Eta to influence DNA repair by challenging parental and Eta-deleted strains of *T. kodakarensis* to common DNA damaging agents. Exposure to UV light limited the growth of both strains, but the strain deleted for Eta was substantially (at least an order of magnitude) more sensitive to UV exposure than the parental strain (Figure 3.7G, top panels). Introduction of the heterocyclic mutagen 4-nitroquinoline 1-oxide (4NQO) similarly limits cellular growth, and again the strain deleted for TK0566 is at least an order of magnitude more sensitive than the parental strain (Figure 3.7G, bottom panels).

3.3 Discussion

Processive transcription necessitates an extremely stable transcription elongation complex. Biochemical and structural studies demonstrate that the overall stability of the TEC is composite, with inputs from RNAP-DNA, DNA-RNA, and RNAP-RNA interactions collectively stabilizing the elongation complex. DNA sequences, and the encoded RNA sequences and structures that form within or adjacent to RNAP can disrupt these contacts and destabilize the

TEC driving transcription termination. Intrinsic termination sequences often suffice to separate independent genes and operons by blocking continued downstream transcription, but not all termination events can be initiated via DNA sequence alone. Scenarios arise in all life where the stable TEC may halt transcription at any position, most likely in response to protein roadblocks or DNA damage, and these arrested TECs must be removed for additional rounds of transcription to occur and to maintain genome stability. In bacteria and archaea, transcription and translation are normally coupled and the uncoupling of these apparatuses offers regulatory potential that is exploited by factors (e.g. Rho in bacteria), that disrupt the TEC and thus limit downstream expression.

We demonstrate, for the first time, archaeal factor-dependent transcription termination and characterize a novel *bona fide* archaeal transcription termination factor *in vivo* and *in vitro*. We purified this biochemical activity directly from cell lysates and then demonstrated that a single protein, now termed Eta, drives TEC disassembly and release of RNA to solution. Our results confirm that factor-dependent transcription termination is conserved in all life, and that the factors capable of disrupting the TEC are all energy-dependent (5, 36, 81–83).

Factor-mediated disruption of the TEC in bacteria and eukarya is discriminatory to ensure that functional TECs are generally not prematurely terminated. Eta-mediated termination results in release of the nascent transcript from stalled or slow elongating TECs to solution, but does so slowly. Eta-mediated termination is not competitive with standard RNAP elongation rates thus functional archaeal TECs are unlikely to be targeted for disruption. In contrast, Eta likely targets RNAPs that are translocationally blocked at sites of DNA damage or that are arrested due to chromatin or other protein roadblocks.

Deletion of TK0566 (encoding Eta) or introduction of mutations that encode inactive variants of Eta to the *T. kodakarensis* genome is possible and results in strains with slow-growth and DNA damage-sensitive phenotypes. Strains lacking Eta permitted investigation of the role of Eta in polar repression of transcription. The low abundance and slow activity of Eta were not

supportive of a general governance role in RNA 3'-end formation as is true for Rho in Bacteria, and we demonstrate that polar-repression occurs to the same extent in strains lacking Eta-mediated termination. The presence of additional "termination active" fractions suggests *T. kodakarensis* encodes additional transcription termination factors and these presumptive transcription termination activities remain to be identified. The increased sensitivity of strains lacking Eta to DNA damage is suggestive of a potential role for Eta in recognition and removal of TECs stalled at DNA lesions. The presence of transcription coupled DNA repair (TCR) in *T. kodakarensis* and any role for Eta in archaeal TCR remains to be determined.

Several models have been proposed for factor-dependent transcription termination and our results suggest that Eta follows a similar mechanism of termination as Mfd (Figure 3.8). Eta likely binds to the upstream DNA of stalled TECs and, using ATP-hydrolysis, translocates along the DNA pushing RNAP forward. In the absence of continued synthesis, RNAP is hypertranslocated and/or the transcription bubble collapses, resulting in disassociation of the TEC and transcription termination.

The conservation of Eta in most archaeal-lineages argues that factor-mediated termination is commonplace in archaeal regulatory strategies. Evidence of additional termination factors likely encoded in *T. kodakarensis* is further supportive of post-initiation regulation strategies underlying archaeal transcription regulation. Continued insight into the mechanism of Eta-mediated transcription termination should provide insight into shared aspects of TEC stability and highlight susceptibilities of the TEC that can be exploited for regulatory control.

In the future, it is necessary to further characterize the mechanism of Eta mediated termination and in Chapter 4 we begin to do such. In addition, it is critical to elucidate the role of Eta *in vivo*. We hypothesize that Eta is involved in transcription-coupled DNA repair and removes RNAPs that are paused/arrested at DNA lesions and potentially recruits DNA repair factors. Additional work is now being done to investigate the role of Eta in TCR but this work is outside the scope of this thesis.

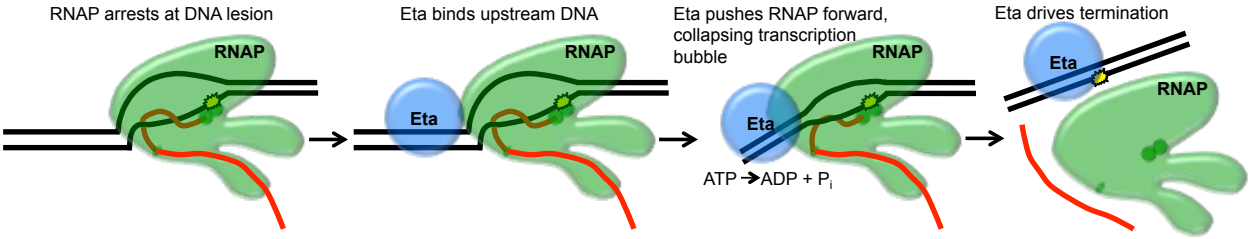


Figure 3.8. Model of Eta-mediated transcription termination. Eta recognizes RNAP arrested at the site of a DNA lesion (yellow). Eta binds to the upstream DNA and uses ATP hydrolysis to push RNAP forward causing transcription bubble collapse and promoting transcription termination.

The identification of Eta arose from a unique screening that used *T. kodakarensis* whole cell lysate and an *in vitro* transcription termination assay. This screening identified Eta and although other termination factors were not identified during the screening (due to high RNase activity) it was clear that there were other termination activities present. In addition, we have shown that Eta is not responsible for polarity and global 3' end formation of RNA transcripts. This data strongly suggests that archaeal genomes encode a second transcription termination factor. Through the combination of bioinformatics, a high-throughput ATPase assay, and protein-protein binding assays we attempt to characterize additional termination factor(s) responsible for global 3' end formation of RNA transcripts. The aforementioned studies are outlined in Chapter 5 of this thesis.

3.5 Methods

Strains and Plasmids.

All of the *T. kodakarensis* strains used in this study were constructed by the same plasmid transformation, homologous recombination, and selection procedures as described previously (71). pQE-80L-Eta was generated to overexpress TK0566 (encoding for Eta). Plasmid variants (termed pQE-80L-Eta$^{D344A/E345A}$ and pQE-80L-Eta$^{\Delta 1\text{-}193}$) were generated via site-directed mutagenesis (Agilent) to overexpress the Eta variant.

Protein purifications.

RNAP, TBP, and TFB were purified as described in ref (56). rEta^{Wt}, rEta$^{D344A/E345A}$, and rEta$^{\Delta 1\text{-}193}$ were purified from Rosetta2 (DE3) cells carrying pQE-80L-Eta, pQE-80L-Eta$^{D344A/E345A}$, or pQE-80L-Eta$^{\Delta 1\text{-}193}$ respectively, grown in LB medium supplemented with 34 µg/ml chloramphenicol and 40 µg/mL of ampicillin. Eta-expression was induced by addition of 0.5 mM final concentration IPTG and cultures were grown overnight (~16 hrs) at 17°C. Biomass was harvested, resuspended and sonicated in lysis buffer (25 mM Tris-HCl pH 6.8, 10% glycerol

(v/v), 0.1 mM EDTA, 100 mM NaCl, 1mM β-ME). Eta was partially purified from clarified cell lysate by heat treatment at 85°C for 20 minutes, followed by passage and fractionation of the cleared supernatant through three separate chromatographic columns (S-sepharose, heparin, and MonoQ; GE Healthcare). An identical purification was done for the $Eta^{D344A/E345A}$ and $rEta^{\Delta 1-193}$ variants. Eta^{wt} and variant-Eta preparations were dialyzed into storage buffer (20 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 100 mM KCl, 50% glycerol) and quantified by Bradford assays (84).

Biochemical Identification of Transcription Termination Activities in *T. kodakarensis* lysates.

Biomass was recovered from *T. kodakarensis* strain TS413 (*rpoL*-HA-his_6) grown at 85°C to mid-exponential phase ($OD_{600} \cong 0.5$). Cells were resuspended and sonicated in lysis buffer (25 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol). Clarified cell lysate was loaded to a Ni^{2+}-charged matrix. Flow-through was collected, diluted with 3 volumes of 25 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 0.1 mM EDTA and loaded to a heparin column. The column was washed with the same buffer containing 160 mM NaCl and a linear gradient over 50 mL to a final concentration of 1 M NaCl was employed followed by an isocratic gradient to a final concentration of 2 M NaCl. Fractions collected were used during the *in vitro* transcription termination assay and deemed active or inactive (see *in vitro* transcription section below). Active fractions were pooled and diluted with 25 mM Tris-HCl pH 8.8, 10 mM $MgCl_2$ to reduce conductivity and loaded to a Q-sepharose column. Fractions were eluted with a linear gradient over 80mL to a final concentration of 400 mM NaCl followed by an isocratic gradient to a concentration of 1 M NaCl. Active fractions were identified, pool, concentrated and loaded to a Superdex 200 column in 25 mM Tris-HCl pH 6.8, 10% glycerol. Active fractions were identified, pooled and loaded on a MonoS column. The sample was eluted over 25 mL to a final concentration of 1 M NaCl. Active fractions were identified, pooled and loaded to a MonoQ column. The column was washed with 5 mM NaCl and a linear gradient over 35 mL to a

concentration of 500 mM NaCl was employed followed by a 10 mL linear gradient with to a concentration of 1 M NaCl. Active fractions were identified and resolved on a 15% discontinuous SDS-PAGE, and four abundant proteins were identified using mass spectrometry.

***In vitro* transcription.**

Plasmids pJW21, pJW29, pJW30 and pJW32 were generated by inserting a gblock (IDT) containing the P_{hmtB} promoter and a 128 bp full-length cassette into the pQE-80L vector. The DNA templates used during *in vitro* transcription were amplified from plasmids using standard molecular biology followed by purification of the linear, biotinylated templates (Agencourt AMPure XP; Beckman Coulter).

In vitro transcription reactions were generally carried out as previously described (56). Briefly, a linear DNA template containing a biotin moiety generated by PCR amplification (10 nM), *T. kodakarensis* RNAP (40 nM), transcription factor B (80 nM), TATA-binding protein (80 nM), and ApC (1.5 μM) were mixed to form a transcription buffer (final concentrations, 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 5 mM DTT, and 250 mM KCl).

