

DISSERTATION

POST-INITIATION ACTIVITIES OF THE ARCHAEL RNA POLYMERASE IN A  
CHROMATIN LANDSCAPE

Submitted by

Travis James Sanders

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 2020

Doctoral Committee:

Advisor: Thomas Santangelo

Jeffrey C. Hansen

Olve B. Peersen

Asa Ben-Hur

Copyright by Travis James Sanders 2020

All Rights Reserved

## ABSTRACT

### POST-INITIATION ACTIVITIES OF THE ARCHAEAL RNA POLYMERASE IN A CHROMATIN LANDSCAPE

The machineries that control transcription initiation and elongation in Archaea and Eukarya are highly homologous. These similarities support the prevailing evolutionary theory of Archaea being the progenitor of Eukarya. Due to the retention of a core transcription apparatus, while lacking complexities of the eukaryotic counterpart, archaeal systems offer the unique potential to study and characterize the basic protein components necessary for transcription. Transcription termination was less well understood in both Archaea and Eukarya. Shared homology of the initiation and elongation phases argued for a homologous method of termination in Archaea and Eukarya. Additionally, both the archaeal and eukaryotic transcription apparatuses are frequently impeded by histone proteins bound to DNA. Like the transcription complex, archaeal histones are a simplified mirror to eukaryotic histones, permitting evaluation of all steps in the transcription cycle in the context of a chromatin landscape. This thesis summarizes the core molecular machineries involved in the regulation of archaeal transcription during elongation and termination in the greater context of archaeal histone-based chromatin. Thus, the discoveries made have contributed to both the transcription and chromatin fields by providing mechanistic details of the core, conserved transcription apparatus in the framework of evolution.

## ACKNOWLEDGMENTS

The last 4 and a half years have been life-changing. The many people guiding and supporting me throughout my time in graduate school have been essential to my success. This acknowledgement section serves as an attempt to thank everyone involved. Though no words are truly sufficient, I hope to express my deepest gratitude.

To my advisor, Tom Santangelo – Thank you for your support both personally and professionally. Your excitement for science and your expectation of greatness has made me a strong, enthusiastic scientist and a skilled communicator. Your guidance in matters of life and in matters of work has helped me through many difficult times. I see you as both a friend and mentor for life.

To Jeff Hansen, Olve Peersen and Asa Ben-Hur, my committee members – Thank you for continually challenging me. The different perspectives elevated the quality of my science and moved my projects forward. I feel lucky to have had a committee that both supported and challenged me.

To all the undergraduates who have helped me along the way, Gladys, Marshall, Paul Sudili, Bree, Barker, Jocelyn, Stavros and Robert – Thank you for teaching me to be a better mentor. Our science was made better by our teamwork and everyday was fun and exciting.

To Julie and Lexi – Thank you for mentoring me in my early graduate student days. Your scientific incite, professional advice and friendship helped me launch into a successful graduate career.

To Craig, Aldy, Kristin, Seré, Bree, Brett and Hallie – Thank you for everything inside and outside of lab. Our long conversations about science and life, the late nights in various bars, the long days in lab and our shared experience have made graduate school so much better. I will take the memories I have of our times in the Santangelo with me wherever I go.

Lastly, to my Mom and Dad, Debbie and Jim – Thank you for always believing in me in whatever I do and pushing me to always strive for something better. I am here because of both of you.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iii
CHAPTER 1: INTRODUCTION.....	1
1.1 The archaeal transcription cycle .....	1
1.2 Archaeal histones and extended, histone-based chromatin .....	4
1.3 Archaeal RNAP relies on conserved transcription factors to overcome histone-based barriers.....	5
1.4 The role of extended, histone-based chromatin in gene expression.....	9
1.5 FttA-mediated transcription termination .....	13
1.6 Thesis rationale .....	17
REFERENCES.....	18
CHAPTER 2: TFS AND SPT4/5 ACCELERATE TRANSCRIPTION THROUGH ARCHAEAAL HISTONE-BASED CHROMATIN.....	31
2.1 Summary.....	31
2.2 Results .....	32
2.3 Discussion .....	41
2.4 Materials and Methods.....	45
REFERENCES.....	49
CHAPTER 3: ARCHAEAAL HISTONE-BASED CHROMATIN STRUCTURE REGULATES GENE EXPRESSION IN <i>THERMOCOCCUS KODAKARENSIS</i> .....	55
3.1 Summary.....	55
3.2 Results .....	55
3.3 Discussion .....	69
3.4 Materials and Methods.....	71
3.5 Data Preprocessing.....	74
REFERENCES.....	76
CHAPTER 4: FTTA IS A CPSF73 HOMOLOGUE THAT TERMINATES TRANSCRIPTION IN ARCHAEA.....	81
4.1 Summary.....	81
4.2 Results .....	83
4.3 Discussion .....	97
4.4 Materials and Methods.....	100
REFERENCES.....	108
CHAPTER 5: DEVELOPING ARCHAEAAL NASCENT ELONGATING TRANSCRIPT SEQUENCING .....	115
5.1 Summary.....	115
5.2 Results .....	116
5.3 Discussion .....	120
5.4 Materials and Methods.....	125
REFERENCES.....	136

## CHAPTER 1: INTRODUCTION

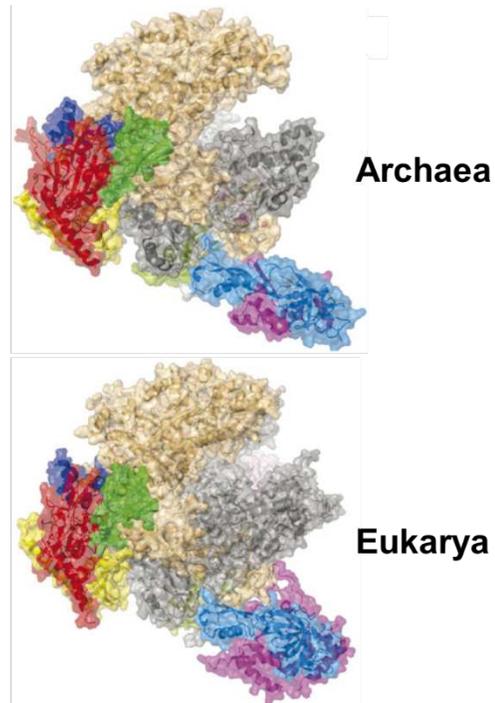
### 1.1 The archaeal transcription cycle

RNA polymerase is essential for life. A conserved set of catalytic subunits makes up the enzymatic core of RNA polymerase in Bacteria, Archaea and Eukarya<sup>1</sup>. Beyond this level of shared homology, Archaea and Eukarya encode an additional RNA polymerase structure known as the stalk domain. Thus, archaeal RNAP (RNAP) and the eukaryotic RNAP II (Pol II) share a degree of structural and functional homology beyond that of the bacterial RNAP<sup>2-7</sup>. The structural properties shared between RNAP and Pol II are reflected in similar regulatory mechanisms of transcription initiation, elongation and termination (Figure 1.1). Archaea utilize a component-reduced transcription apparatus when compared to that of Eukarya. This simplified system offers an ideal platform to dissect the basic mechanistic principles of transcription.

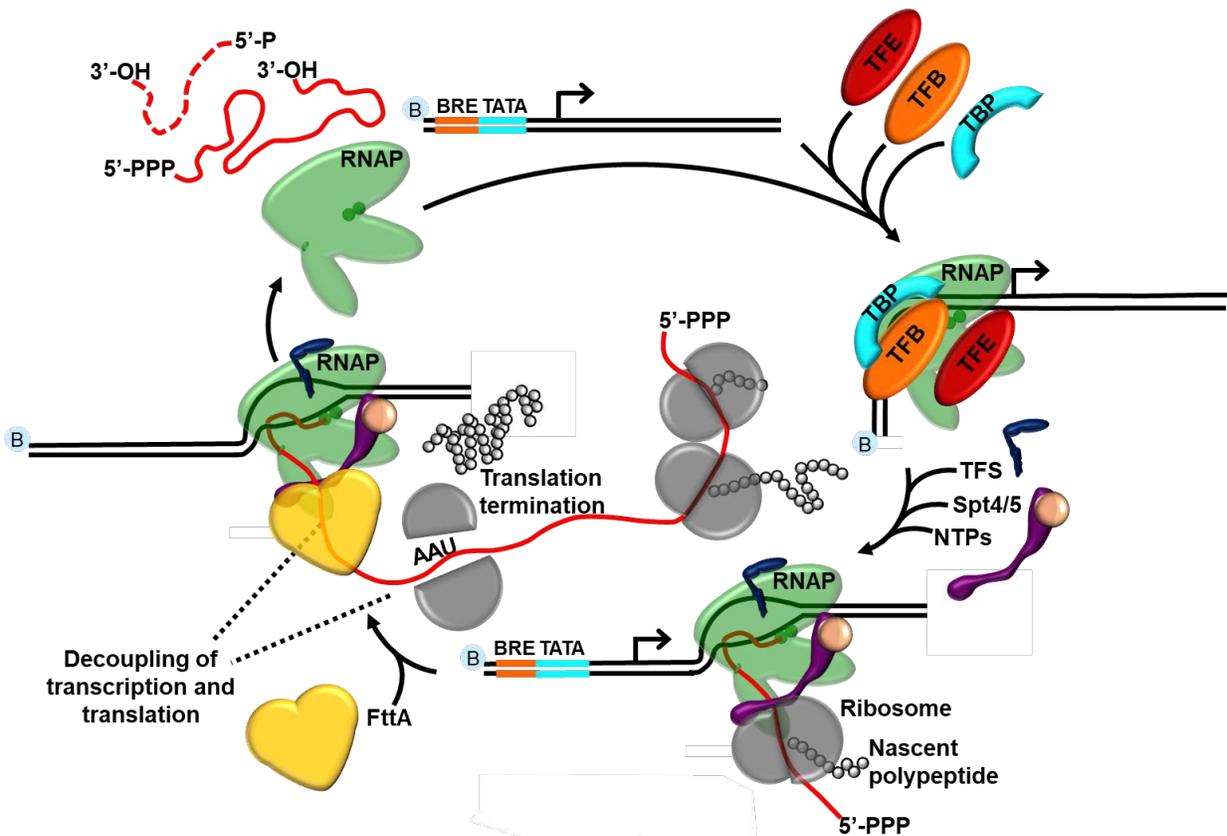
Archaeal transcription is demarcated into three phases: initiation, elongation and termination (Figure 1.2)<sup>8-10</sup>. Transcription in Archaea initiates at the promoter region of genes. Archaeal promoters typically consist of a B recognition element and a TATA box. Conserved transcription initiation factors transcription factor B (TFB, homologous to eukaryotic TFIIB) and TA-TA-binding protein (TBP, homologous to the eukaryotic TBP) recruit and position RNAP on DNA near the transcription start site<sup>10-12</sup>. Though TFB and TBP are sufficient for transcription initiation, transcription factor E (TFE, homologous to eukaryotic TFIIE $\alpha$ ) has been shown to participate in transcription initiation<sup>13-16</sup>. Early elongation is characterized by a release of RNAP (often with TFE) from initiation factors TFB and TBP into the gene body. Continued, stable elongation is characterized by Spt5

	Bacter	Archaea	Eukaryotes			Plants	
			RNAPII	RNAPIII	RNAPI	RNAPIV	RNAPV
RNAP subunits	$\beta'$	Rpo1 (A)	RPB1	C160	A190	NRPD1	NRPE1
	$\beta$	Rpo2 (B)	RPB2	C128	A135	NRPD/E2	NRPD/E2
	$\alpha$	Rpo3 (D)	RPB3	AC40	AC40	RPB3(+1)	RPB3(+1)
	$\alpha$	Rpo11 (L)	RPB11	AC19	AC19	RPB11	RPB11
	$\omega$	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
		Rpo13*	RPB9	C11	A12	NRPD9b	RPB9
	transcription factors			TFII $\alpha$	C53	A49	
			TFII $\beta$	C37	A34.5		
		TFE $\alpha$	TFIIE $\alpha$	C82			
		TFE $\beta$ /C34*	TFIIE $\beta$	C34			
				C31			
		TBP	TBP	TBP	TBP		
		TFB	TFIIB	Btf-1	TAF1B		
			TFIIA				
			TFIIF				
		TFS	TFIIS		TFIIS		
		Spt4	Spt4		Spt4		
		NusG	Spt5	Spt5	Spt5		
		NusA	NusA				
	<i>rho</i>						
	<i>Sigma</i>						
	<i>Gre</i>						

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



**Figure 1.1 RNA polymerase is universally conserved.** (left) The core subunits of RNA polymerase are retained in Bacteria, Archaea and Eukarya. Archaea and Eukarya encode additional RNA polymerase subunits, most notably the additional stalk domain (Archaea RpoF/E, Eukarya RPB4/7). Archaea utilize a reduced set of transcription initiation factors (minimally TFB and TBP) compared to the complexity of the eukaryotic system. A single transcription factor, Spt5 (NusG in Bacteria) is universally conserved. The cleavage stimulatory activity of TFS (TFIIS in Eukarya) is analogous to the GreA/GreB in Bacteria. (right) Crystal structures of the archaeal RNAP and eukaryotic Pol II demonstrate the high degree of structural homology.