Figure 3.1C: To obtain stalled elongation complexes, starting NTPs (200 μM ATP, 200 μM UTP, 10 μM CTP, and 10 μCi [α-^{32}P]-CTP, 3,000Ci/mmol) were added to the reactions that were incubated at 85°C for 5 minutes. TEC_{+376} complexes were captured with streptavidin-coated magnetic beads (Promega) and washed three times in wash buffer (20 mM Tris–HCl pH 8.0, 0.1 mM EDTA, 250 mM KCl, 4 mM $MgCl_2$, 20 μg/ml BSA). TEC_{+376} complexes were resuspended in 250 mM KCl, 20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 2 mM DTT, then incubated at 85˚C -/+ 5 mM ATP in the presence or absence of partially purified lysates from *T. kodakarensis*. After 5 minutes, rNTPs (each rNTP at a final concentration of 200 μM) were added to the reaction and incubation at 85°C was continued for another 10 minutes. 5 reaction volumes of 1.2X STOP buffer (0.6 M Tris-HCl pH 8.0, 12 mM EDTA) containing 7 μg tRNA

(total) was added to the reaction which was then subjected to an equal volume phenol/chloroform/isoamyl alcohol (25:24:1, by vol) extraction. Radiolabeled RNA transcripts were precipitated from the aqueous phase with alcohol. Templates and RNA transcripts were separated on 10%, 12%, 15%, or 20% denaturing polyacrylamide gels containing 8M urea (National Diagnostic Urea Gel). Radiolabeled RNA was detected using phosphorimaging (GE Healthcare) and analyzed using GE Imagequant Software 5.2.

Figure 3.2, Figure 3.3A and Figure 3.5A: To obtain stalled elongation complexes, starting NTPs (200 μM ATP, 200 μM GTP, 10 μM UTP, and 10 μCi [α-^{32}P]-UTP) were added to the reactions and incubated at 85°C for 5 minutes. TECs $_{+58}$ were captured with Ni^{2+}-coated magnetic beads and washed three times in wash buffer then resuspended in a modified transcription buffer containing low concentrations of ATP, GTP, and UTP (20 mM Tris-HCl pH 8, 250 mM KCl, 10 mM $MgCl_2$, 2 mM DTT, 10 μM ATP, 10 μM GTP, and 10 μM UTP). 10 μL aliquouts were combined with equal volume reactions containing 15 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 2 mM DTT, +/- 4 mM energy source (dATP, AMP-PNP, or ADP), +/- 500 nM purified Eta or Eta-variant. Reactions were incubated at 85°C for 5 minutes (or in Figure 3.3A for 0, 1, 3, and 5 minutes), and then streptavidin-coated magnetic beads were used to separate reactions into pellet and supernatant fractions. Pellet and supernatant fractions were incubated with STOP buffer, extracted, and RNA transcripts purified as above. RNA transcripts were quantified using GE Imagequant Software 5.2. Release was calculated by quantifying transcripts in the supernatant divided by transcripts quantified in the supernatant and pellet.

Figure 3.3C: Washed TECs$_{+58}$ were generated as above, then 10 μl aliquots were combined with equal volume reactions containing 15 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 2 mM DTT, 4 mM dATP, and +/- 500 nM Eta. Reactions in lanes 1&2 and lanes 13&14 were supplemented with 10μM each of UTP, GTP, ATP. The concentrations of NTPs provided to allow elongation to +128 during 7 minutes of incubation at 85˚C is listed in the Figureure. Streptavidin-coated magnetic beads were used to separate reactions into pellet and supernatant

fractions. Pellet and supernatant fractions were incubated with STOP buffer, extracted, and RNA transcripts purified and analyzed as above.

Figure 3.4B: Washed $TECs_{+58}$ were generated on ^{32}P-labeled, biotinylated templates as above, then TEC_{+58} complexes were resuspended in digestion buffer (1X NEB Buffer 2.1, +/- 8U Ssp1-HF) and incubated at 37°C for 30 minutes. Streptavidin-coated magnetic beads were used to separate reactions into pellet and supernatant fractions. Pellet and supernatant fractions were incubated with STOP buffer, extracted, and DNA templates and RNA transcripts purified and analyzed as above.

Figure 3.4C: Washed $TECs_{+58}$ were generated on ^{32}P-labeled, biotinylated templates as above, then TEC_{+58} complexes were resuspended in digestion buffer (1X NEB Buffer 2.1, +/- 8U Ssp1-HF) and incubated at 37°C for 30 minutes. Streptavidin-coated magnetic beads were used to separate reactions into pellet and supernatant fractions, and the pellet fraction was discarded. TEC_{+58} complexes retained in the supernatant were captured with Ni^{2+}-coated magnetic beads and washed two times in wash buffer then resuspended in a modified transcription buffer containing low concentrations of ATP, GTP, and UTP (20 mM Tris-HCl pH 8, 250 mM KCl, 10 mM $MgCl_2$, 2 mM DTT, 10 μM ATP, 10 μM GTP, and 10 μM UTP, 4 mM dATP, +/- 500 nM Eta). Reactions were incubated at 85°C for 7 minutes followed by the addition of rNTPS (each rNTP at a final concentration of 200μM) to the reaction and incubation was extended at 85°C for 5 additional minutes. 5 volumes of 1.2X STOP buffer containing 7 μg tRNA (total) was added, the reactions were extracted, RNA transcripts purified and analyzed as above.

Figure 3.5B: To obtain stalled elongation complexes, starting NTPs (200 μM ATP, 200 μM GTP, 10 μM UTP, and 10 μCi [α-^{32}P]-UTP) were added to the reactions and incubated at 85°C for 5 minutes. TEC_{+20} complexes were captured with Ni^{2+}-coated magnetic beads and washed three times in wash buffer then resuspended in a modified transcription buffer containing low concentrations of ATP, GTP, and UTP (20 mM Tris-HCl pH 8, 250 mM KCl, 10 mM $MgCl_2$, 2 mM DTT, 10 μM ATP, 10 μM GTP, and 10 μM UTP). 10 μL aliquouts were

combined with equal volume reactions containing 15 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 2 mM DTT, 4 mM dATP, +/- 500 nM purified Eta. Reactions were incubated at 85°C for 5 minutes, and then streptavidin-coated magnetic beads were used to separate reactions into pellet and supernatant fractions. Pellet and supernatant fractions were incubated with STOP buffer, extracted, and RNA transcripts purified and analyzed as above.

Figure 3.5C: To obtain stalled elongation complexes, starting NTPs (200 μM ATP, 200 μM GTP, 10 μM UTP, and 10 μCi [α-^{32}P]-UTP) were added to the reactions and incubated at 85°C for 5 minutes with $RNAP^{Wt}$ or $RNAP^{\Delta stalk}$. $TECs_{+58}$ were captured with Ni^{2+}-coated magnetic beads and washed three times in wash buffer then resuspended in RNase buffer (1X NEB Buffer 2, +/- 8U RNase I_f). Reactions were incubated at 37°C for ten minutes and washed two times in wash buffer and resuspended in transcription buffer (20 mM Tris-HCl pH 8, 250 mM KCl, 10 mM $MgCl_2$, 2 mM DTT, 4 mM dATP, +/- 500 nM Eta). Some of the reactions were incubated at 85°C for 7 minutes and then streptavidin-coated magnetic beads were used to separate all reactions into pellet and supernatant fractions. Pellet and supernatant fractions were incubated with STOP buffer, extracted, and RNA transcripts purified and analyzed as above.

Western blots.

rEta was used as antigen to prepare polyclonal antibodies in rabbits at Cocalico Biologicals, Inc. In Figure 5D, 15 μg of mid-log clarified cell lysate from *T. kodakarensis* strains TS559 and ΔTK0566 were resolved via SDS-PAGE, transferred to PVDF, and probed with primary anti-Eta antibodies followed by secondary anti-rabbit IgG-alkaline-phosphatase conjugates (Promega). In Figure 5E, the specified amounts of purified Eta or extracts from lysed strain TS559 *T. kodakarensis* are listed. These proteins were resolved via SDS-PAGE, transferred to PVDF, and probed with primary anti-Eta antibodies followed by secondary anti-rabbit IgG-HRP conjugates (Promega).

Polarity assay.

Plasmids and strains generated for the polarity assay were constructed as previously described except the parental strain was TS559 or ΔTK0566 (52). β-glycosidase activity was measured for each strain as previously described (52). Percent activity was calculated by comparing activities of identical strains (TS559 or ΔTK0566) with and without stop codons in PF1848.

DNA damage assays.

UV irradiation assays: *T. kodakarensis* strains TS559 (parental) and ΔTK0566 were grown to mid-exponential phase ($OD_{600\ nm}$ = 0.6) in ASW-YT-pyruvate media containing agmatine (71). ~1 x 10^8 cells were anaerobically harvested, collected via centrifugation, and resuspended in 100 µL of 0.8X ASW. Washed cells were serially diluted with 0.8X ASW and spotted (in duplicate) onto ASW-YT plates that were allowed to dry for 5 minutes. A portable, 254 nm UV lamp was used to provide a dose of 25 J/m^2 to one plate, then both plates were incubated under anaerobic conditions at 85˚C for 18 hrs to permit colony growth. Cells were visualized by transferring the resultant colonies to PVDF followed by staining with Coomassie Brilliant Blue.

4-Nitroquinoline 1-oxide (4NQO) sensitivity assays: *T. kodakarensis* strains TS559 (parental) and ΔTK0566 were grown to mid-exponential phase ($OD_{600\ nm}$ = 0.6) in ASW-YT-pyruvate media containing agmatine (71). ~1 x 10^8 cells were anaerobically harvested, collected via centrifugation, and resuspended in 100 µL of 0.8X ASW. Washed cells were serially diluted with 0.8X ASW and spotted (in duplicate) onto ASW-YT plates lacking or containing 10 µM 4NQO. Plates were incubated at 85°C for 18h and the cells were visualized by transferring to PVDF followed by staining with Coomassie Brilliant Blue.

Isolation of His_6-Eta containing complexes and MS-identification of binding partners.

T. kodakarensis cells were harvested by centrifugation from 5 L cultures grown to mid-exponential phase at 85°C in ASW-YT medium supplemented with 5 g sodium pyruvate/L. The cells were resuspended in 30 ml of buffer A (25 mM Tris-HCl pH 8, 500 mM NaCl, 10 mM imidazole, 10% glycerol) and lysed by sonication. Clarified lysate was loaded onto a 1-ml HiTRAP chelating column (GE Healthcare) preequilibrated with $NiSO_4$. The column was washed with buffer A, and proteins were eluted using a linear imidazole gradient from buffer A to 60% buffer B (25 mM Tris-HCl pH 8, 100 mM NaCl, 150 mM imidazole, 10% glycerol). Fractions that contained the tagged protein were identified by Western blotting, pooled, and dialyzed against buffer C (10 mM Tris-HCl pH 8). The pooled samples were lyophilized and resuspended in 10mM Tris-HCl pH 8.