**Figure 1.2 The archaeal transcription cycle consists of 3 steps: initiation, elongation and termination.** Promoter-directed assembly of pre-initiation complexes required RNAP, TBP and TFB and is often assisted by TFE. *De novo* RNA synthesis permits promoter escape and transcription initiation factors are replaced by transcription elongation factors TFS and Spt4-Spt5. The absence of a nuclear compartment permits translation initiation and the normal coupling of the archaeal transcription and translation apparatuses throughout transcription of the gene or operon, but this coupling is disrupted by translation termination. The exposed, nascent transcript likely permits the loading of FttA to transcription complexes and FttA activity mediates cleavage of nascent transcripts and release of RNAP to solution.

(and its binding partner Spt4) outcompeting TFE for the clamp domain of RNAP<sup>13,14</sup>. Once bound, Spt5-Spt4 stabilize RNAP in a processive conformation<sup>6,17,18</sup>. Processivity of RNAP is further bolstered by the cleavage stimulatory activity of TFS. Upon stochastic backtracking, or barrier induced backtracking of RNAP along the DNA, TFS stimulates RNAP to endonucleolytically cleave the RNA in the active site to generate a new 3'-OH, ultimately allowing for continued polymerization of RNA<sup>17,19–21</sup>.

Archaea are prokaryotic<sup>22,23</sup>. The lack of subcellular membrane compartmentalization allows the physical coupling of transcription and translation<sup>24,25</sup>. As such, the continuation of transcription is often accompanied by translation. The translating ribosome is likely linked to transcribing RNAP by the KOW domain of Spt5. Any transcription elongation stimulatory effect of coupling the translating ribosome to RNAP still requires extensive study. Upon the uncoupling of transcription and translation, preempted by the termination of translation, RNAP becomes susceptible to factor-mediated transcription termination<sup>8,25–29</sup>. The factor that terminates transcription in Archaea (FttA, homologous to the eukaryotic CPSF73 subunit of the cleavage and polyadenylation specificity complex) cleaves the nascent transcript and terminates transcription, recycling RNAP for additional rounds of transcription<sup>8,30</sup>.

## **1.2 Archaeal histones and extended, histone-based chromatin**

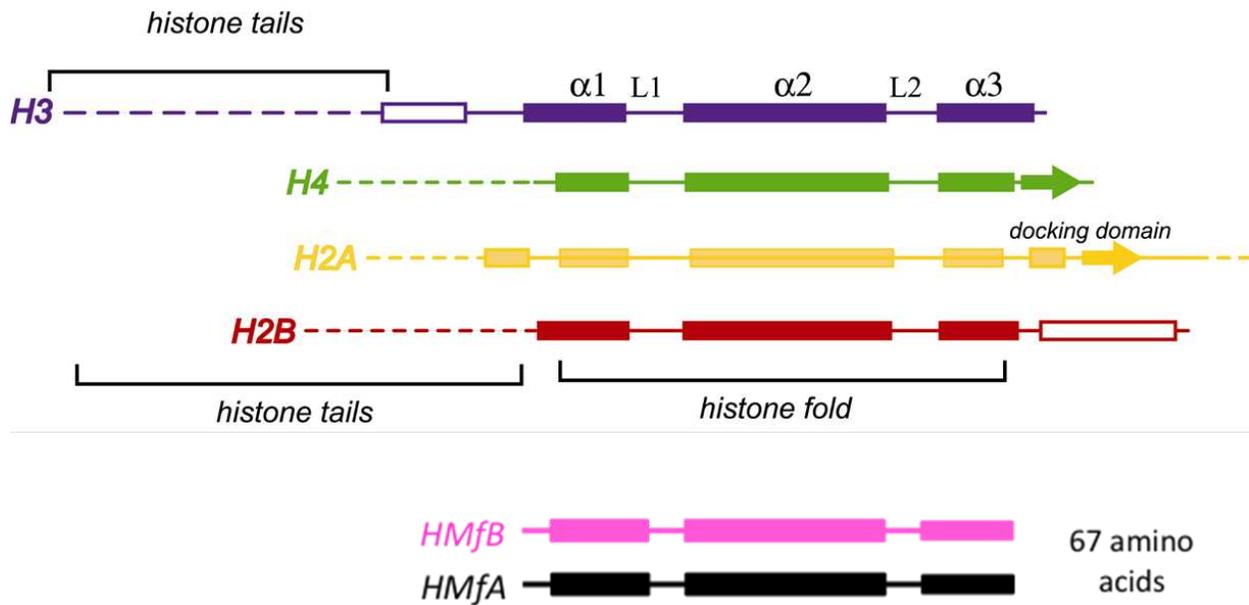
As in Eukarya, most Archaea encode histone proteins to organize their genomes, and the resultant chromatin structures alter transcription elongation rates, change positions of transcription pausing, and can impact elongation-termination decisions<sup>31–40</sup>. Thus, like the eukaryotic nuclear RNAPs, the archaeal RNAP regularly encounters – and must overcome – histone-based barriers to transcription elongation. Despite

interactions of the core histone fold with DNA being nearly identical between Eukarya and Archaea, the macro units of genome organization have diverged<sup>37,39-42</sup>. Most archaeal histones have no N-terminal tails or C-terminal extensions, allowing for continued polymerization of histone dimers (hetero- or homo-) to form an extended histone polymer that is not known to be post-translationally modified (Figure 1.3)<sup>33,40,43</sup>. Histone-based archaeal chromatin allows study of RNAP elongation kinetics in the absence of epigenetic complexity while still retaining the regulatory implications of histone-bound DNA.

### **1.3 Archaeal RNAP relies on conserved transcription factors to overcome histone-based barriers**

Given the absence of conserved chromatin remodeling/modification machinery in Archaea, conserved transcription factors likely play the dominant role in aiding transcription through histone-bound DNA. All archaea encode Spt5 (and its binding partner Spt4; Spt4-Spt5 is also commonly termed DSIF) and TFS<sup>17,33,44</sup>. Any contributions of TFS or Spt4-Spt5 in aiding transcription through a chromatin barrier likely result from distinct mechanisms. Spt5, often in complex with Spt4, is known to bind directly to RNAP and regulate the transition to stable elongation, promoter-proximal pausing, and processive elongation. Spt5 binds to RNAP via the clamp domain and facilitates increased processivity likely by retaining the clamp domain in a closed configuration<sup>18,45</sup>. Alignment of the template strand in the active center of RNAP is likely influenced by Spt4-Spt5 association with RNAP, and Spt4-Spt5 helps prevent arrest of Pol II via interactions with downstream DNA<sup>46-48</sup>.

Bacterial RNAP, archaeal RNAP, and eukaryotic RNAP I/II/III all exhibit intrinsic



**Figure 1.3 The core histone motif is highly conserved.** (top) The canonical eukaryotic histones (H2A, H2B, H3 and H4) contain the core histone fold as well as N-terminal and C-terminal extensions that are the site of prolific post-translational modification. (bottom) The vast majority of archaeal histones are simply the core histone fold.

endonuclease activity that can be stimulated by interactions with GreA and GreB (Bacteria), TFS (Archaea) and TFIIIS (also termed SII, Eukarya – Pol II)<sup>21,49–54</sup>. TFS and TFIIIS are homologous, whereas GreA/GreB are analogous in function. All cleavage stimulatory factors bind RNAP near the secondary channel and extend protein fingers towards the active center of RNAP. The tips of these protein extensions typically contain and donate acidic residues to the RNAP active center that stimulate endonuclease activity<sup>19–21</sup>. TFS interacts with RNAP through the secondary channel and primarily acts to restore catalytic activity to transcription elongation complexes (TECs) that have paused and subsequently reverse translocated to position an internal phosphodiester linkage in the bipartite active center of RNAP<sup>55</sup>. The lack of an RNA 3' -OH group in the active site of RNAP prevents further transcription. Continued pausing of RNAP may lead to DNA replication complex/TEC collisions, resulting in genomic instabilities<sup>56</sup>. Such 'backtracked' complexes are commonly the result of TECs encountering a strongly-bound protein or protein complex - such as histone-bound DNA – that slows/blocks RNAP forward translocation, thus stimulating TECs to reverse translocate<sup>17</sup>. Backtracked TECs may resume productive elongation either by spontaneous isomerization back to the forward position, thereby reestablishing the 3' end of the RNA in the active center, or by endonucleolytic cleavage of the RNA from a backtracked position that shortens the RNA and generates a new 3' end in the active center of RNAP<sup>46,57,58</sup>. 1-Dimensional diffusion forward to restore the RNA 3' end to the active center is stochastic and limited to shallow backtracking events. Endonucleolytic cleavage of the RNA to provide the active site with a new 3' end is independent of backtracking depth and is enhanced by protein factors, including TFS. Resumption of

elongation permits RNAP to approach the downstream barrier again, and if the protein roadblock is retained, multiple rounds of backtracking, subsequent cleavage, and elongation may occur.

Histone proteins are present in vivo to fully saturate the archaeal genome. In vitro addition of histone-proteins to DNA templates that contain stalled TECs results in the spontaneous assembly of chromatin structures that dramatically decrease ensemble transcription elongation rates<sup>17,5940,43</sup>. Different mechanisms of stabilizing TECs via binding of TFS, Spt4- or Spt5-alone, and Spt4-Spt5 complexes facilitated elongation through histone-bound DNAs and protein-free templates.