The proteins were identified by 1D SDS-PAGE fractionation followed by LC/MSMS at the Ohio State University mass spectrometry facility. A MASCOT score of >100 was considered meaningful. To obtain such a score, a minimum of two unique peptide fragments usually had to be identified from the same protein. Protein isolation and mass spectrometry analyses of lysates from *T. kodakarensis* TS559 were also undertaken. From these controls, several *T. kodakarensis* proteins were identified that bound and eluted from the Ni^{2+}-charged matrix in the absence of a His_6-tagged protein. The proteins identified in the Eta sample that had MASCOT scores of >100 and were not also present in the control samples are listed in Table 3.1.

In vitro RNAP-Eta binding study.

8 µg of purified RNA polymerase, 8 µg of recombinant Eta, and 8 µg of both RNAP & Eta were resuspended in Buffer A (25m M Tris-HCl pH 8, 250 mM NaCl, and 10% glycerol) to a final volume of 50 µL. The protein(s) were incubated at 85°C for 10 minutes and loaded to Ni-NTA magnetic beads (Qiagen). Samples were washed three times with 250µL of Buffer A and eluted

three times with 50µL Buffer B (25 mM Tris-HCl pH 8, 100 mM NaCl, 10% glycerol, 500 mM imidiazole). Samples were resolved on a 4-20% SDS-PAGE (Biorad).

REFERENCES

1. Li W, Giles C, Li S (2014) Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res* 42(11):7069–83.
2. Hirtreiter A, et al. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* 38(12):4040–51.
3. Jiang Y, Liu M, Spencer CA, Price DH (2004) Involvement of Transcription Termination Factor 2 in Mitotic Repression of Transcription Elongation. *Mol Cell* 14(3):375–386.
4. Fish RN, Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* 1577(2):287–307.
5. Fong N, et al. (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* 60(2):256–267.
6. Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* 97(13):7090–5.
7. Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* 34(Pt 6):1062–6.
8. Windgassen T a, et al. (2014) Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res* 42(20):12707–21.
9. Hein PP, et al. (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* 21(9):794–802.
10. Yakhnin A V, Murakami KS, Babitzke P (2016) NusG is a Sequence-specific RNA Polymerase Pause Factor that Binds to the Non-template DNA Within the Paused Transcription Bubble. *J Biol Chem*. doi:10.1074/jbc.M115.704189.
11. Ha KS, Toulokhonov I, Vassylyev DG, Landick R (2010) The NusA N-terminal domain is

necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J Mol Biol* 401(5):708–725.

12. Bochkareva A, Yuzenkova Y, Tadigotla VR, Zenkin N (2012) Factor-independent transcription pausing caused by recognition of the RNA-DNA hybrid sequence. *EMBO J* 31(3):630–9.
13. James K, Gamba P, Cockell SJ, Zenkin N (2016) Misincorporation by RNA polymerase is a major source of transcription pausing *in vivo*. *Nucleic Acids Res* 45(3):gkw969.
14. Dulmage KA, Todor H, Schmid AK (2015) Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *MBio* 6(5):e00649-15.
15. Tadigotla VR, et al. (2006) Thermodynamic and kinetic modeling of transcriptional pausing. *Proc Natl Acad Sci* 103(12):4439–4444.
16. Todor H, et al. (2014) A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon *H alobacterium salinarum*. *Mol Microbiol* 93(6):n/a-n/a.
17. Darnell CL, Schmid AK (2015) Systems biology approaches to defining transcription regulatory networks in halophilic archaea. *Methods* 86:102–114.
18. Kim M, Park S, Lee S-J (2016) Global transcriptional regulator TrmB family members in prokaryotes. *J Microbiol* 54(10):639–645.
19. Lee TI, et al. (2002) Transcriptional Regulatory Networks in Saccharomyces cerevisiae. *Science (80-)* 298(5594).
20. Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 176(2):269–75.
21. Ouhammouch M, Werner F, Weinzierl RO, Geiduschek EP (2004) A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* 279(50):51719–51721.
22. Pritchett MA, Wilkinson SP, Geiduschek EP, Ouhammouch M (2009) Hybrid Ptr2-like

activators of archaeal transcription. *Mol Microbiol* 74(3):582–93.

23. Grohmann D, et al. (2011) The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* 43(2):263–274.
24. Grohmann D, Klose D, Fielden D, Werner F (2011) FRET (fluorescence resonance energy transfer) sheds light on transcription. *Biochem Soc Trans* 39(1):122–127.
25. Hausner W, Thomm M (2001) Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. *J Bacteriol* 183(10):3025–31.
26. Blombach F, et al. (2015) Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife* 4. doi:10.7554/eLife.08378.
27. Langer D, Hain J, Thuriaux P, Zillig W (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92(13):5768–72.
28. Hartzog GA, Fu J (2013) The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim Biophys Acta* 1829(1):105–15.
29. Grunberg S, Bartlett MS, Naji S, Thomm M (2007) Transcription factor E is a part of transcription elongation complexes. *J Biol Chem* 282(49):35482–35490.
30. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 30(7):1302–10.
31. Xie Y, Reeve JN (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol* 186(11):3492–8.
32. Gehring AM, Walker JE, Santangelo TJ (2016) Transcription Regulation in Archaea. *J Bacteriol* 198(14):1906–1917.
33. Walker JE, Santangelo TJ (2015) Analyses of in vivo interactions between transcription factors and the archaeal RNA polymerase. *Methods*. doi:10.1016/j.ymeth.2015.05.023.
34. Park J-S, Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription

termination. *Proc Natl Acad Sci U S A* 103(13):4870–5.

35. Cambray G, et al. (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res* 41(9):5139–48.

36. Proudfoot NJ, et al. (2016) Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 352(6291):aad9926.

37. Epshtein V, Dutta D, Wade J, Nudler E (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463(7278):245–9.

38. Santangelo TJ, Reeve JN (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol* 355(2):196–210.

39. Arimbasseri AG, Rijal K, Maraia RJ Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta* 1829(3–4):318–30.

40. Arimbasseri AG, Maraia RJ (2015) Biochemical Analysis of Transcription Termination by RNA Polymerase III from Yeast Saccharomyces cerevisiae. *Methods Mol Biol* 1276:185–98.

41. Santangelo TJ, Cubonová L, Skinner KM, Reeve JN (2009) Archaeal intrinsic transcription termination in vivo. *J Bacteriol* 191(22):7102–8.

42. Kireeva ML, Komissarova N, Waugh DS, Kashlev M (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* 275(9):6530–6.

43. Komissarova N, Becker J, Solter S, Kireeva M, Kashlev M (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell* 10(5):1151–62.

44. Roberti M, Bruni F, Polosa PL, Gadaleta MN, Cantatore P (2006) The Drosophila termination factor DmTTF regulates in vivo mitochondrial transcription. *Nucleic Acids Res* 34(7):2109–2116.

45. Nudler E, Mustaev A, Lukhtanov E, Goldfarb A (1997) The RNA-DNA hybrid maintains

the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89(1):33–41.

46. Dar D, Prasse D, Schmitz RA, Sorek R (2016) Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat Microbiol* 1(10):16143.

47. Guo J, Turek ME, Price DH (2014) Regulation of RNA Polymerase II Termination by Phosphorylation of Gdown1. *J Biol Chem* 289(18):12657–12665.

48. Brenneis M, Hering O, Lange C, Soppa J (2007) Experimental Characterization of Cis-Acting Elements Important for Translation and Transcription in Halophilic Archaea. *PLoS Genet* 3(12):e229.

49. Wolf YI, Rogozin IB, Kondrashov AS, Koonin E V (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res* 11(3):356–72.

50. Ermolaeva MD, White O, Salzberg SL (2001) Prediction of operons in microbial genomes. *Nucleic Acids Res* 29(5):1216–21.

51. Price MN, Huang KH, Alm EJ, Arkin AP (2005) A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res* 33(3):880–92.

52. Santangelo TJ, et al. (2008) Polarity in archaeal operon transcription in Thermococcus kodakaraensis. *J Bacteriol* 190(6):2244–8.

53. Stantial N, Dumpe J, Pietrosimone K, Baltazar F, Crowley DJ (2016) Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst)* 41:63–8.

54. ROBERTS JW (1969) Termination Factor for RNA Synthesis. *Nature* 224(5225):1168–1174.

55. Selby CP, Witkin EM, Sancar A (1991) Escherichia coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proc Natl Acad Sci U S A* 88(24):11574–8.

56. Gehring AM, Santangelo TJ (2015) Manipulating archaeal systems to permit analyses of

transcription elongation-termination decisions in vitro. *Methods Mol Biol* 1276:263–79.

57. Sakakibara N, Kelman LM, Kelman Z (2009) Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol* 72(2):286–96.

58. Fujiwara A, et al. (2016) Application of a Euryarchaeota-Specific Helicase from Thermococcus kodakarensis for Noise Reduction in PCR. *Appl Environ Microbiol* 82(10):3022–31.

59. Chamieh H, Ibrahim H, Kozah J (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* 576(1):214–228.

60. Atomi H, Fukui T, Kanai T, Morikawa M, Imanaka T (2004) Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea* 1(4):263–7.

61. Santangelo TJ, Cubonova L, James CL, Reeve JN (2007) TFB1 or TFB2 is sufficient for Thermococcus kodakaraensis viability and for basal transcription in vitro. *J Mol Biol* 367(2):344–357.

62. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1(8):945–51.

63. Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* 6(7):519–529.

64. Li Z, Santangelo TJ, Cuboňová L, Reeve JN, Kelman Z (2010) Affinity purification of an archaeal DNA replication protein network. *MBio* 1(5). doi:10.1128/mBio.00221-10.

65. Grohmann D, Hirtreiter A, Werner F (2009) Molecular mechanisms of archaeal RNA polymerase. *Biochem Soc Trans* 37(Pt 1):12–17.

66. Grohmann D, Hirtreiter A, Werner F (2009) RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro. *Biochem J* 421(3):339–343.

67. Meka H, Werner F, Cordell SC, Onesti S, Brick P (2005) Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res* 33(19):6435–6444.

68. Újvári A, Luse DS (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* 13(1):49–54.

69. Hirata A, et al. (2008) Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a Thermococcus kodakarensis mutant lacking subunit F is temperature-sensitive. *Mol Microbiol* 70(3):623–633.

70. Sato T, Fukui T, Atomi H, Imanaka T (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *J Bacteriol* 185(1):210–20.

71. Hileman TH, Santangelo TJ (2012) Genetics Techniques for Thermococcus kodakarensis. *Front Microbiol* 3:195.

72. Farkas JA, Picking JW, Santangelo TJ (2013) Genetic techniques for the archaea. *Annu Rev Genet* 47:539–61.

73. Spaans SK, van der Oost J, Kengen SWM (2015) The chromosome copy number of the hyperthermophilic archaeon Thermococcus kodakarensis KOD1. *Extremophiles*. doi:10.1007/s00792-015-0750-5.

74. French SL, Santangelo TJ, Beyer AL, Reeve JN (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol* 24(4):893–895.

75. Grundy FJ, Henkin TM (2006) From Ribosome to Riboswitch: Control of Gene Expression in Bacteria by RNA Structural Rearrangements. *Crit Rev Biochem Mol Biol* 41(6):329–338.

76. Adhya S, Gottesman M, De Crombrugghe B (1974) Release of polarity in Escherichia coli by gene N of phage lambda: termination and antitermination of transcription. *Proc Natl Acad Sci U S A* 71(6):2534–8.

77. Ciampi MS (2006) Rho-dependent terminators and transcription termination. *Microbiology* 152(Pt 9):2515–28.

78. Korn LJ, Yanofsky C (1976) Polarity suppressors defective in transcription termination at the attenuator of the tryptophan operon of Escherichia coli have altered rho factor. *J Mol Biol* 106(2):231–41.

79. Cohen SE, et al. (2010) Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc Natl Acad Sci U S A* 107(35):15517–15522.

80. Darrigo C, et al. (2016) The Bacterial Mfd Protein Prevents DNA Damage Induced by the Host Nitrogen Immune Response in a NER-Independent but RecBC-Dependent Pathway. *PLoS One* 11(10):e0163321.

81. D'Heygère F, Schwartz A, Coste F, Castaing B, Boudvillain M (2015) ATP-dependent motor activity of the transcription termination factor Rho from Mycobacterium tuberculosis. *Nucleic Acids Res*:gkv505-.

82. Park J-S, Marr MT, Roberts JW (2002) E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109(6):757–67.

83. Park J, Kang M, Kim M (2015) Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Res* 43(5):2625–37.

84. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–54.

CHAPTER 4

STRUCTURAL STUDIES OF ETA[4]

4.1 Introduction

Helicases and translocases are conserved in all Domains and are necessary for a diverse array of critical cellular processes (1–4). Helicases and translocases both bind DNA and/or RNA to remodel nucleic acids or nucleic acid-protein complexes. All known transcription termination factors are helicases or translocases that are able to remodel nucleic acid complexes (i.e., mediate transcription termination) through the binding and hydrolysis of ATP (5–8). The structures of bacterial and eukaryotic transcription termination factors are known (9–12); however, the structure for archaeal transcription termination factors remains elusive. Resolving structures of transcription factors opens a plethora of structure-function studies that provide invaluable mechanistic details of transcription regulation.

Recently, a novel archaeal transcription termination factor, Eta was characterized in the euryarchaeal organism *Thermococcus kodakarensis* (8). Similar to other known termination factors, Eta is annotated as a helicase and the mechanism of Eta-mediated termination has been partially resolved (1, 8). Briefly, Eta-mediates transcription termination through binding the upstream DNA of the transcription elongation complex (TEC) and in an ATP-dependent manner releases the RNA transcript and RNAP into solution. Eta is non-competitive with transcription elongation rates and thus is only able to mediate transcription termination on paused or stalled RNAPs. Cells deleted for *eta* exhibit hypersensitivity to DNA damaging agents when compared to parental strains. Mechanistically and phenotypically, Eta is very similar to the bacterial

[4] The work in this chapter is in preparation for submission.

Santangelo TJ, Murakami KS, Mohammad ZQ, Luyties O, and I conceived the content and contributed in experimental design, data interpretation, collection and analysis.

transcription-repair coupling factor Mfd. We predict that similar to Mfd, Eta plays a role in transcription-coupled DNA repair (TCR) in euryarchaea. We are continuing to elucidate the *in vivo* role of Eta; however, many mechanistic questions of Eta-mediated termination still remained unanswered.

In collaboration with the Murakami Lab at Pennsylvania State University, we resolved the crystal structure of Eta and by extension the first crystal structure of an archaeal transcription termination factor. Here we report the archaeal crystal structure of Eta at 4 Å resolution (Figure 4.1). Our results provide further evidence that Eta has several domains that are homologous to the bacterial transcription-repair coupling factor Mfd. After analysis of the Eta crystal-structure we employed several structure-function studies that provide information on the domains of Eta responsible for transcription termination.

4.2 Results & Discussion

T. kodakarensis Eta purification and crystallization.

Eta and a helicase from *Pyrococcus furiosus*, termed Hjm (Holiday junction migration), share 33% sequence identity. Eta and Hjm are not true homologs and the N-terminal region of Eta (residues 1-193) shares no sequence identity to Hjm. Nonetheless, the crystal structure for Hjm was previously solved (13) and due to the high sequence identity we did not anticipate problems with crystallizing Eta. Eta crystals were successfully formed; however, during the crystallization screening process protein aggregation was observed frequently. Attempts to resolve the crystal structure of full-length Eta were unsuccessful. After repeated unsuccessful attempts we predicted that the observed aggregation was due to the non-conserved N-terminal domain of Eta misfolding in *E. coli*.

To determine if an N-terminal deletion variant would reduce aggregation, an Eta variant was constructed wherein the first 193aa were deleted (Eta $^{\Delta 1\text{-}193}$). Attempts to resolve the crystal structure of Eta $^{\Delta 1\text{-}193}$ were successful and the first crystal structure of Eta was resolved at 4 Å

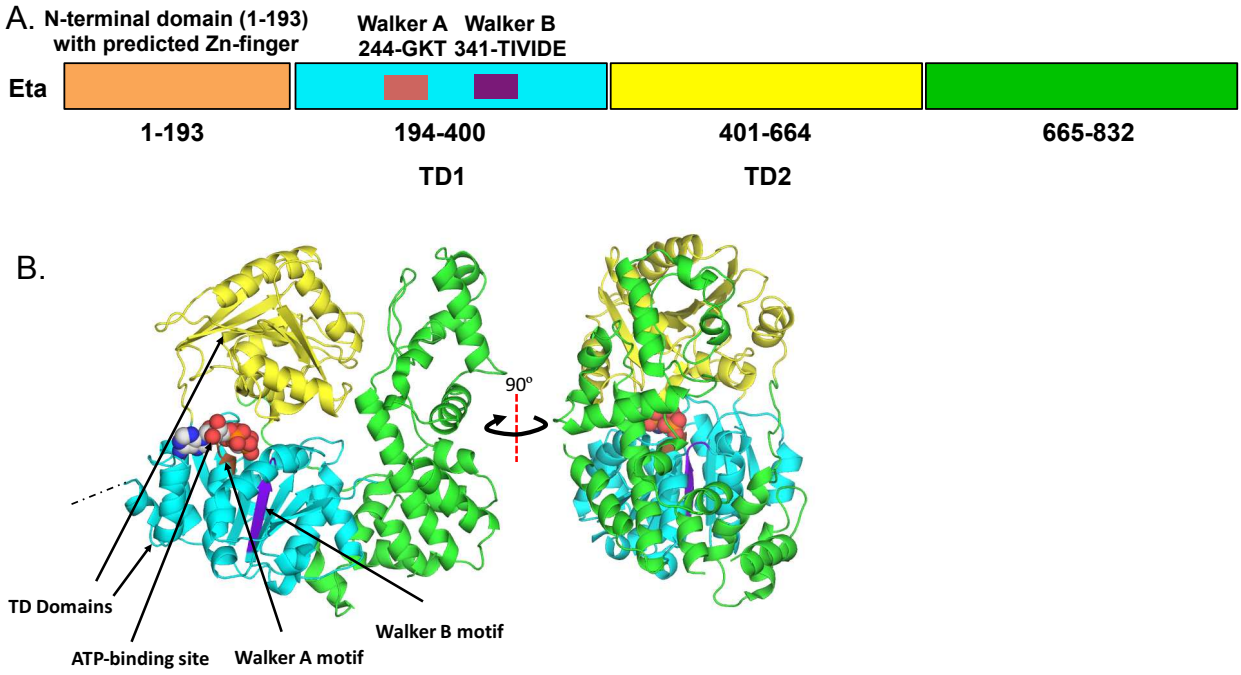


Figure 4.1. Eta crystal structure. A. Linear representation of Eta highlighting the N-terminal domain (orange), translocase domain 1 (TD1; blue), translocase domain 2 (TD2; yellow), and the C-terminal domain (green). **B.** The crystal structure of Eta at 4 Å resolution highlighting domains using the same color scheme as in A. The translocase domains (TD), Walker A and B motifs, and the ATP binding site are shown.

resolution (Figure 4.1). The crystal structure of Eta $^{\Delta 1\text{-}193}$ contains two translocase domains and a less conserved C-terminal domain (Figure 4.1). The translocase domains are predicted to be involved in binding to the DNA to mediate transcription termination and the N-terminal domain could be involved in binding RNAP to mediate transcription termination. Although, the crystal structure of the N-terminal domain was not resolved the N-terminal domain contains 'CXXC' motif characteristic of a Zn^{2+} finger motif and thus we predict the N-terminal domain to be a Zn^{2+} finger motif. The crystal structure of Eta allows for a wealth of structure-function studies to further elucidate the mechanism of Eta mediated termination.

Eta structure-function studies.

Previously, we showed that an Eta variant deleted for residues 1-193 (Eta $^{\Delta 1\text{-}193}$/EtaM2) retained full termination activity *in vitro* (8). To determine if residues 1-193 have a role *in vivo* a *T. kodakarensis* strain was constructed wherein the first 579 nt of *eta* were deleted. Deletion of the complete locus or deletion of just the N-terminal domain of Eta resulted in slowed and limited growth when compared to the parental strain TS559 (Figure 4.2A). Although the N-terminal region is not required for Eta-mediated transcription termination, the N-terminal region is likely involved in a critical role *in vivo* and may potentially be involved in recruiting DNA repair factors.

From the Eta crystal structure, the C-terminal domain (residues 573-832) do not fold into a highly conserved domain. To determine if the C-terminal domain is necessary for Eta-mediated termination two Eta variants (Eta^{M3} and Eta^{M4}) (Figure 4.2B; upper panel) were purified and tested for termination activity. Both Eta variants showed no termination activity (Figure 4.2B; lower panel) and thus we conclude that the C-terminal domain of Eta is essential for Eta-mediated termination. The C-terminal domain may be necessary for binding to RNAP and/or the DNA and follow up studies are needed to describe the role of the C-terminus of Eta.