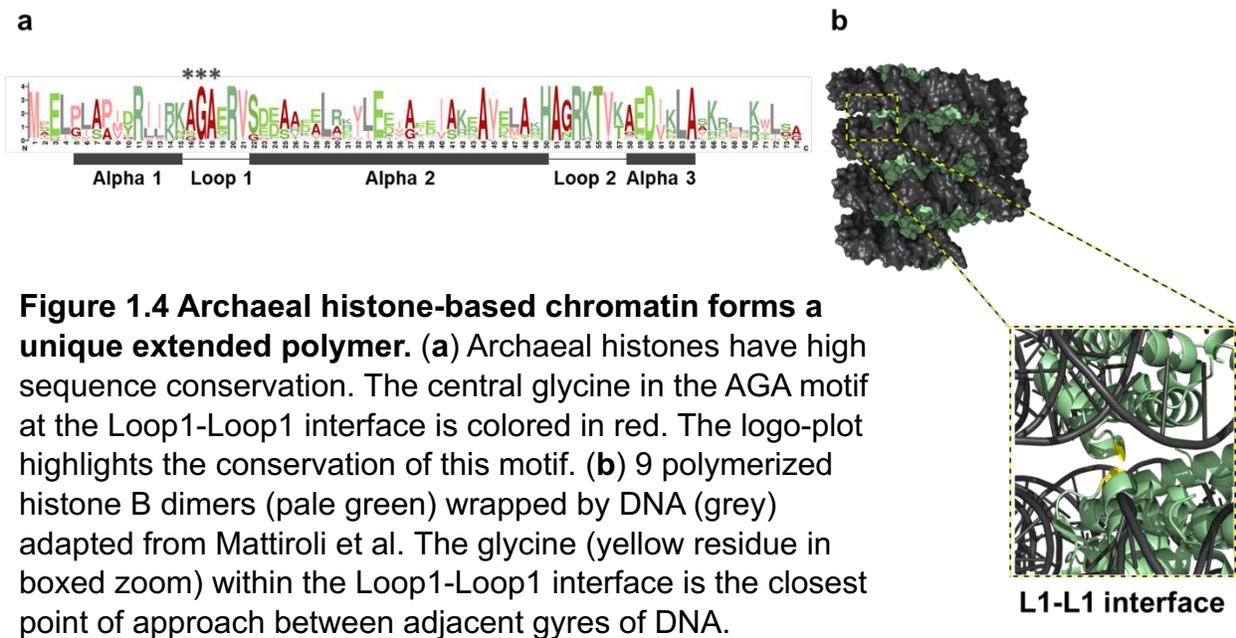
TFS stimulates endonucleolytic transcript cleavage and by reactivating backtracked TECs dramatically increases elongation rates on both protein-free and histone-bound DNA templates. Spt4 or spt5 alone has minimal effects on transcription elongation rates on protein-free or histone-bound templates and does not stimulate endonucleolytic cleavage of backtracked TECs nor reduce backtracking to any appreciable extent. Reconstitution of an spt4-spt5 complex does result in modest increases in the elongation rates on protein-free and histone-bound templates, suggesting that stabilizing TECs through clamp-closure can also assist in elongation through chromatin.

TECs repeatedly collide with protein barriers, backtrack and reattempt to transcribe through the barrier after regeneration of the RNA 3' end within the active center of RNAP. TFS-stimulating endonucleolytic cleavage appears to be the most efficient mechanism to restore a catalytically active RNAP that can repetitively attempt to translocate through downstream protein-roadblocks. Spt4-Spt5 likely enhance

elongation rates by binding to the clamp domain of RNAP, while making additional contacts with the downstream non-template strand stabilizing the TEC, facilitating processive elongation through chromatin by reducing pausing.

#### **1.4 The role of extended, histone-based chromatin in gene expression**

Despite the congruence of core histone-architecture between the two Domains, the chromatin structures in Archaea and Eukarya vary significantly<sup>32,40,43,60</sup>. In eukaryotic chromatin, DNA is wrapped around an octameric complex of histone heterodimers to form discrete nucleosomes<sup>32</sup>. In archaeal histone-based chromatin, DNA is wrapped around a symmetrical and semi-continuous helical ramp of homo- or heterogeneous archaeal histone dimers resulting in a superhelix<sup>40,61</sup>. The polymerization of histone dimers in Archaea does not resemble the nucleosome-nucleosome interactions observed in eukaryotes<sup>40,43</sup>. Alignments and analyses of >120 unique archaeal histone sequences reveal that the vast majority of archaeal histone proteins retain only the residues comprising the core eukaryotic histone-fold and that many amino acids are highly conserved in positions that form the histone-fold or mediate DNA interactions. The core histone fold consists of three alpha helices ( $\alpha 1$ ,  $\alpha 2$  and  $\alpha 3$ ) connected through two flexible loops (L1 and L2). Perhaps surprisingly, a conserved alanine-glycine-alanine (AGA) motif in L1 was identified that was not obviously involved in protein-DNA interactions or supporting the core histone-fold (Figure 1.4). The crystal structure of archaeal histone-based chromatin provided the first clues to the importance of the highly conserved motif in L1, wherein the small side chains of these nearly-invariant amino acids permitted the close proximity of adjacent gyres of DNA within the superhelix at the Loop1-Loop1 (L1-L1) interface. The



**Figure 1.4 Archaeal histone-based chromatin forms a unique extended polymer.** (a) Archaeal histones have high sequence conservation. The central glycine in the AGA motif at the Loop1-Loop1 interface is colored in red. The logo-plot highlights the conservation of this motif. (b) 9 polymerized histone B dimers (pale green) wrapped by DNA (grey) adapted from Mattioli et al. The glycine (yellow residue in boxed zoom) within the Loop1-Loop1 interface is the closest point of approach between adjacent gyres of DNA.

conservation of these residues was thus hypothesized to support the biologically-important, tight-packing of archaeal histone-based chromatin architectures<sup>40,62</sup>.

Archaeal histone-based chromatin digestion patterns reveal the existence of histone-DNA superstructures that vary in length from protection of ~60 bp of DNA by a histone tetramer to protection of ~300 bp of DNA by a semi-continuous polymer of ~10 symmetrically bound histone dimers (here termed a histone polymer; each additional histone dimer can wrap ~30 bp of DNA to extend the chromatin superstructure)<sup>40,63,64</sup>. While dynamic changes in histone polymer length are expected, it is also likely that specific DNA sequences and loci are more likely to retain extended chromatin architectures that play important roles in regulating gene expression. The presence of several histone isoforms in many Archaea suggests the potential for variation in superhelix composition or length, but the viability of strains with a single histone demonstrate that even homopolymers permit formation of extended histone-based architectures in Archaea<sup>37,42,65,66</sup>. The biological importance of the extended superhelical chromatin structures is supported by the evolutionary retention of the AGA motif in L1 which permits close association of adjacent superhelical gyres<sup>40,62</sup>. Importantly, allelic substitutions of the central glycine within the L1-AGA motif abrogated formation of the archaeal histone-based chromatin-superhelix, decreased cellular fitness and resulted in the loss of adaptive gene expression – measured for a single operon – when actively growing cultures were moved to a new metabolic environment<sup>40</sup>.

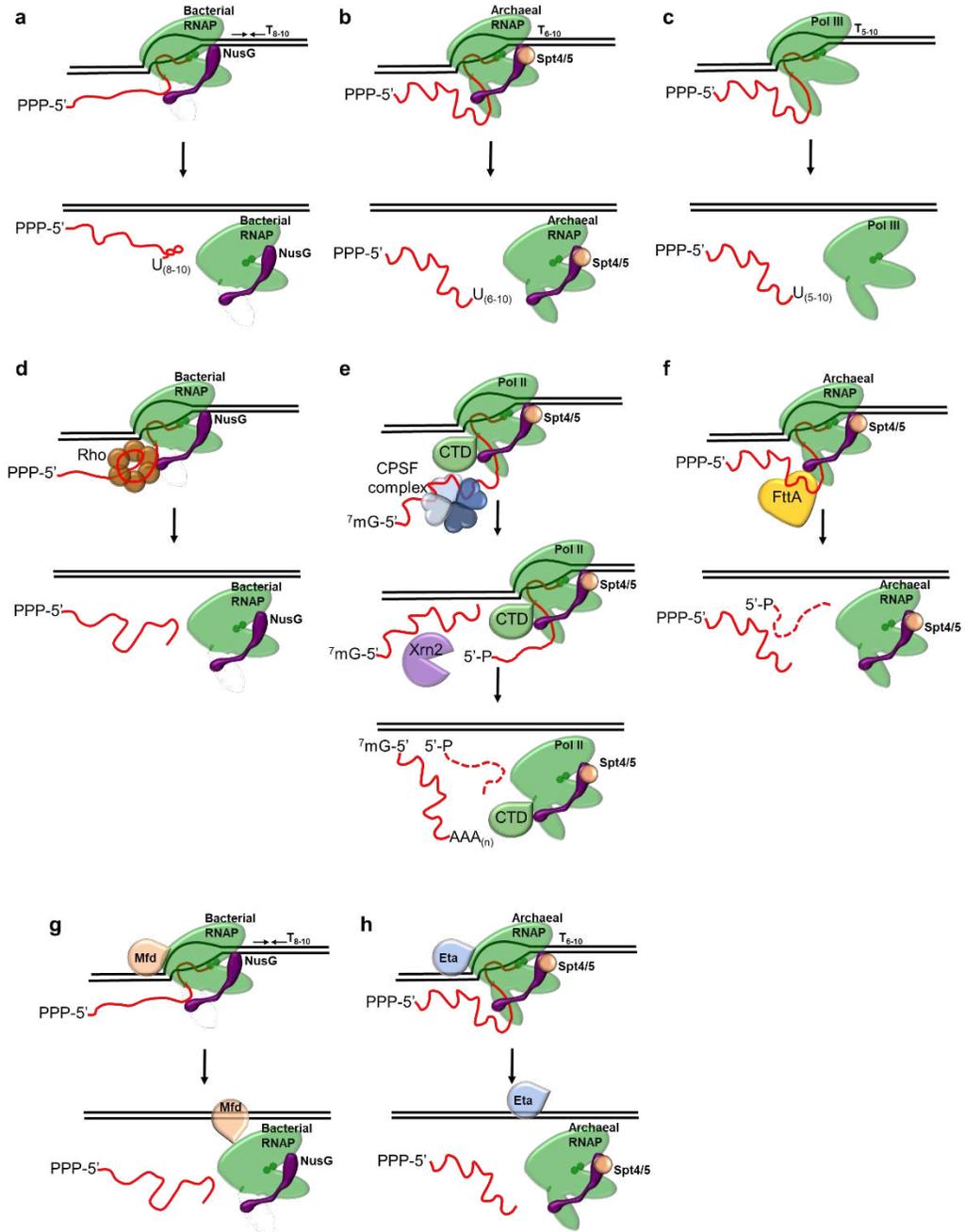
The importance of superhelix formation in modulating gene expression suggests unique archaeal gene expression regulation strategies that take advantage of mechanisms to retain or abolish extended archaeal histone-based chromatin

structures<sup>40,63,64,66</sup>. To evaluate the normal contribution of histone polymers strains of the model archaeon *Thermococcus kodakarensis*<sup>67</sup> wherein histone variants impacted global genomic architecture and quantified the gene expression changes resultant from modified chromatin architectures were generated. *T. kodakarensis* normally encodes two closely related histone isoforms, termed Histone A (HTkA) and Histone B (HTkB), but strains encoding only a single histone variant are viable<sup>12,65,68,69</sup>. To evaluate the consequences of altering extended chromatin landscapes in archaeal cells, strains of *T. kodakarensis* were constructed to encode only HTkA in WT form, HTkA<sup>WT</sup>, only HTkA with a single amino acid substitution (G17 to D), HTkA<sup>G17D</sup>, or retain HTkB in WT form and express HTkA<sup>G17D</sup>. Retention of a single WT histone variants is sufficient to maintain an extended chromatin architecture, but that extended chromatin structures are abolished in strains encoding only HTkA<sup>G17D</sup>. Comparative differential RNA-seq analyses of strains with unique chromatin landscapes reveals substantial and genome-wide variations in gene expression patterns, underscoring the importance of extended chromatin structures in normal regulation of gene expression. Expression differences observed in strains lacking extended chromatin architectures suggest architectural changes in chromatin are most impactful for the proper regulated expression of chemotaxis-, motility- and proviral-encoding regions of the genome. The global regulatory potential of an extended archaeal histone based-chromatin superstructure is thus confirmed and offers impactful routes to control archaeal gene expression by modulating chromatin architectures and histone-based superhelix formation.