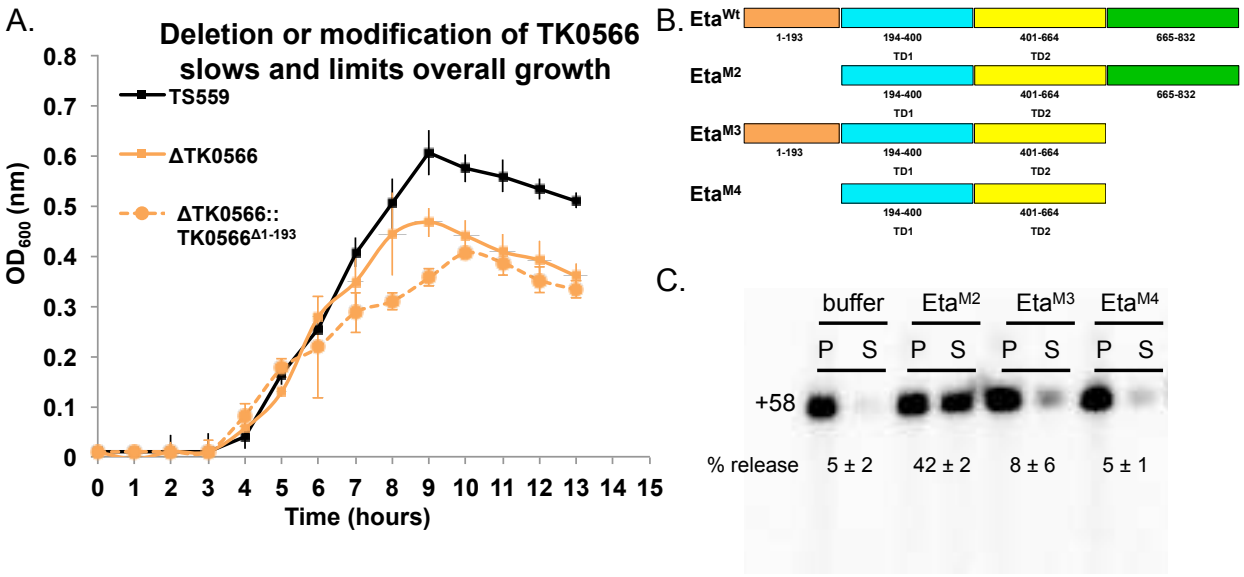


Figure 4.2.Eta structure-function studies. A. Deletion of full-length Eta or residues 1-193 of Eta hinders growth rate and final cell densities. The curves and errors shown represent means and standard averages of triplicate technical repeats of triplicate biological samples. **B.** Linear representation of Eta and Eta variants shown in C. **C.** Eta^{M3} and Eta^{M4} do not retain termination activity *in vitro*.

Eta structurally resembles Mfd.

Eta shares many similarities to the bacterial transcription-repair coupling factor Mfd (Figure 4.3). For example, both *eta* and *mfd* are dispensable to cells; however, the deletion of both genes renders cells hypersentivitve to DNA damaging agents (8, 14). Structurally, both Eta and Mfd are annotated as superfamily II helicases. They both contain translocase domains and an ATP binding site with the highly conserved Walker A & B motifs that are responsible for the binding and hydrolysis of ATP. Mechanistically, both Eta and Mfd recognize paused/arrested RNAP and bind to the upstream edge of DNA to mediate transcription termination (15). The RNAP interacting domain (RID) has been elucidated for Mfd and is shown in orange in the model (Figure 4.3); however, the RID of Eta remains to be identified.

4.3 Future Directions

Our collaborators are currently optimizing crystallization parameters of Eta $^{\Delta 1\text{-}193}$ in attempts to resolve the Eta crystal structure at 1-2 Å resolution. Once this data has been acquired we plan to publish the Eta crystal structure as well as additional structure-function studies.

To further investigate the function of Eta numerous variants have been purified. Currently, we are targeting two portions of a conserved winged-helix domain in the translocase domain 2 portion of Eta. An archaeal helicase Hel308, with some structural similarities to Eta, is predicted to be involved in DNA replication and contains a conserved winged-helix domain. Studies with the helicase Hel308 determined that the winged-helix domains were critical for the binding to dsDNA (16). Based of this data, we predicted that the winged-helix domain of Eta may also be essential for binding to dsDNA. For one Eta variant we deleted the entire winged-helix domain (Eta^{M5}; deleted for residues 531-651). For the other Eta variant we only partially deleted the winged-helix domain (residues Eta^{M6}; deleted for residues 630-642) (Figure 4.4).

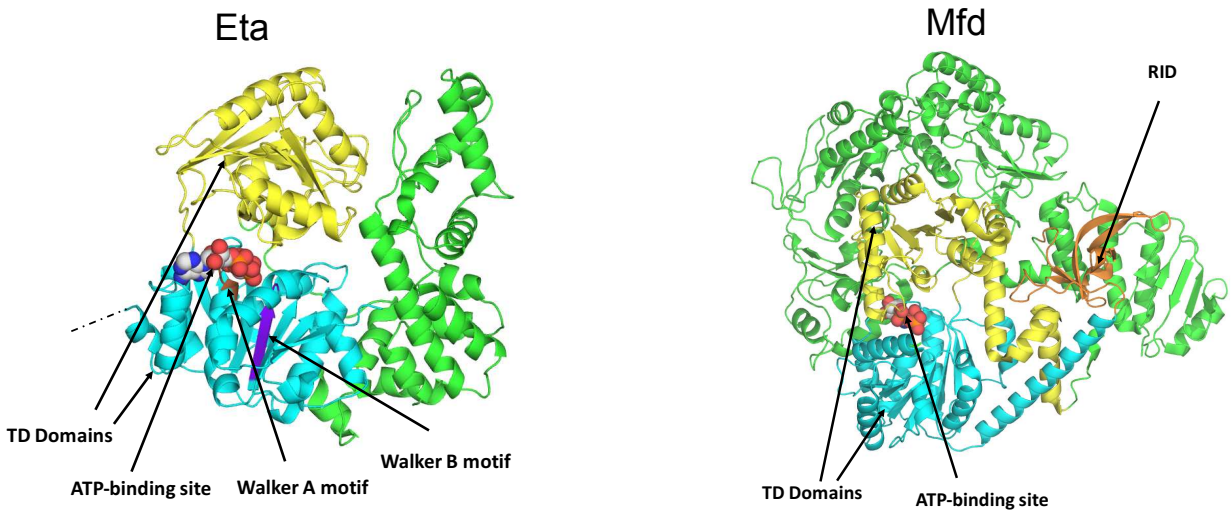


Figure 4.3. Eta resembles Mfd. Structural similarities between Eta and the bacterial transcription coupling repair factor Mfd. The Mfd RNAP interacting domain (RID) is shown in orange.

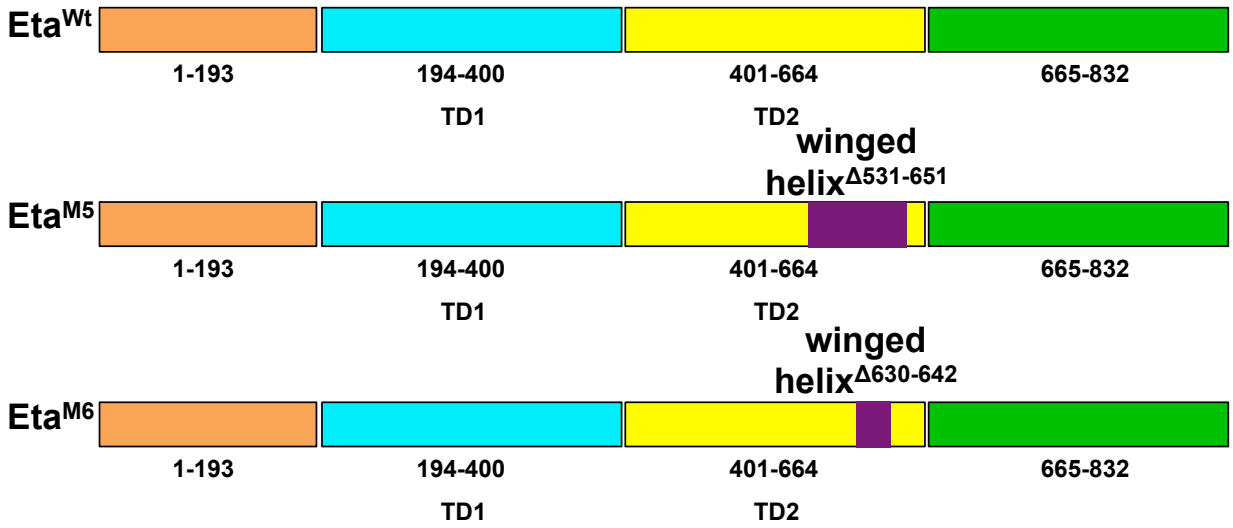


Figure 4.4 Eta variants. Linear representation of Eta and Eta variants. The portions of the winged-helix domain deleted for each variant (Eta^{M5} and Eta^{M6}) are highlighted in purple.

The Eta variants have been purified and in the near future will be analyzed for termination activity and tested for the ability to bind to dsDNA in the absence of the TEC. In addition to testing the activity of Eta variants *in vitro* we plan to construct appropriate *T. kodakarensis* strains encoding for Eta variants.

A structure of the TEC in complex with a transcription termination factor has not been resolved in any Domain. In collaboration with Katsuhiko Murakami we plan to resolve the first Cyro-EM structure of the TEC in complex with a termination factor (Eta). Our lab has a procedure in hand to acquire large amounts of Eta in complex with the TEC that can be used for Cyro-EM studies. Solving the structure of Eta in complex with the TEC would provide great insight into Eta domains binding to RNAP and/or DNA and will further our understanding of the mechanism of Eta-mediated termination. Continued structure-function studies as well as Cyro-EM studies to resolve the Eta-TEC structure will undoubtedly have a broad impact for both the transcription community as well as the DNA repair community.

4.4 Methods

Purification and crystallization of Eta.

Eta and Eta variants (Eta^{Wt}, $Eta^{\Delta 1\text{-}193}$, $Eta^{\Delta 665\text{-}832}$, $Eta^{\Delta 1\text{-}193\&\Delta 665\text{-}832}$) were purified as previously described in (8).

A selenomethionine (SeMet) substituted $Eta^{\Delta 1\text{-}193}$ complex was prepared by suppression of methionine biosynthesis during culture growth followed by an identical purification scheme as referenced previously. Both native Eta^{Wt} and SeMet-labeled $Eta^{\Delta 1\text{-}193}$ were concentrated to 10mg/mL with buffer (0.1M Tris, pH 8, 4M NaCl and 4% ethylene glycol) for crystallization. Crystals were acquired using the “hanging drop” method at room temperature (22ºC). The crystals were flash frozen by liquid nitrogen. To acquire the phase information native crystals from Eta^{Wt} were soaked in 2mM TaBr harvested, and then flash frozen in liquid nitrogen.

Structure determination of Eta.

The X-ray crystallographic datasets were obtained from two different crystal sets a) native crystals soaked in Tantalum bromide (TaBr) cluster, b) Seleno-methionine (SeMet) labeled Eta $^{\Delta 1\text{-}193}$. TaBr cluster is utilized for the preparation of heavy-atom derivatives for structure determination of biological macromolecules by X-ray analysis. The cluster is an electron-rich compound that brings significant changes in crystal diffraction required for phase calculation in single and multiple isomorphous replacement (SIR and MIR) experiments and in anomalous dispersion (SAD and MAD) experiments. Both SAD/MAD and SIR/MIR approaches using SeMet and TaBr.