## 1.5 FttA-mediated transcription termination

The extreme stability required of transcription elongation complexes (TECs) to processively transcribe large, chromatinized genomic regions necessitates robust mechanisms to terminate transcription. Transcription termination, driven either by DNA sequence and encoded RNA structures (e.g. intrinsic termination) or by protein factors (e.g. factor-dependent termination) ensures rapid dissociation of RNA polymerase (RNAP) from the DNA template to maintain efficient transcription, recycle RNAP, and generate RNA 3' ends (Figure 1.5)<sup>8,70-75</sup>. Intrinsic termination can efficiently disrupt bacterial, archaeal, and eukaryotic RNA Pol III (Figure 1.5a-c) complexes while eukaryotic RNA Pol I and Pol II require factor-mediated termination events<sup>26,76-78</sup>. While often prevalent within genomes, intrinsic termination sequences are typically not sufficiently abundant to mediate all termination events, and when TECs become arrested at DNA lesions, specific proteins or protein complexes are required to release RNAP from the DNA template<sup>26,28,30,71,74,79-82</sup>. For termination factors to serve as global governance regulators of aberrant and run-on transcription each must efficiently recognize TECs and compete with continued elongation to mediate release of the nascent transcript. While such proteins have been identified in bacteria and eukaryotic species, no kinetically-efficient mechanism of factor-dependent transcription termination had been described in Archaea<sup>8,10,44</sup>.

In Bacteria, the ATP-dependent helicase, Rho, acts to terminate transcription upon the uncoupling of transcription and translation by recognizing a ribosome-free C-rich RNA sequence<sup>26-28,83-85</sup>. Once bound, Rho hydrolyzes ATP to translocate in the 5'-3' direction, ultimately encountering RNAP and terminating transcription (Figure 1.5d).



**Figure 1.5 Transcription termination mechanisms commonly employed in Bacteria, Eukarya and Archaea.**

Rho-mediated termination is aided by the activity of universally conserved transcription elongation factor NusG (Spt5 in Eukarya). NusG both allosterically activates Rho and kinetically couples Rho termination activity with transcription elongation<sup>83,86</sup>. The DNA repair coupling factor, Mfd, acts to terminate stalled RNAP at sites of DNA damage (Figure 1.5g)<sup>79,80</sup>. Mfd recruits repair enzymes and removes RNAP from the DNA template by way of translocation forward. An analogous mechanism of DNA repair associated transcription termination by the protein factor Eta has been described in the Euryarchaea (Figure 1.5h)<sup>82</sup>.

In Eukarya, the CPSF73 subunit of the cleavage and polyadenylation complex cleaves RNA downstream of the polyadenylation signal<sup>87-89</sup>, allowing Xrn2/Rat1 to bind and degrade the uncapped RNA associated with Pol II<sup>70,71,73,74,90</sup>. CPSF73, as part of the RNA 3'-maturation machinery, initiates cleavage of the nascent transcript, a necessary step for mRNA maturation that in-of-itself does not direct transcription termination; Pol II can continue transcription for thousands of base pairs beyond the cleavage position<sup>91,92</sup>. Cleavage by the CPSF complex instead generates a 5'-uncapped, monophosphate end that permits Rat1/Xrn2-mediated degradation of the nascent transcript up to the boundary imposed by Pol II. Although Rat1/Xrn2 is not capable of directly disrupting the Pol II TEC, the combined activities of CPSF and Xrn2 are necessary for normal termination patterns in Eukarya (Figure 1.5e)<sup>74</sup>.

The prokaryotic nature and observation of polarity (defined as reduced expression of a downstream gene in an operon controlled by premature transcription termination during expression of upstream genes) in Archaea suggested a Rho-like mechanism wherein a protein factor recognizes RNA sequence elements upon

uncoupling of transcription from translation<sup>24,93</sup>. Archaea are the progenitors of Eukarya and archaeal information processing systems are typically component simplified versions of the more complex eukaryotic systems that dictate replication, repair and transcription within a chromatin landscape<sup>22,23,94</sup>. The conserved structures of multi-subunit RNAPs and the shared mechanisms of initiation and elongation between the archaeal and eukaryotic transcription apparatuses argued that the final phase of transcription – termination – would also be conserved. Only a core set of 129 genes are conserved in all sequenced archaeal genomes and the vast majority of these have obvious eukaryotic, rather than bacterial homologues<sup>95</sup>. One of these universally conserved and essential archaeal proteins is an obvious orthologue of a subunit of the cleavage and polyadenylation specificity factor (CPSF) complex<sup>96–98</sup>. The homology of most archaeal transcription components to eukaryotic factors argued that the archaeal CPSF homologue (aCPSF) might function as the Factor that terminates transcription in Archaea (FttA) (Figure 1.5f).

Using components purified from the model archaeal species *Thermococcus kodakarensis* we established that FttA can disrupt the otherwise extremely stable TEC and is thus a bona fide archaeal transcription termination factor. Identification and characterization of FttA completes the archaeal transcription cycle and links transcription regulation between the three Domains. FttA-mediated transcription termination shares many attributes with Rho-mediated transcription termination, but FttA retains a structure with obvious homology to the eukaryotic CPSF73 protein. FttA is kinetically coupled to RNAP by the only universally conserved transcription elongation factor, Spt5 and the archaeal-eukaryotic specific RNAP stalk domain. FttA provides a

missing-link between prokaryotic and eukaryotic transcription regulation and provides the most parsimonious link to the evolution of the processing activities involved in RNA 3'-end formation in Eukarya<sup>8,30,97</sup>.

## **1.6 Thesis rationale**

The work presented here completes the archaeal transcription cycle while exploring the depth of gene expression strategies used during elongation and termination. The role of conserved transcription factors TFS and Spt4-Spt5 will be explored in the context of chromatin. Further, the perturbation of conserved 3-dimensional chromatin structure in *T. kodakarensis* contributes to the understanding of histone-based chromatin both evolutionarily and structurally. Finally, the discovery and characterization of FttA as the universally conserved archaeal transcription termination factor completes the archaeal transcription cycle. These works support continued study of the transcription apparatus. Development of Nascent Elongating Transcript sequencing (NET-seq)<sup>50,91,99</sup> in combination with elongation factor/termination factor/histone-based chromatin genetic perturbation will facilitate the future study of global RNAP position in the context of transcription elongation, termination and protein barriers.

## REFERENCES

1. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Publ. Gr.* **9**, 85–98 (2011).
2. Grohmann, D., Hirtreiter, A. & Werner, F. Molecular mechanisms of archaeal RNA polymerase. *Biochem. Soc. Trans.* **37**, 12–17 (2009).
3. Nagy, J., Grohmann, D., Cheung, A. C. M., Schulz, S., Smollett, K., Werner, F. & Michaelis, J. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun.* **6**, 6161 (2015).
4. Werner, F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* **16**, 247–250 (2008).
5. Hirata, A., Kanai, T., Santangelo, T. J., Tajiri, M., Manabe, K., Reeve, J. N., Imanaka, T. & Murakami, K. S. 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**, 623–633 (2008).
6. Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* **2**, 17021 (2017).
7. Reich, C., Zeller, M., Milkereit, P., Hausner, W., Cramer, P., Tschochner, H. & Thomm, M. The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable *in vivo* and *in vitro*. *Mol. Microbiol.* **71**, 989–1002 (2009).
8. Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker,

- J. E. & Santangelo, T. J. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* **5**, 545–553 (2020).
9. Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P. & Thomm, M. The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.* **41**, 7048–7059 (2013).
  10. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription Regulation in Archaea. *J. Bacteriol.* **198**, 1906–1917 (2016).
  11. Santangelo, T. J., Cubonová, L., James, C. L. & Reeve, J. N. TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro. *J. Mol. Biol.* **367**, 344–57 (2007).
  12. Gehring, A. M. & Santangelo, T. J. in *Methods Mol. Biol.* **1276**, 263–279 (2015).
  13. Schulz, S., Gietl, A., Smollett, K., Tinnefeld, P., Werner, F. & Grohmann, D. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl. Acad. Sci.* **113**, E1816–E1825 (2016).
  14. Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J. & Werner, F. The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol. Cell* **43**, 263–274 (2011).
  15. Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C., Smollett, K. L., Albers, S. V., Kay, C. W. M., Thalassinou, K. & Werner, F. Archaeal TFE $\alpha/\beta$  is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife* **4**, (2015).
  16. Blombach, F., Ausiannikava, D., Figueiredo, A. M., Soloviev, Z., Prentice, T.,

- Zhang, M., Zhou, N., Thalassinou, K., Allers, T. & Werner, F. Structural and functional adaptation of *Haloferax volcanii* TFEI. *Nucleic Acids Res.* **46**, 2308–2320 (2018).
17. Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019). doi:10.1111/mmi.14191
  18. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
  19. Hausner, W., Lange, U. & Musfeldt, M. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**, 12393–9 (2000).
  20. Fouqueau, T., Blombach, F., Hartman, R., Cheung, A. C. M., Young, M. J. & Werner, F. The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nat. Commun.* **8**, 1914 (2017).
  21. Lange, U. & Hausner, W. Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol. Microbiol.* **52**, 1133–1143 (2004).
  22. Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A. & Ettema, T. J. G. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).

23. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
24. Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T. & Reeve, J. N. Polarity in Archaeal Operon Transcription in *Thermococcus kodakaraensis*. *J. Bacteriol.* **190**, 2244–2248 (2008).
25. Korn, L. J. & Yanofsky, C. Polarity suppressors defective in transcription termination at the attenuator of the tryptophan operon of *Escherichia coli* have altered rho factor. *J. Mol. Biol.* **106**, 231–41 (1976).
26. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85**, 319–47 (2016).
27. Strauß, M., Vitiello, C., Schweimer, K., Gottesman, M., Rösch, P. & Knauer, S. H. Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res.* **44**, 5971–5982 (2016).
28. Hart, C. M. & Roberts, J. W. Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *J. Biol. Chem.* **266**, 24140–8 (1991).
29. Mitra, P., Ghosh, G., Hafeezunnisa, M. & Sen, R. Rho Protein: Roles and Mechanisms. *Annu. Rev. Microbiol.* **71**, 687–709 (2017).
30. Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., Li, L., Zheng, X. & Dong, X. The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. *Nucleic Acids Res.* **48**, (2020).
31. Chang, C. H. & Luse, D. S. The H3/H4 tetramer blocks transcript elongation by

- RNA polymerase II in vitro. *J. Biol. Chem.* **272**, 23427–34 (1997).
32. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
  33. Sanders, T. J., Marshall, C. J. & Santangelo, T. J. The Role of Archaeal Chromatin in Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.05.006
  34. Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-I. & Kurumizaka, H. Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595–598 (2018).
  35. Luse, D. S., Spangler, L. C. & Újvári, A. Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors. *J. Biol. Chem.* **286**, 6040–8 (2011).
  36. Kim, J., Guermah, M. & Roeder, R. G. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
  37. Sandman, K. & Reeve, J. N. Chromosome packaging by archaeal histones. *Adv. Appl. Microbiol.* **50**, 75–99 (2001).
  38. Nishida, H. & Oshima, T. Archaeal histone distribution is associated with archaeal genome base composition. *J. Gen. Appl. Microbiol.* **63**, 28–35 (2017).
  39. Decanniere, K., Babu, A. M., Sandman, K., Reeve, J. N. & Heinemann, U. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* **303**, 35–47 (2000).
  40. Mattioli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart,

- B. W., Byrne, K. R., Lee, T., Ahn, N. G., Santangelo, T. J., Reeve, J. N. & Luger, K. Structure of histone-based chromatin in Archaea. *Science* (80-. ). **357**, 609–612 (2017).
41. Richmond, R. K., Sargent, D. F., Richmond, T. J., Luger, K. & Ma, A. W. Crystal structure of the nucleosome ° resolution core particle at 2 . 8 A. **7**, 251–260 (1997).
  42. Sandman, K. & Reeve, J. N. Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.* **9**, 520–525 (2006).
  43. Bhattacharyya, S., Mattioli, F. & Luger, K. Archaeal DNA on the histone merry-go-round. *FEBS J.* **285**, 3168–3174 (2018).
  44. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.06.020
  45. Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P. & Werner, F. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
  46. Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. & Cramer, P. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560**, 607–612 (2018).
  47. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S. & Handa, H. DSIF, a novel transcription elongation factor that regulates RNA polymerase II

- processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–56 (1998).
48. Guo, G., Gao, Y., Zhu, Z., Zhao, D., Liu, Z., Zhou, H., Niu, L. & Teng, M. Structural and biochemical insights into the DNA-binding mode of Mj Spt4p:Spt5 complex at the exit tunnel of RNAPII. *J. Struct. Biol.* **192**, 418–425 (2015).
  49. Ishibashi, T., Dangkulwanich, M., Coello, Y., Lionberger, T. A., Lubkowska, L., Ponticelli, A. S., Kashlev, M. & Bustamante, C. Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics. *Proc. Natl. Acad. Sci.* **111**, 3419–3424 (2014).
  50. Imashimizu, M., Takahashi, H., Oshima, T., McIntosh, C., Bubunencko, M., Court, D. L. & Kashlev, M. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.* **16**, (2015).
  51. Toulmé, F., Mosrin-Huaman, C., Sparkowski, J., Das, A., Leng, M. & Rachid Rahmouni, A. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* **19**, 6853–6859 (2000).
  52. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334 (2003).
  53. Marr, M. T. & Roberts, J. W. Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol. Cell* **6**, 1275–1285 (2000).
  54. Chédin, S., Riva, M., Schultz, P., Sentenac, A. & Caries, C. The RNA cleavage activity of RNA polymerase III is mediated an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev.* **12**, 3857–3871 (1998).