The datasets were collected at the Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY), and the data was processed by HKL2000. The anomalous peaks from SeMet Eta $^{\Delta 1\text{-}193}$ and TaBr Eta^{Wt} were used for molecular replacement followed by rigid body and positional refinements with non-crystallographic symmetry by using the program Phenix. The resulting maps contained additional electron densities, which allowed the structure to be built in using Coot.

***In vitro* Transcription.**

In vitro transcription termination assays were carried out as previously described in (8).

REFERENCES

1. Chamieh H, Ibrahim H, Kozah J (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* 576(1):214–228.
2. Patel SS, Donmez I (2006) Mechanisms of Helicases. *J Biol Chem* 281(27):18265–18268.
3. Lohman TM, Tomko EJ, Wu CG (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* 9(5):391–401.
4. Jarmoskaite I, Russell R (2014) RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* 83:697–725.
5. Han Z, Porrua O (2017) Helicases as transcription termination factors: different solutions for a common problem. *Transcription*:00–00.
6. Smith AJ, Pernstich C, Savery NJ (2012) Multipartite control of the DNA translocase, Mfd. *Nucleic Acids Res* 40(20):10408–16.
7. Epshtein V, Dutta D, Wade J, Nudler E (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463(7278):245–9.
8. Walker JE, Luyties O, Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci*:201704028.
9. Deaconescu AM, et al. (2006) Structural Basis for Bacterial Transcription-Coupled DNA Repair. *Cell* 124(3):507–520.
10. Westblade LF, et al. (2010) Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res* 38(22):8357–69.
11. Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114(1):135–46.
12. Richter H, Katic I, Gut H, Großhans H (2016) Structural basis and function of XRN2 binding by XTB domains. *Nat Struct Mol Biol* 23(2):164–171.

13. Fujikane R, Shinagawa H, Ishino Y (2006) The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork. *Genes Cells* 11(2):99–110.

14. Cohen SE, et al. (2010) Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc Natl Acad Sci U S A* 107(35):15517–22.

15. Roberts J, Park J-S (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol* 7(2):120–125.

16. Northall SJ, et al. (2017) DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst)* 57:125–132.

CHAPTER 5

TRANSCRIPTION AND TRANSLATION ARE COUPLED IN ARCHAEA[5]

5.1 Introduction

Bacteria and *Archaea* lack a nuclear membrane, thus permitting the transcription and translation apparatuses to have access to each other and functionally interact (1, 2). Translation initiates before transcription is complete, and ribosome binding to nascent transcripts is normally essential for completion of transcription (3–5). Disrupting the normally-tight association of ribosomes and RNAP, and thus exposure of nascent transcripts to solution, results in polar repression of downstream transcription (3, 6). Coupling of the translation and transcription apparatuses is an essential process and increases genome stability and regulates genome-wide gene expression (3, 7–11). Coupling limits RNA polymerase (RNAP) backtracking and prevents premature transcription termination (3, 7, 11–13). When transcription initiates aberrantly, or when translation lags or is halted due to a premature nonsense codon, the lack of coupling signals factors to terminate transcription (Figure 5.1). In addition the rate of transcription elongation is regulated by the rate of translation elongation (2, 5, 7, 12).

Decades of research in the bacterial community has shown that the RNAP-bound, tightly associated elongation factor NusG (homologous to Spt5 in *Archaea*) makes direct contact with the ribosomal protein S10 and that this molecular interaction is the basis of coupling the transcription and translation apparatuses (14–16). In the absence of an S10-NusG interaction, RNAP and the trailing ribosome can become disengaged, and when sufficient (~70nt) RNA is exposed to solution then the transcription termination factor Rho will access the exposed RNA and mediate transcription termination, resulting in polarity (17–19).

[5] Santangelo TJ, and I conceived the content and contributed in experimental design. Lynch ER, Selena J and I contributed in experimental data interpretation, collection, and analysis.

NusG is composed of two domains: a NusG N-terminal domain (NGN) and a Kyprides-Ouzounis-Woese (KOW) C-terminal domain (Figure 5.1) (20, 21). The NGN domain binds RNAP while the KOW domain binds the S10 ribosomal subunit. In addition to the role in coupling, NusG generally stimulates transcription elongation rates, increases RNAP processivity, and stimulates Rho-dependent transcription termination (9, 22–25). A recent cryo-EM structure provides an alternative, NusG-S10 independent mechanism of transcription and translation coupling, but this "expressome" structure has not yet been biologically challenged to determine whether it represents a natural alternative to the NusG-S10 mediated-coupling (26).

Due to coupling of transcription and translation and the conservation of the NusG/Spt5 in *Bacteria* and *Archaea*, we hypothesized that archaeal organisms use the same interaction (Spt5 & S10) to couple the transcription and translation apparatuses. We also predicted that archaeal Spt5 may interact with the transcription termination factor(s) responsible for polarity. Bioinformatic analysis shows no obvious homolog of a Rho like factor encoded in archaeal genomes but we predicted there are Rho analogues encoded in archaeal genomes. Our studies sought to identify and further characterize the binding partners of Spt5 in the hopes of further characterizing the coupling of transcription and translation, as well as to identify any archaeal-encoded polarity factor(s).

As an alternative approach, we also used bioinformatics to search for other potential transcription termination factors. There are no eukaryotic or bacterial transcription termination factors encoded in archaeal genomes, but there is a homologue of the eukaryotic Cleavage and Polyadenylation Specificity Factor (CPSF) encoded in archaeal genomes. In eukaryotes, CPSF is part of a complex that is responsible for RNA cleavage at the poly-A site encoded in mRNA (27). Cleavage of the mRNA permits the transcription termination factor Xrn2 to mediate transcription termination of RNA polymerase II (Pol II) (28–30). We predicted that CPSF may have a role in transcription termination in *Archaea* and further that CPSF might be involved in

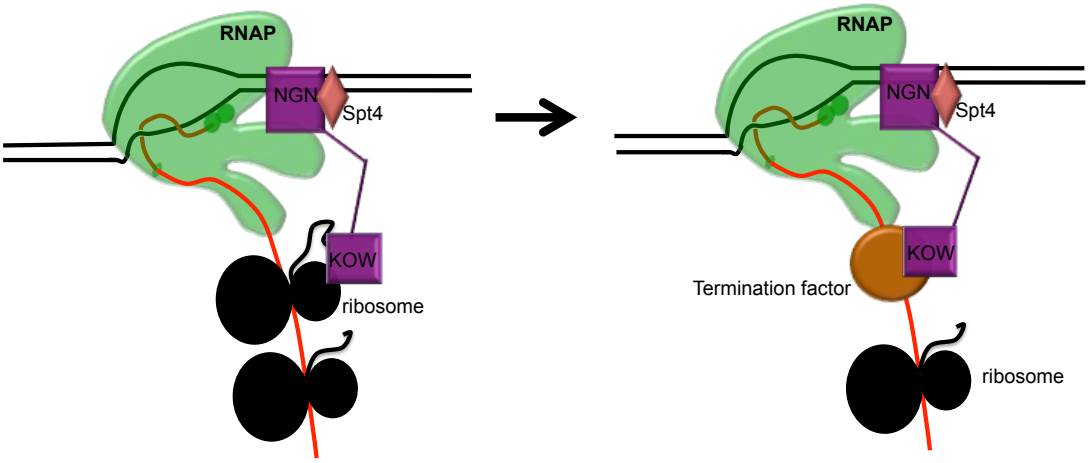


Figure 5.1. Coupling and uncoupling of transcription and translation. Transcription and translation are coupled in *Bacteria* and *Archaea* and it is hypothesized that Spt5 (purple) is the coupling factor that binds to both RNAP (green) and the leading ribosome (black). When transcription and translation become uncoupled a termination factor, termed Rho in bacteria, binds to the Spt5-KOW domain to mediate transcription termination.

global 3' end formation of RNA transcripts. CPSF contains no conserved domains predicted to bind or hydrolyze ATP, and given that all known transcription termination factors require ATP hydrolysis to stimulate termination (31–33), it is unlikely that CPSF is the sole factor involved in transcription termination. Instead, our studies of CPSF were aimed at identifying binding partners in the hopes of isolating additional factors responsible for transcription termination. Here we begin to characterize the *in vitro* activity of CPSF from *T. kodakarensis*.

5.2 Results and Discussion

Isolation of His_6-Spt5 containing complexes.

To determine interaction partners of Spt5 – and hopefully identify potential coupling and polarity factors – we used an *in vivo* pull-down to isolate Spt5 directly from whole cells under mild conditions. Established genetic techniques (34, 35) were employed to generate a strain of *T. kodakarensis* wherein sequences encoding affinity- and epitope-tags were appended to the C-terminus of TK1419 (encoding Spt5) (36). The resultant strain produced full-length, C-terminally tagged Spt5 from the natural TK1419 locus under normal regulatory control. Cell lysis, followed by gentle nickel-affinity purifications of complexes containing His6-tagged Spt5 were performed, and the identity of co-purifying partners revealed by mass spectrometry (Table 5.1). Several subunits of RNA polymerase and the archaea-eukarya conserved elongation factor Spt4 were identified, supporting the well-established Spt4-Spt5-RNAP interactions *in vivo*. Interestingly, the S10 ribosomal subunit (TK0307) was identified thus adumbrating a role of Spt5-S10 interactions in the coupling of transcription and translation. In attempts to identify an uncoupling factor (i.e., termination factor) we searched for helicase and ATPase enzymes that were identified. An ATPase (TK2042) was identified and is highly conserved in *euryarchaea*. Although a pull-down is suggestive of Spt5 interacting partners, it is not conclusive evidence of Spt5 directly interacting with the identified proteins. Follow up studies were done specifically with S10 to determine if Spt5 and S10 are directly interacting.

Table 5.1. Proteins copurifying with His_6-Eta from cellular lysates.

Gene	Annotation	MASCOT
TK1419	Spt5	40,970
TK1736	DNA directed RNA polymerase subunit F	31,862
TK1083	DNA directed RNA polymerase subunit beta	4,321
TK1082	DNA directed RNA polymerase subunit alpha	1,875
TK1081	DNA directed RNA polymerase subunit A''	890
TK2042	ATPase	772
TK1698	Spt4	696
TK1769	Transcriptional Regulator	680
TK1503	DNA directed RNA polymerase subunit D	678
TK0137	Nucleic-acid binding protein	533
TK0883	hypothetical protein	310
TK0307	30S ribosomal protein S10	280

Characterization of the Spt5-S10 interaction.

The bacterial NusG-NGN domain interacts with RNAP and the NusG-KOW domain interacts with the S10 ribosomal subunit. We predicted that *Archaea* coupling also might be reliant on Spt5 interacting with RNAP and the S10 ribosomal subunit. To determine if Spt5 interacts with the S10 ribosomal subunit we employed a binding assay using purified components. To ensure our assay worked properly we also investigated the well-established Spt4-Spt5 interaction.