55. Symersky, J., Perederina, A., Vassilyeva, M. N., Svetlov, V., Artsimovitch, I. & Vassilyev, D. G. Regulation through the RNA polymerase secondary channel: Structural and functional variability of the coiled-coil transcription factors. *J. Biol. Chem.* **281**, 1309–1312 (2006).
56. Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* **20**, 412–418 (2013).
57. Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M. & Sekine, S. I. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science (80-. )*. **357**, 921–924 (2017).
58. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7090–7095 (2000).
59. Xie, Y. & Reeve, J. N. Transcription by an Archaeal RNA Polymerase Is Slowed but Not Blocked by an Archaeal Nucleosome. *J. Bacteriol.* **186**, 3492–3498 (2004).
60. Sandman, K. & Reeve, J. N. Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.* **173**, 165–9 (2000).
61. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
62. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
63. Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.-P., Reeve, J. N.,

- Santangelo, T. J. & Fondufe-Mittendorf, Y. N. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* **14**, 391 (2013).
64. Rojec, M., Hocher, A., Stevens, K. M., Merckenschlager, M. & Warnecke, T. Chromatinization of escherichia coli with archaeal histones. *Elife* **8**, (2019).
65. Cubonova, L., Katano, M., Kanai, T., Atomi, H., Reeve, J. N. & Santangelo, T. J. An Archaeal Histone Is Required for Transformation of *Thermococcus kodakarensis*. *J. Bacteriol.* **194**, 6864–6874 (2012).
66. Dulmage, K. A., Todor, H. & Schmid, A. K. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *MBio* **6**, e00649-15 (2015).
67. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**, 352–363 (2005).
68. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *BIO-PROTOCOL* **7**, (2017).
69. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic Techniques for the Archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).
70. Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., Nguyen, T., Karp, S. & Bentley, D. L. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest

- Widespread Kinetic Competition. *Mol. Cell* **60**, 256–267 (2015).
71. Luo, W. & Bentley, D. A Ribonucleolytic Rat Torpedoes RNA Polymerase II. *Cell* **119**, 911–914 (2004).
  72. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M. & Bentley, D. L. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–61 (1997).
  73. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K. & Bentley, D. L. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Mol. Cell* **76**, 896-908.e4 (2019).
  74. Eaton, J. D., Francis, L., Davidson, L. & West, S. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145 (2020).
  75. Svetlov, V. & Nudler, E. Towards the unified principles of transcription termination. *EMBO J.* **39**, (2020).
  76. Santangelo, T. J. & Reeve, J. N. Archaeal RNA Polymerase is Sensitive to Intrinsic Termination Directed by Transcribed and Remote Sequences. *J. Mol. Biol.* **355**, 196–210 (2006).
  77. Santangelo, T. J., Cubonova, L., Skinner, K. M. & Reeve, J. N. Archaeal Intrinsic Transcription Termination In Vivo. *J. Bacteriol.* **191**, 7102–7108 (2009).
  78. Dar, D., Shamir, M., Mellin, J. R., Koutero, M., Stern-Ginossar, N., Cossart, P. & Sorek, R. Term-seq reveals abundant ribo-regulation of antibiotics resistance in

- bacteria. *Science* (80-. ). **352**, aad9822–aad9822 (2016).
79. Park, J.-S., Marr, M. T. & Roberts, J. W. E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**, 757–67 (2002).
80. Le, T. T., Yang, Y., Tan, C., Suhanovsky, M. M., Fulbright, R. M., Inman, J. T., Li, M., Lee, J., Perelman, S., Roberts, J. W., Deaconescu, A. M. & Wang, M. D. Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell* **172**, 344-357.e15 (2018).
81. West, S., Gromak, N. & Proudfoot, N. J. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525 (2004).
82. Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci.* **114**, E6767–E6773 (2017).
83. Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**, 935–8 (2008).
84. Epshtein, V., Dutta, D., Wade, J. & Nudler, E. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**, 245–249 (2010).
85. Roberts, J. W. Mechanisms of Bacterial Transcription Termination. *J. Mol. Biol.* **431**, 4030–4039 (2019).
86. Dar, D., Prasse, D., Schmitz, R. A. & Sorek, R. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat. Microbiol.* **1**, 16143 (2016).

87. Clerici, M., Faini, M., Muckenfuss, L. M., Aebersold, R. & Jinek, M. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat. Struct. Mol. Biol.* **25**, 135–138 (2018).
88. Kolev, N. G., Yario, T. A., Benson, E. & Steitz, J. A. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO Rep.* **9**, 1013–1018 (2008).
89. Fusby, B., Kim, S., Erickson, B., Kim, H., Peterson, M. L. & Bentley, D. L. Coordination of RNA Polymerase II Pausing and 3' end processing factor recruitment with alternative polyadenylation. *Mol. Cell. Biol.* **36**, MCB.00898-15 (2015).
90. Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K. C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J. & Cramer, P. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* **66**, 38-49.e6 (2017).
91. Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).
92. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
93. French, S. L., Santangelo, T. J., Beyer, A. L. & Reeve, J. N. Transcription and Translation are Coupled in Archaea. *Mol. Biol. Evol.* **24**, 893–895 (2007).
94. López-García, P. & Moreira, D. Cultured Asgard Archaea Shed Light on

- Eukaryogenesis. *Cell* **181**, 232–235 (2020).
95. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46 (2012).
  96. Silva, A. P. G., Chechik, M., Byrne, R. T., Waterman, D. G., Ng, C. L., Dodson, E. J., Koonin, E. V., Antson, A. A. & Smits, C. Structure and activity of a novel archaeal  $\beta$ -CASP protein with N-terminal KH domains. *Structure* **19**, 622–632 (2011).
  97. Phung, D. K., Rinaldi, D., Langendijk-Genevaux, P. S., Quentin, Y., Carpousis, A. J. & Clouet-d'Orval, B. Archaeal  $\beta$ -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res.* **41**, 1091–1103 (2013).
  98. Chavarria, N. E., Hwang, S., Cao, S., Fu, X., Holman, M., Elbanna, D., Rodriguez, S., Arrington, D., Englert, M., Uthandi, S., Söll, D. & Maupin-Furlow, J. A. Archaeal Tuc1/Ncs6 homolog required for wobble uridine tRNA thiolation is associated with ubiquitin-proteasome, translation, and RNA processing system homologs. *PLoS One* **9**, (2014).
  99. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).

## CHAPTER 2: TFS AND SPT4/5 ACCELERATE TRANSCRIPTION THROUGH ARCHAEOAL HISTONE-BASED CHROMATIN<sup>1</sup>

### 2.1 Summary

Transcription complexes must remain processively engaged with DNA templates to generate long transcripts, but transcription elongation is neither uniform nor continuous<sup>2-7</sup>. RNA polymerase frequently encounters - and must surmount - translocation barriers for continued transcription. In Eukarya and most Archaea, DNA-bound histone proteins represent the most common and troublesome barrier to transcription elongation<sup>1,3,8-12</sup>. Eukaryotes encode a plethora of chromatin-remodeling complexes, histone-modification enzymes and transcription elongation factors to aid transcription through nucleosomes, while archaea seemingly lack machinery to remodel/modify histone-based chromatin and thus must rely on elongation factors to accelerate transcription through chromatin-barriers<sup>1,12-15</sup>. Only two transcription elongation factors – TFS (TFIIS in Eukarya) and the Spt4-Spt5 complex - are universally encoded in archaeal genomes, and here we demonstrate that both elongation factors, via different mechanisms, can accelerate transcription through archaeal histone-based chromatin<sup>1,13,15</sup>. TFS-enhanced cleavage of RNAs in backtracked transcription complexes reactivates stalled RNAPs and dramatically accelerates transcription through histone-barriers<sup>1,16-18</sup>, while Spt4-Spt5 mediated changes to clamp-domain dynamics play a lesser-role in stabilizing transcription elongation<sup>19-24</sup>. The essentiality of both

---

<sup>1</sup> This chapter was published in March 2019 under the same title.

Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019).

conserved transcription elongation factors establishes that both mechanisms to accelerate transcription through downstream protein barriers are required for efficient transcription in vivo.

## **2.2 Results**

### ***The DNA of *Thermococcus kodakarensis* is completely bound by histones.***

To ensure that in vitro studies on chromatin templates accurately reflected in vivo conditions we determined the concentration of histone proteins in *T. kodakarensis*. Quantitative Western blots, using DNaseI treated cellular lysates derived from *T. kodakarensis* cells and polyclonal antibodies that recognize both histone isoforms (HTkA and HTkB) revealed the steady-state abundance of histone proteins in vivo (Figure 2.1).

Establishing Western blot signal intensity curves using known concentrations of highly-purified HTkA and HTkB allowed us to establish the total number of histone molecules in cellular lysates. *T. kodakarensis* polyploid, retaining ~9-17 genomes per cell, however, cells retain sufficient histone proteins to completely bind all genomes. Thus, histone proteins were added to in vitro transcription reactions in saturating amounts.