Transcription elongation factors Spt5, Spt5 variants and Spt4-his_6 and the ribosomal subunit S10-his_6 were expressed recombinantly and purified for use during a nickel-resin based *in vitro* binding assay. The nickel-resin binding assay is reliant on only one of the proteins of interest containing an affinity his_6 tag that binds to the nickel slurry and the other protein of interest containing no affinity tag. Spt4-his_6 was used as a positive control as it is well established in the literature that Spt5 and Spt4 interact (23, 37–39). Our results corroborate this data as Spt4-his_6 and Spt5 co-eluted (Figure 5.2A).

To determine if Spt5 and S10-his_6 interact, we used the same assay as described above. We concluded that Spt5 and S10-his_6 did not interact as only S10-his_6 was present in the elutions (data not shown). Previously, it was shown that Spt5 can auto-inhibit itself through the NGN domain interacting with the KOW domain (40, 41). To ensure that the NGN domain was not auto-inhibiting the KOW-domain from interacting with S10 an Spt5 variant was constructed (Spt5-KOW) wherein the NGN domain was removed. Disappointingly, Spt5-KOW and S10-his_6 still did not interact and only S10-his_6 was present in the elutions (Figure 5.2B). We conclude that Spt5-KOW and S10 most likely do not interact *in vivo* and perhaps Spt5-KOW interacts with another subunit of the ribosome to couple transcription and translation. Based off these results, archaeal coupling does not follow the classic model of bacterial coupling. Very recent cyro-EM evidence from the bacterial community suggests RNAP and the ribosome

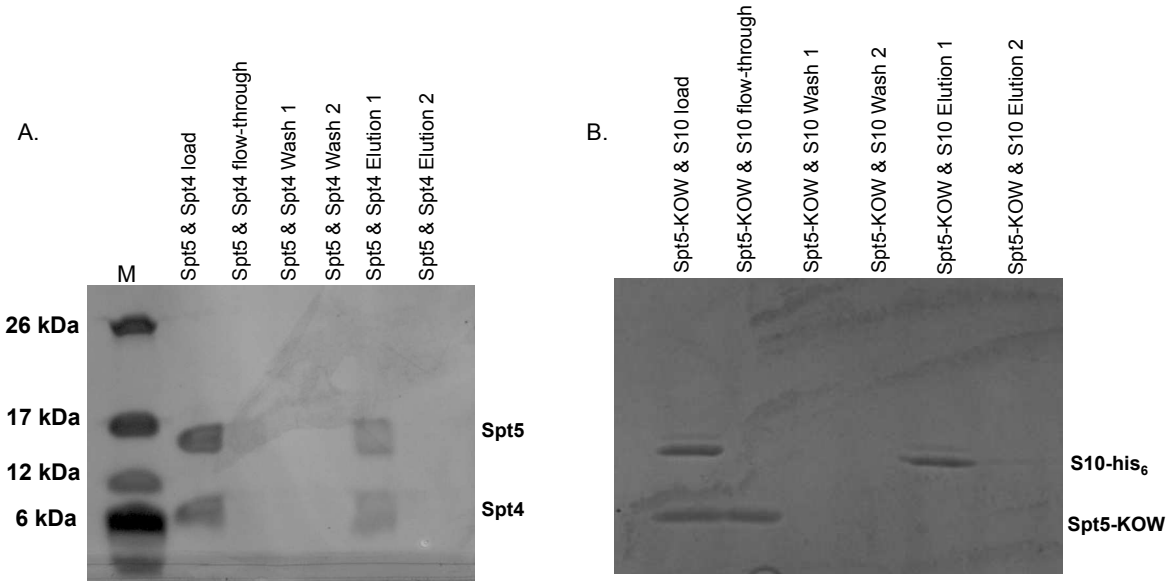


Figure 5.2. Spt5 interacts with Spt4 but not S10. A. A pull-down experiment was employed with Spt5 (17kDa) and Spt4-his_6 (7kDa). The portion of the sample that did not stick to the resin is indicated as 'flow-through' and the portion of the sample that eluted when a high imidazole buffer was added is indicated as 'elution'. Samples were resolved on an SDS-PAGE with a marker shown in kDa. **B.** A pull-down experiment was employed with Spt5-KOW (7kDa) and S10-his_6 (12kDa). The portion of the sample that did not stick to the resin is indicated as 'flow-through' and the portion of the sample that eluted when a high imidazole buffer was added is indicated as 'elution'. Samples were resolved on an SDS-PAGE.

directly interact to facilitate coupling and Spt5 may not be required for coupling (26). This model has not been validated *in vivo* and thus the classic model of Spt5-S10 coupling is still largely accepted in the field. The archaeal RNAP and ribosome may directly interact to facilitate coupling or Spt5 may interact with a different subunit of the ribosome to facilitate coupling. Follow up studies now need to be employed to determine the mechanism of transcription and translation coupling in *Archaea*.

Characterization of potential transcription termination factors.

An ATPase (TK2042) was identified during the Spt5 *in vivo* pull-down and we are now targeting this ATPase to determine if it has a role in transcription termination. We have recombinantly purified the ATPase and future studies will focus on characterization of the ATPase in the well-defined *in vitro* transcription termination assay (33). Current efforts are focused on determining if the ATPase (encoded by TK2042) is able to release the nascent transcript and RNAP into solution thus defining the ATPase as a *bone fide* transcription termination factor. Ongoing efforts are attempting to knock-out TK2042 from the *T. kodakarensis* genome to determine if the ATPase encoded has a critical role.

Attempts to delete the TK1428 locus (encoding for CPSF) from the *T. kodakarensis* genome have not been successful, hinting at an essential role for CPSF. To determine if purified CPSF affected transcription elongation or stimulated transcription termination, we recombinantly expressed and purified CPSF. Preliminary *in vitro* transcription assays with CPSF demonstrate a cleavage activity associated with CPSF addition, but no obvious termination activity. A DNA template was used such that a long RNA transcript would be produced to satisfy any RNA length or sequence requirements CPSF may have (Figure 5.3A). A “failure-to-chase” assay was used in preliminary studies wherein $TECs_{+376}$ were formed, washed, and then incubated +/- an energy source (dATP) and +/- CPSF. After incubation, addition of all four rNTPs permits intact $TECs_{+376}$ to restart transcription and elongate to the end of the template, resulting in full-length

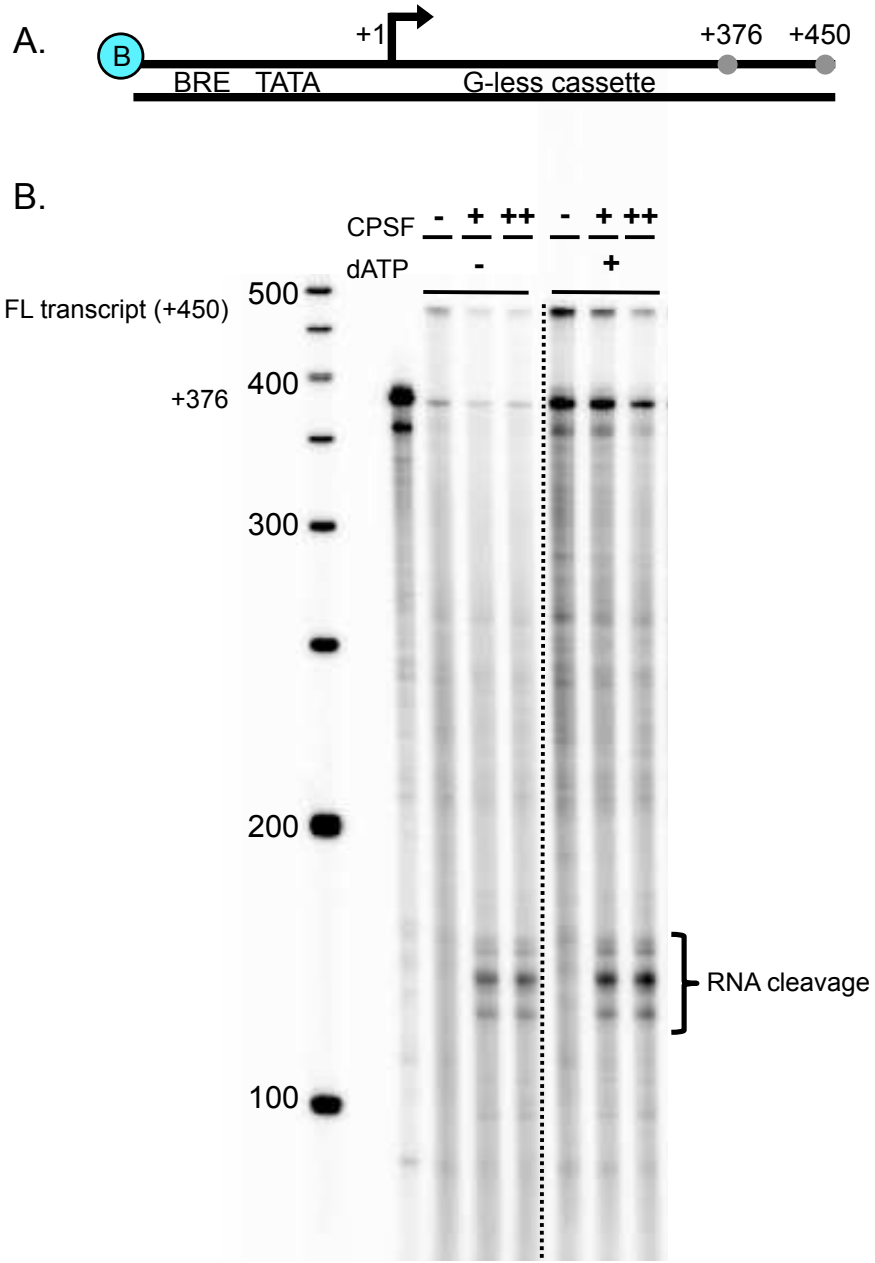


Figure 5.3. CPSF has RNA cleavage activity. A. DNA templates contain a biotin moiety (blue B), a strong promoter, P_{hmtB}, a 376bp G-less cassette, and permit elongation to produce a full-length transcript of +450. **B.** Stalled TECs at the end of a G-less cassette (TEC_{+376}) were incubated in the presence or absence of an energy source (dATP) and +/- CPSF. CPSF RNA cleavage activity is observed ~150nt in both the presence and absence of dATP.

(+450 nt) transcripts. We predicted that in the absence of CPSF all $TECs_{+376}$ would resume elongation and transcribe a full-length transcript, and if CPSF had termination activity then only some of the $TECs_{+376}$ would be able to resume elongation. Our results show that ~50% of complexes were unable to transcribe full-length transcripts in the absence of CPSF. We were surprised by this result as previous assays using a shorter DNA template showed that nearly 100% of complexes are able to form full-length transcripts under the same conditions. Nonetheless, when CPSF is present in reactions RNA cleavage is observed in both the absence and presence of an energy source dATP at around ~150nt. (Figure 5.3B). Despite RNA cleavage activity, we are unable to provide evidence that isolated CPSF has a role in directly mediating transcription termination.

Preliminary results from an *in vitro* transcription termination assay where RNA transcript release into solution is observed reveal that CPSF most likely does not have a direct role in transcription termination (data not shown). We are currently optimizing assay conditions to determine if CPSF is responsible for transcription termination or if the role of CPSF is exclusively to cleave the RNA. Future studies will further characterize the activities of CPSF as well as determine the *in vivo* binding partners of CPSF.

5.3 Methods

Isolation of His_6-Spt5 complexes and MS identification of binding partners.

T. kodakarensis cells were harvested by centrifugation from 5-L cultures grown to mid-exponential phase at 85 °C in ASW-YT medium supplemented with 5 g of sodium pyruvate/L. The cells were resuspended in 30 mL of Buffer A (25 mM Tris·HCl pH 8, 500 mM NaCl, 10 mM imidazole, and 10% glycerol) and lysed by sonication. Clarified lysate was loaded onto a 1-mL HiTRAP chelating column (GE Healthcare) preequilibrated with $NiSO_4$. The column was washed with Buffer A, and proteins were eluted using a linear imidazole gradient from Buffer A to 60% Buffer B (25 mM Tris·HCl pH 8, 100 mM NaCl, 150 mM imidazole, and 10% glycerol). Fractions

that contained the tagged protein were identified by Western blot analysis, pooled, and dialyzed against Buffer C (10 mM Tris·HCl pH 8). The pooled samples were lyophilized and resuspended in 10 mM Tris·HCl pH 8. The proteins were identified by 1D SDS/PAGE fractionation followed by LC/MSMS at the Ohio State University MS facility. A MASCOT score of > 100 was considered meaningful. To obtain such a score, a minimum of two unique peptide fragments usually had to be identified from the same protein. Protein isolation and MS analyses of lysates from *T. kodakarensis* TS559 (parental strain) were also undertaken. From these controls, several *T. kodakarensis* proteins that bound and eluted from the Ni2+-charged matrix in the absence of a His6-tagged protein were identified. The proteins identified in the Spt5 sample that had a MASCOT score >100 and were not also present in the control samples are listed in Table 5.1.

Protein purifications.

The pQE-80L plasmids encoding Spt5 (pTS317), Spt5-KOW (pJW19), Spt4 –his_6 (pTS542), CPSF-his_6 (pAKL1), ATPase-his_6 (TK2042; pOL1) and S10-his_6 (pJW12) were transformed into *E. coli* Rosetta 2 cells. Cells were grown at 37˚C until an OD_{600} of 0.5-0.7 was reached. The cells were induced with 0.4mM isopropyl-β-D-thiogalactopyranoside and incubated at 37˚C for three hours. Cell biomass was collected after induction and lysed using sonication. Cell lysate was heat treated at 75˚C to denature most of the *E. coli* proteins while the thermostable recombinant proteins remained folded.

Heat-treated cell-lysate for Spt4, CPSF, ATPase (TK2042) and S10 was loaded onto a 1-mL HiTRAP chelating column (GE Healthcare) preequilibrated with $NiSO_4$. The column was washed with 20 column volumes (CV) in lysis buffer (25mM Tris-HCl pH 8.0, 10% glycerol (v/v), and 1M NaCl). Proteins were eluted over a linear gradient in Buffer B (25mM Tris-HCl pH 8.0, 10% glycerol (v/v), 0.1M NaCl, and 250mM imidazole). Purified proteins were dialyzed into storage buffer (25mM Tris-HCl pH 8.0, 1mM EDTA, 200mM KCl, 50% glycerol, and 1mM β-Me).

Heat treated cell-lysate for Spt5 and Spt5-KOW was loaded to an anion exchange column. The column was washed with 20CV lysis buffer (25 mM Tris-HCL pH 8.0, 10 mM MgCl2, and 0.1 mM EDTA). Proteins were eluted via a linear gradient with elution buffer (1 M KCl, 25 mM Tris-HCl pH 8.0, 10 mM MgCl2, and 0.1 mM EDTA. Purified proteins were dialyzed into storage buffer (25mM Tris-HCl pH 8.0, 1mM EDTA, 200mM KCl, 50% glycerol, and 1mM β-Me) and stored at -20˚C.

***In vitro* binding studies.**

8µg of each purified protein (Spt4-his6, Spt5, Spt5-KOW, S10 ribosomal subunit) were resuspendend in Buffer A (25 mM Tris HCl pH 8, 250 mM NaCl, and 10% glycerol) to a final volume of 50 µL. The proteins were incubated at 85°C for 10 minutes and loaded to Ni-NTA magnetic beads (Qiagen). Samples were washed two times with 250 µL Buffer A and then eluted two times with Buffer B (25 mM Tris HCl pH 8, 100 mM NaCl, 10% glycerol, and 500 mM imidazole). Samples were resolved on a 4-20% SDS-PAGE (Bio-Rad).

***In vitro* transcription.**

In vitro transcription assays were carried out as previously described (33).

REFERENCES

1. French SL, Santangelo TJ, Beyer AL, Reeve JN (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol* 24(4):893–895.
2. Mäkelä J, Lloyd-Price J, Yli-Harja O, Ribeiro AS (2011) Stochastic sequence-level model of coupled transcription and translation in prokaryotes. *BMC Bioinformatics* 12(1):121.
3. Santangelo TJ, et al. (2008) Polarity in archaeal operon transcription in Thermococcus kodakaraensis. *J Bacteriol* 190(6):2244–8.
4. Korn LJ, Yanofsky C (1976) Polarity suppressors defective in transcription termination at the attenuator of the tryptophan operon of Escherichia coli have altered rho factor. *J Mol Biol* 106(2):231–41.
5. Bastet L, et al. (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Res* 45(12):7474–7486.
6. Adhya S, Gottesman M, De Crombrugghe B (1974) Release of polarity in Escherichia coli by gene N of phage lambda: termination and antitermination of transcription. *Proc Natl Acad Sci U S A* 71(6):2534–8.
7. Li R, Zhang Q, Li J, Shi H (2016) Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. *Nucleic Acids Res* 44(6):2554–63.
8. Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328(5977):504–8.
9. Cardinale CJ, et al. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320(5878):935–8.
10. Grundy FJ, Henkin TM (2006) From Ribosome to Riboswitch: Control of Gene Expression in Bacteria by RNA Structural Rearrangements. *Crit Rev Biochem Mol Biol* 41(6):329–338.

11. Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E (2011) Linking RNA Polymerase Backtracking to Genome Instability in E. coli. *Cell* 146:533–543.
12. Nudler E (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149(7):1438–45.
13. Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114(1):135–46.
14. Gottesman ME, Rösch P (2010) A NusE : NusG Complex Links Transcription and Translation. 283(APRIL):501–505.
15. Drögemüller J, et al. (2015) Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci Rep* 5:16428.
16. Burmann BM, Scheckenhofer U, Schweimer K, Rösch P (2011) Domain interactions of the transcription-translation coupling factor Escherichia coli NusG are intermolecular and transient. *Biochem J* 435(3):783–9.
17. Epshtein V, Dutta D, Wade J, Nudler E (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463(7278):245–9.
18. Kriner MA, Sevostyanova A, Groisman EA (2016) Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho. *Trends Biochem Sci*. doi:10.1016/j.tibs.2016.05.012.
19. Leela JK, Syeda AH, Anupama K, Gowrishankar J (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proc Natl Acad Sci U S A* 110(1):258–63.
20. Zhou H, Liu Q, Gao Y, Teng M, Niu L (2009) Crystal structure of NusG N-terminal (NGN) domain from Methanocaldococcus jannaschii and its interaction with rpoE''. *Proteins* 76(4):787–93.
21. Liu B, Steitz TA (2017) Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res* 45(2):968–974.

22. Tomar SK, Artsimovitch I (2013) NusG-Spt5 proteins-Universal tools for transcription modification and communication. *Chem Rev* 113(11):8604–19.

23. Hirtreiter A, et al. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* 38(12):4040–51.

24. Hartzog G a, Kaplan CD (2011) Competing for the clamp: promoting RNA polymerase processivity and managing the transition from initiation to elongation. *Mol Cell* 43(2):161–3.

25. Chalissery J, et al. (2011) Interaction Surface of the Transcription Terminator Rho Required to Form a Complex with the C-Terminal Domain of the Antiterminator NusG. *J Mol Biol* 405(1):49–64.

26. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P (2017) Architecture of a transcribing-translating expressome. *Science (80-)* 356(6334):194–197.

27. Nag A, Narsinh K, Martinson HG (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat Struct Mol Biol* 14(7):662–669.

28. Richter H, Katic I, Gut H, Großhans H (2016) Structural basis and function of XRN2 binding by XTB domains. *Nat Struct Mol Biol* 23(2):164–171.

29. Fong N, et al. (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* 60(2):256–267.

30. Chalamcharla VR, Folco HD, Dhakshnamoorthy J, Grewal SIS (2015) Conserved factor Dhp1/Rat1/Xrn2 triggers premature transcription termination and nucleates heterochromatin to promote gene silencing. *Proc Natl Acad Sci* 112(51):201522127.

31. D'Heygère F, Schwartz A, Coste F, Castaing B, Boudvillain M (2015) ATP-dependent motor activity of the transcription termination factor Rho from Mycobacterium tuberculosis. *Nucleic Acids Res*:gkv505-.

32. Han Z, Porrua O (2017) Helicases as transcription termination factors: different solutions for a common problem. *Transcription*:00–00.

33. Walker JE, Luyties O, Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci*:201704028.

34. Farkas JA, Picking JW, Santangelo TJ (2013) Genetic techniques for the archaea. *Annu Rev Genet* 47:539–61.

35. Hileman TH, Santangelo TJ (2012) Genetics Techniques for Thermococcus kodakarensis. *Front Microbiol* 3:195.

36. Li Z, Santangelo TJ, Cuboňová L, Reeve JN, Kelman Z (2010) Affinity purification of an archaeal DNA replication protein network. *MBio* 1(5). doi:10.1128/mBio.00221-10.

37. Blythe AJ, et al. (2016) The yeast transcription elongation factor Spt4/5 is a sequence-specific RNA binding protein. *Protein Sci* 25(9):1710–21.

38. Hartzog GA, Fu J (2013) The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim Biophys Acta* 1829(1):105–15.

39. Guo M, et al. (2008) Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure* 16(11):1649–58.

40. Drögemüller J, et al. (2016) OUP accepted manuscript. *Nucleic Acids Res*. doi:10.1093/nar/gkw1111.

41. An Autoinhibited State in the Structure of Thermotoga maritima NusG (2013) *Structure* 21(3):365–375.