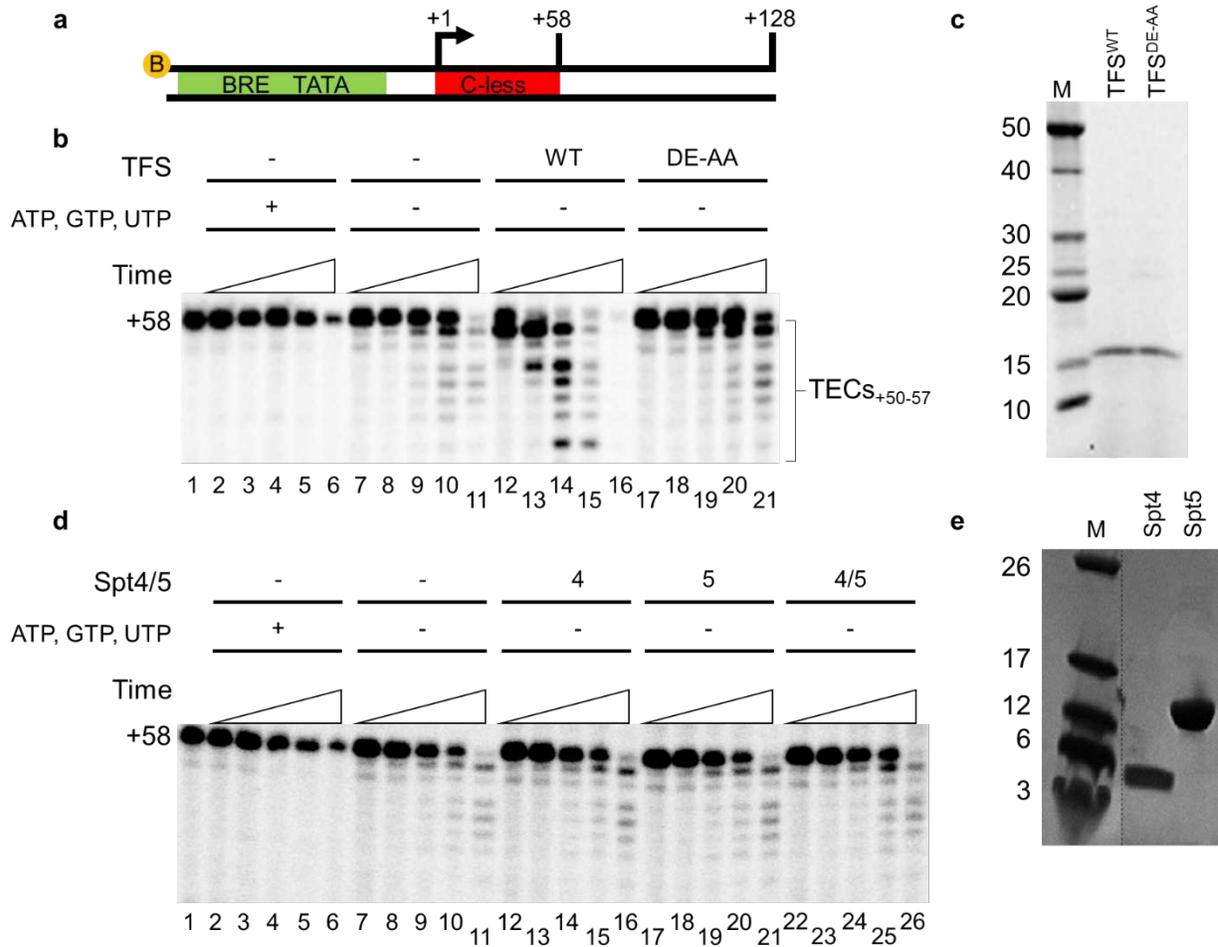
### ***Transcription Factor S (TFS), but not Spt4-Spt5, stimulates archaeal RNAP RNA-cleavage activity.***

The component-purified, promoter-directed transcription system from *T. kodakarensis* permits formation of stalled TECs at defined template positions via nucleotide deprivation (Figure 2.2a)<sup>25</sup>. When conditions do not permit continued polymerization, TECs<sub>+58</sub> spontaneously backtrack and slowly cleave nascent transcripts

Replica	A			B			C			HTkB (ng)				HTkA (ng)			
Cells lysed (x10 <sup>9</sup> )	0.45	1.4	2.2	0.63	1.9	3.1	0.39	1.2	1.9	10	100	300	500	10	100	300	500



**Figure 2.1 The genomes of *T. kodakarensis* are completely bound by histone proteins.** Quantitative Western blot analyses of DNaseI treated *T. kodakarensis* lysates with polyclonal anti-HTkA antibodies demonstrates histone protein levels – HTkA and HTkB - are sufficient to bind the entirety of the the *T. kodakarensis* genomes.



**Figure 2.2 TFS, but not Spt4-Spt5, stimulates intrinsic RNAP endonuclease activity.**

(a) Biotinylated DNA templates permit promoter directed transcription to generate stalled TECs at the end of a 58 bp C-less cassette. (b) TECs<sub>+58</sub> spontaneously backtrack and cleave nascent transcripts (lanes 7-11) to yield TECs<sub>+50-57</sub>. When NTPs are present, TECs rapidly re-elongate to +58 (lanes 2-6). The rate of nascent transcript cleavage is stimulated by addition of TFS<sup>WT</sup> (lanes 12-16) but not by addition of TFS<sup>DE-AA</sup> (lanes 17-21). (c) Coomassie-stained, SDS-PAGE of purified TFS<sup>WT</sup> and the inactive mutant TFS<sup>DE-AA</sup>. Lane M contains size standards labeled in Kda to the left. (d) TEC backtracking and nascent transcript cleavage is unaffected by the addition of Spt4, Spt5 or the Spt4-Spt5 complex. (e) Coomassie-stained, SDS-PAGE of purified Spt4 and Spt5. Lane M contains size standards labeled in Kda to the left.

to generate a range of TECs with transcripts ranging from +50-58 (Figure 2.2b, lanes 7-11). When TECs<sub>+58</sub> are provided with even low concentrations of ATP, GTP, and UTP, any TECs that backtrack and cleave their transcripts immediately resynthesize to +58 (Figure 2.2b, lanes 2-6). The position of TECs on such templates is thus dynamic, and addition of TFS dramatically stimulated transcript cleavage in backtracked TECs (Figure 2.2b, lanes 12-16). A TFS variant, wherein two conserved acidic residues were replaced with alanines (TFS<sup>DE-AA</sup>), was unable to produce the same cleavage stimulatory effect as TFS<sup>WT</sup> and even slightly impeded RNAP endonuclease activity (Figure 2.2b, lanes 17-21). The inability of TFS<sup>DE-AA</sup> to properly donate acidic residues to the active site of RNAP abrogates its function as a cleavage stimulatory factor.

Backtracking can result from extended pausing and the configuration of mobile-domains of RNAP is known to modulate the propensity to pause and the duration of pausing<sup>2,16,26-28</sup>. We thus examined whether addition of Spt4 and/or Spt5 would influence the efficiency of RNA cleavage, with transcript cleavage also serving as a proxy for the propensity to, and depth of backtracking. In contrast to the stimulated RNA cleavage observed with TFS<sup>WT</sup>, the addition of Spt4 and/or Spt5 did not influence RNA cleavage nor affect the rate, depth, or propensity for backtracking (Figure 2.2d).

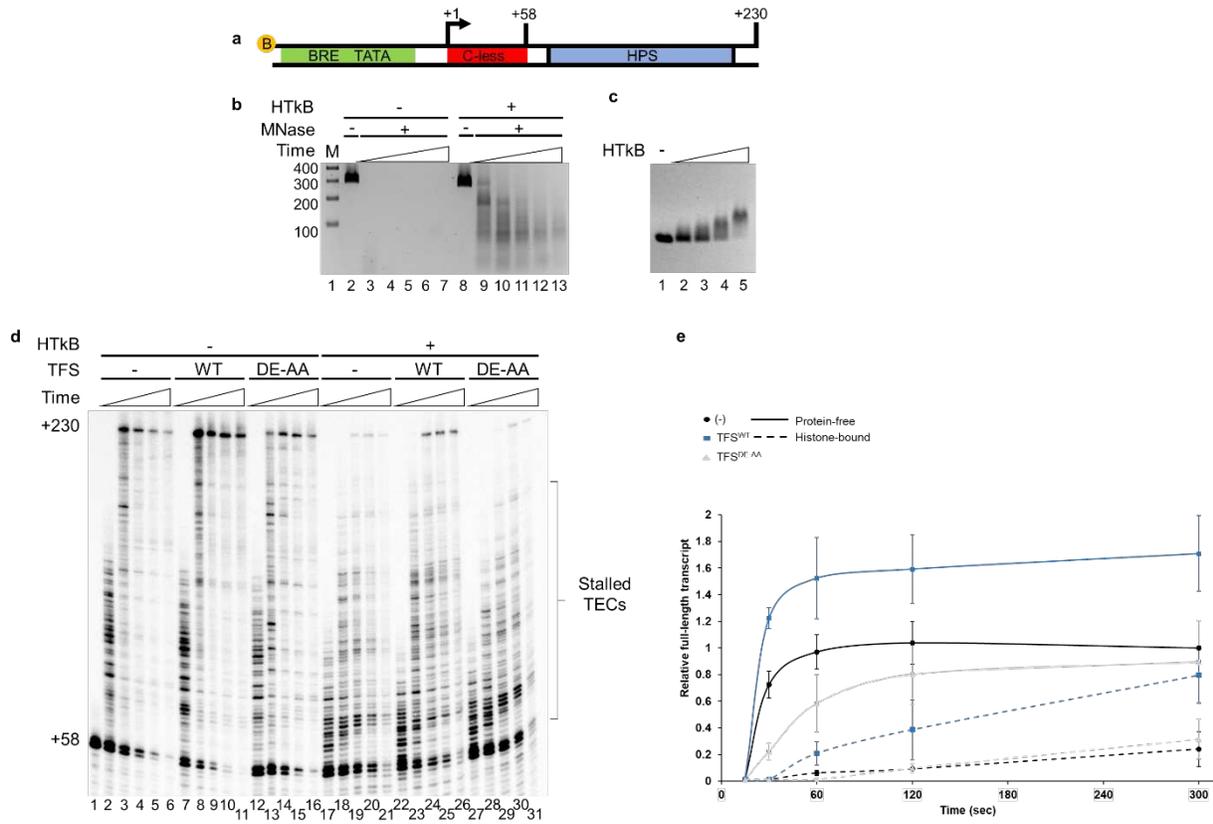
### ***Archaeal transcription is impeded by histone-based chromatin.***

Archaeal chromatin can be formed with a single histone protein and archaeal histones spontaneously bind DNA to assemble chromatin in vitro at the same positions and density as in vivo<sup>10,29</sup>. These attributes, and the simplicity of the archaeal transcription apparatus, permitted us to first assemble washed, promoter-proximal, nucleotide-deprived, stalled TECs (one TEC per template) on DNA templates that could

then be spontaneously bound by archaeal histones to assemble chromatin structures immediately downstream of the TECs<sup>8</sup>. Histone-based chromatin structures were assembled and positioned by incorporating two, 60-bp tandemly repeated, SELEX-derived, optimized archaeal histone-positioning sequences (HPSs) downstream of the stalled TECs (Figure 2.3a)<sup>6,8</sup>. The HPSs supported the spontaneous and consistent binding of histones that hindered MNase digestion and resulted in characteristic protection patterns derived from archaeal chromatin (Figure 2.3b). Histone-binding was monitored via electro-mobility shift assays (Figure 2.3c), and at saturating histone concentrations (>1 histone dimer per 30 bp), the DNA templates were completely histone bound.

By first generating and isolating TECs<sub>+58</sub> on templates that were homogeneously histone-bound or remain protein-free, the rates of elongation on each template upon NTP addition were determined (Figure 2.3d). Transcription rapidly restarts regardless of the presence/absence of downstream protein-barriers and the population of TECs quickly becomes nonsynchronous; template positions that hinder elongation and direct pausing are evident on protein-free templates, but the bulk of TECs reach the end of the template to generate +230 nt transcripts without substantial delay.

Addition of either histone from *T. kodakarensis* (Histone A = TK1413, HTkA; Histone B = TK2289, HTkB) results in dramatically altered elongation patterns and rates of transcription. The overall elongation rate and resultant pace of RNA synthesis from archaea TECs are dramatically reduced when DNA is bound by archaeal histones. Compared to transcription on protein-free templates, chromatin blocked ~80% of TECs from generating full-length transcripts, and the blocked TECs often arrested or paused



**Figure 2.3 TFS increases the rate of elongation and full-length transcript production on protein-free and histone-bound templates. (a)** TECs<sub>+58</sub> can be generated on biotinylated C-less cassettes, washed, and then incubated with HTkB to generate well-positioned downstream histone barriers. **(b)** Ethidium-bromide stained, agarose electrophoresis demonstrates that HTkB-binding protects DNAs from MNase digestion (lanes 9-13) under conditions where protein-free templates are rapidly degraded (lanes 3-7). Lane M contains DNA size standards in base pairs labeled to the left. **(c)** Native electrophoresis and ethidium bromide staining demonstrate that HTkB binding fully saturates and shifts the DNA templates. **(d)** On both protein-free (lanes 2-16) and histone-bound templates (lanes 17-31), the addition of TFS<sup>WT</sup> but not TFS<sup>DE-AA</sup> accelerates ensemble elongation rates and stimulates production of full-length +230 nt transcripts. **(e)** Quantification of full-length transcript levels (n ≥ 3) in the absence and presence of TFS on protein-free and histone-bound templates demonstrates that TFS accelerates transcription and increases full-length transcript yields. The amount of full-length transcripts at 5 minutes on protein-free templates in the absence of TFS is set to 1.0.

for extended periods at multiple positions. The inhibition of transcription elongation observed for archaeal components almost exactly matches the impediment observed of Pol II transcribing through a well-positioned nucleosome<sup>11,30</sup>. The initial collision results in the greatest obstacle, and when the TEC escapes this initial collision, transcription pauses every ~10-15 bp while the TEC traverses the histone complex. The duration of the initial pause is much greater than subsequent pauses, implying that the rate limiting step to transcription on chromatin templates – in both archaea and eukaryotes – is disrupting the first set of conserved histone-DNA contacts.

***TFS increase the rate of elongation and full-length transcript production on protein-free and histone-bound templates by reactivating backtracked complexes.***

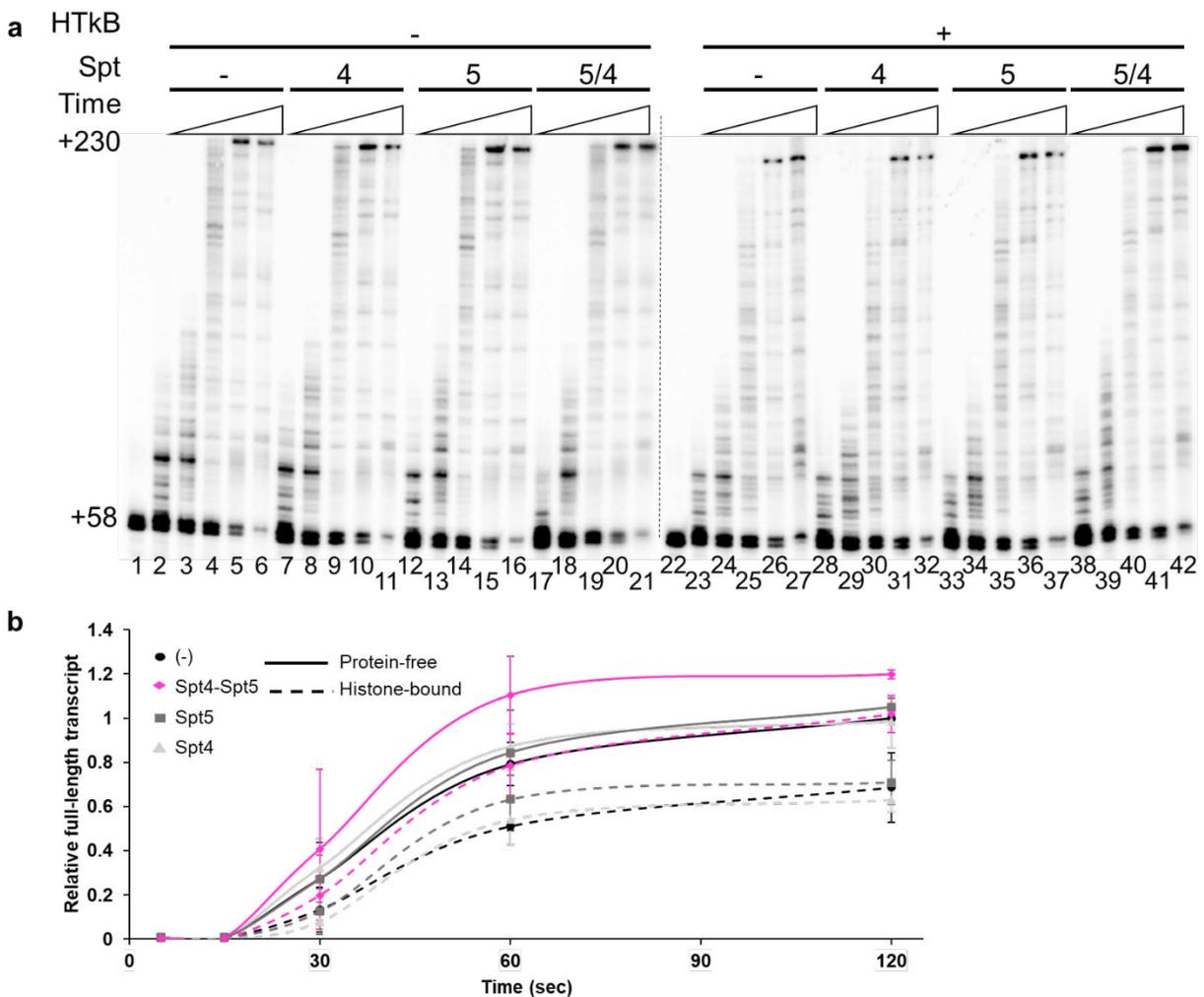
To assess the role TFS and its cleavage-stimulatory activity on the rate and efficiency of transcription through chromatin, purified TFS<sup>WT</sup> and TFS<sup>DE-AA</sup> were added at the time of transcription restart on protein-free and histone-bound templates. On templates lacking chromatin, TFS<sup>WT</sup>, but not TFS<sup>DE-AA</sup> increased the rate and total amount of full-length transcript production (Figure 2.3d, lanes 7-11 & 8-12, respectively). Addition of TFS<sup>WT</sup> reproducibly resulted in nearly-twice the total number of full-length transcripts, and these additional full-length transcripts were generated from TECs that were paused/arrested when TFS was not present or when TFS<sup>DE-AA</sup> was added.

The addition of TFS<sup>WT</sup> to TECs elongating through chromatin dramatically increased both the percentage of full-length transcripts and the rate of full-length transcript production from histone-bound templates. Just ~20% of the full-length transcripts generated on protein-free templates were observed on chromatin templates in the absence of TFS, and the rate of full-length transcript production was reduced by

~20-25-fold. Addition of TFS<sup>WT</sup> restored ~90% of full-length transcript production and increased the ensemble rate of elongation through chromatin ~4-fold (Figure 2.3e). As observed on protein-free DNA, addition of TFS<sup>DE-AA</sup> to TECs transcribing chromatin had no discernable effects, demonstrating that the transcript-cleavage-stimulatory function of TFS is critical for accelerating transcription on histone-bound DNA. The stimulatory activities of TFS<sup>WT</sup> on protein-free and histone-bound DNA suggest that RNAP often pauses - in response to DNA sequence motifs and proteinaceous-roadblocks – long enough to backtrack. These backtracking events hinder elongation, and that cleavage and synthesis cycles more rapidly recover active TECs than spontaneous isomerization to restore the RNA 3' end in the active center of RNAP.

***Spt5 and Spt4 together, but not individually, facilitate elongation through chromatin.***

Spt4-Spt5 binds directly to RNAP and stabilizes a closed-clamp configuration that facilitates elongation and TEC stability<sup>19,23,31,32</sup>. As pausing and subsequent backtracking can be influenced by inter-domain movements of RNAP, we sought to determine whether Spt4-Spt5 binding to RNAP would reduce pausing or accelerate transcription on protein-free and histone-bound templates. Addition of either Spt4 or Spt5 alone (Figure 2.4a, lanes 7-11 & lanes 12-16, respectively; Figure 2.4b) did not alter the ensemble rate of elongation nor the total production of full-length transcripts on protein-free templates. When added together, the Spt4-Spt5 complex (Figure 2.4a, lanes 17-21; Figure 2.4b) resulted in a ~20% increase in total transcript production and a reproducible, but small increase in elongation rate. A ~40% increase in total full-length transcript production and a modest (~1.5-fold) acceleration of elongation rate



**Figure 2.4 The Spt4-Spt5 complex, but neither Spt4 or Spt5 alone, increases the rate of elongation and full-length transcript production on protein-free and histone-bound templates.** (a) TECs<sub>+58</sub> were assembled, washed, and HTkB then added (lanes 23-42) or left out (lanes 2-21) before elongation restart in the presence or absence of Spt4 and/or Spt5. Only the Spt4-Spt5 complex accelerates ensemble elongation rates and stimulates production of full-length +230 nt transcripts. (b) Quantification of full-length transcript levels ( $n \geq 3$ ) in the absence and presence of Spt4 and/or Spt5 on protein-free and histone-bound templates demonstrates that the Spt4-Spt5 complex accelerates transcription and increases full-length transcript yields. The amount of full-length transcripts at 2 minutes on protein-free templates in the absence of either factor is set to 1.0.

were obtained when the Spt4-Spt5 complex was added to TECs transcribing chromatin, but neither factor alone resulted in increased elongation rates or proficiency on histone-bound DNA [Spt4 (Figure 2.4a, lanes 29-33) or Spt5 (Figure 2.4a, lanes 34-38)]. Thus, although not as robust as the improvements observed based on addition of TFS<sup>WT</sup>, the conserved Spt4-Spt5 complex does aid in transcription elongation rates and efficiencies on both protein-free and histone-bound DNAs (Figure 2.4b).

### 2.3 Discussion

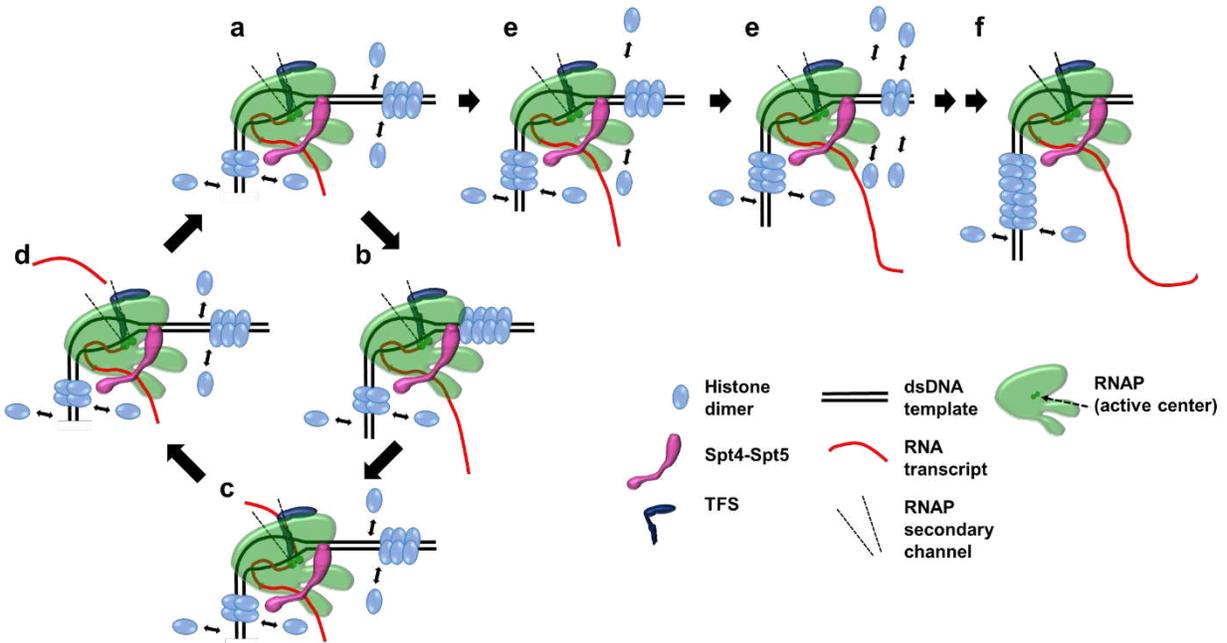
The activities of all macromolecular complexes that interact with DNA must access DNA in a nucleoid- or chromatin-context, and for processive assemblies – such as replication or transcription complexes – the obstacles presented by DNA-bound proteins can dramatically alter elongation rates. Eukaryotic- and most archaeal- genomes are bound by histone proteins, and in *T. kodakarensis*, histone protein concentrations are sufficiently high to bind the entirety of the genomes<sup>33</sup>. The replicative MCM helicase is sufficient powerful enough to easily unwind DNA bound by archaeal histones<sup>34</sup>. In contrast, the rate of transcription on archaeal histone-bound DNAs is dramatically slowed compared to protein-free transcription elongation.

While eukaryotic genomes encode a wealth of factors to both perturb and strengthen the interactions of histones with DNA, the main barrier to the expression of genes is the core histone fold and its interactions with DNA<sup>3,11,20,35–37</sup>. No histone modification machinery has been identified in Archaea, nor has evidence of post-translational modification of archaeal histones emerged<sup>12,14,15,38</sup>. The ability of archaeal histones to form hetero- or homo-dimeric pairings suggests there may be a yet

undiscovered method of regulation in the third domain. Our ability to assemble transcription elongation complexes with or without histones provides a quantitative assay to characterize potential novel and conserved archaeal factors that may accelerate transcription through histone-bound DNA.

Archaeal genomes typically retain only two known and conserved transcription factors – TFS and Spt4-Spt5 – which are known to function through different mechanisms to promote transcription elongation<sup>13,39</sup>. We demonstrate that addition of TFS stimulates the endonucleolytic cleavage activity of RNAP, and that such stimulation is critical to accelerate transcription elongation through the normally very restrictive barrier presented by assemblages of archaeal histones bound to DNA. Addition of TFS triples the ensemble elongation rate and nearly completely restores production of full-length transcripts on histone-bound templates. The ability of TFS to stimulate cleavage of nascent transcripts is critical for such stimulation, as a TFS variant that does not contribute acidic residues to the RNAP active site does not alter elongation kinetics or full-length transcript production. Spt4-Spt5, when combined, also modestly increase the rate of elongation and promote production of full-length transcripts on both protein-free and histone-bound DNAs. Neither Spt4 nor Spt5 alone influences RNA cleavage, and only together can Spt4-Spt5 stabilize RNAP such that intrinsic and histone-induced pausing is reduced.

We present a model of transcription elongation through chromatin templates that incorporates the known activities of TFS, Spt4, and Spt5 to explain the mechanistic features that facilitate transcription elongation through a chromatin barrier (Figure 2.5). Our experimental results support a model wherein repeated rounds of backtracking,



**Figure 2.5 Model for transcription-factor aided elongation through archaeal histone-based chromatin.** (a) TFS and Spt4-Spt5 are associated with RNAP, while the downstream histone-based chromatin landscape is dynamic. (b) Collision with a downstream histone-DNA barrier results in RNAP pausing. (c) Extended pausing results in RNAP backtracking and movement of the RNA 3'-end from the active center to the secondary channel. (d) TFS stimulated endonucleolytic transcript cleavage by RNAP generates a new 3' OH in the RNAP active site. (e) In repeated rounds of synthesis and cleavage the histone-based chromatin landscape has shifted allowing further progression by RNAP. (f) Spt4-Spt5 likely reduce pausing while traversing histone-bound DNA by stabilizing the closed-clamp configuration that may stimulate forward translocation. (g) The combinatorial activities of TFS and Spt4-Spt5 allow RNAP to transcribe the full-length of a DNA template through a shifting histone landscape.

cleavage, and elongation permit TECs to continually approach the histone-imposed barrier leading to opportunistic traversal of the chromatin barrier. When the chromatin barrier is sufficiently dynamic, the TEC – in complex with TFS and Spt4-Spt5 – can displace histone proteins and continue elongation to the end of the template (Figure 2.5, panels a-f). Uninterrupted elongation is uncommon, and most TECs pause (Figure 2.5b) upon encountering a stable downstream histone-barrier. Extended pausing results in backtracking (Figure 2.5c), and the depth of which may be limited by any upstream chromatin landscape. TFS-stimulated cleavage of the nascent transcript (Figure 2.5d) reestablishes a 3' OH in the RNAP active center. Due to the spontaneous nature of archaeal histone interactions with DNA, both upstream and downstream segments of DNA are histone-bound. Most TECs will go through multiple rounds of elongation, pausing, backtracking, cleavage and synthesis to traverse a chromatin template. The activities of the Spt4-Spt5 complex decrease pausing likely by stabilizing a closed clamp configuration, limiting jaw and clamp movements, and through interactions with the non-template strand. The totality of these interactions increases RNAP processivity through the remaining barrier while not interfering with the backtracking rescue activity of TFS.

Our results demonstrate the importance of the universally conserved protein Spt5 and the conserved activity of TFS in modulating various aspects of RNAP activity to overcome both sequence-induced pauses and histone-induced barriers. The ability of TFS (or Gre factors in Bacteria) to restore RNAP from a backtracked position, and Spt4-Spt5 interactions to maintain a closed-clamp, processive confirmation of RNAP likely stem from LUCA and remain in all extant organisms to facilitate transcription elongation through protein-based roadblocks. The evolution of histone-tails and extensions, and

post-translational histone-modifications likely further impacts transcription elongation and it is plausible that these eukaryotic chromatin modifications necessitated the evolution of eukaryotic chromatin-remodeling complexes.

## **2.4 Materials and Methods**

### ***Protein Purifications.***

RNAP, TBP, TFB, HTkA and HTkB were purified as described previously<sup>25</sup>. Spt4, Spt5, TFS<sup>WT</sup>, and TFS<sup>DE-AA</sup> were purified from Rosetta2 *E. coli* cells (Millipore Sigma) containing modified pQE-80L expression vectors (Qiagen) containing His<sub>6</sub>-Spt4-, Spt5-, TFS<sup>WT</sup>- or TFS<sup>DE-AA</sup> encoding sequences, respectively, grown in LB medium with 30 µg/mL chloramphenicol and 100 µg/mL ampicillin. Expression of each protein was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside, and cultures were grown for an additional 3 h at 37°C with shaking (~225 rpm). Biomass was harvested, resuspended and lysed via sonication (3 mL/g of pellet) in lysis buffer (25 mM Tris-HCl pH 8.0, 50 mM NaCl). Cellular lysates were clarified by centrifugation, (~20,000 x g, 20 min, 4°C). TFS proteins were partially purified by heating the clarified cell lysates to 85° C for 30 min, followed by passage and fractionation of cleared supernatants through a cellulose phosphate column. The column was equilibrated in 25 mM Tris-HCl pH 8.0, 50 mM NaCl, and resolved with a linear gradient of 50 mM – 1 M NaCl in 25 mM Tris-HCl pH 8.0. Fractions containing TFS were identified by SDS-PAGE, pooled, dialyzed into storage buffer (25mM Tris-HCl pH 8.0, 100mM KCl, 10mM β-ME, and 50% Glycerol) and quantified using a Bradford Assay. His<sub>6</sub>-Spt4 was partially purified by heating clarified cell lysate at 85° C for 30 min, followed by passage and fractionation of the cleared supernatant through a 1 mL Hi-TRAP chelating column (GE Healthcare)

preequilibrated with NiSO<sub>4</sub>. The column was washed in 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, and 10% glycerol and resolved with a linear imidazole gradient to 60% 25 mM Tris-HCl pH 8.0, 500 mM imidazole, 100 mM NaCl and 10% glycerol. Spt4 containing fractions were identified by SDS-PAGE, pooled, and dialyzed into storage buffer (25 mM Tris-HCl pH 8.0, 100 mM KCl, 10 mM β-ME, and 50% glycerol). Spt5 was partially purified by heating clarified cell lysate at 85° C for 30 min, followed by passage and fractionation of cleared supernatant through a S-100 size exclusion column (GE Healthcare) equilibrated in 25 mM Tris-HCl pH 7.4, 200 mM NaCl. Spt5 containing fractions were identified by SDS-PAGE, pooled, and dialyzed into storage buffer (25 mM Tris-HCl pH 8.0, 100 mM KCl, 10 mM β-ME, and 50% glycerol).

#### ***DNA template construction.***

Double-stranded biotinylated DNA templates used in transcription reactions were PCR amplified from plasmids and gel purified as described<sup>25</sup>.

#### ***In vitro transcription reactions.***

Assembly of preinitiation complexes (PICs) and elongation via NTP deprivation was carried out as described previously<sup>6,25,40</sup>. To obtain stalled TEC<sub>+58</sub>, PICs were incubated with 200 μM ATP, 200 μM GTP, 10 μM UTP and 10 μCi [α-<sup>32</sup>P]-UTP for 3 min at 85°C, then chilled to 4°C and biotinylated templates were captured with streptavidin coated paramagnetic particles (Promega). TECs<sub>+58</sub> were thrice washed in 100 μl WB (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 M KCl) then resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, and ± 10 μM ATP/GTP/UTP. For reactions on protein-free templates, aliquots of washed TECs were combined with equal volume reactions containing 10 mM Tris-HCl pH 8.0, 1 mM MgCl<sub>2</sub>, 120 mM KCl, 8 mM

DTT,  $\pm$  Spt4,  $\pm$  Spt5 or  $\pm$  Spt4-Spt5, -or-  $\pm$  TFS<sup>WT</sup> or  $\pm$  TFS<sup>DE-AA</sup>. To ensure sufficient saturation of TECs, TFS was added at  $\sim$ 9  $\mu$ M, whereas Spt4 and/or Spt5 were added at  $\sim$ 6  $\mu$ M each. Elongation was reinitiated at 85°C by the addition of 25  $\mu$ M NTPs and reaction aliquots were removed over a time course into 5 volumes of 1.2X STOP buffer (0.6 M Tris-HCl pH 8.0, 12 mM EDTA). Radiolabeled transcripts were recovered by addition of 7  $\mu$ g of tRNA (total), equal volume phenol/chloroform/isoamyl alcohol (25:24:1, by volumes) extractions, and precipitations of the aqueous phase with 2.6 volumes 100% ethanol. Precipitated transcripts were resuspended in 95% formamide, 1X TBE, heated to 99°C for 5 minutes, rapidly chilled on ice, loaded and resolved in 10-20% polyacrylamide, 8M urea, 1X TBE denaturing gels. Radiolabeled RNA was detected using phosphorimaging (GE Healthcare). Gel images were analyzed using GE Imagequant 5.2 software.

To obtain TECs on histone-bound templates, TEC<sub>+58</sub> were generated and captured as described above. TEC<sub>+58</sub> were then resuspended in 20 mM Tris-HCl pH 8.0, 100 mM KCl, 4 mM MgCl<sub>2</sub>, 3 mM DTT, 10  $\mu$ M each of ATP, GTP, and UTP, and saturating ( $>$ 1 histone dimer per 30 bp) amounts of HTkB for 30 min at 4°C. Reactions were incubated at 85°C for 2 minutes prior to transcription restart by addition of 25  $\mu$ M NTPs (for Figure 2.3) or 100  $\mu$ M NTPs (for Figure 2.4)  $\pm$  Spt4,  $\pm$  Spt5 or  $\pm$  Spt4-Spt5, -or-  $\pm$  TFS<sup>WT</sup> or  $\pm$  TFS<sup>DE-AA</sup>. Reactions were processed as above.

### ***Western Blot Analysis and Histone Quantification.***

HTkA was used as an antigen to prepare polyclonal antibodies in rabbits (Cocalico Biologicals). Known amounts of HTkA and HTkB were loaded into gels as comparative quantification standards in adjacent lanes to DNaseI-treated clarified cell

lysates. Proteins were separated via SDS-PAGE, transferred to PVDF membranes, and probed with primary anti-HTkA antibodies. Addition of an IgG-HRP conjugated anti-rabbit secondary antibody allowed for detection by chemiluminescent ECL western blotting substrate (Thermo Fisher Scientific). A linear regression of HTkA and HTkB signal intensity to HTkA or HTkB amount in ng was generated. Steady-state histone levels were sufficient to coat >85% percent of the genomes (assuming 30 bp/histone-dimer) even when 19 genomes were assumed per cell. *T. kodakarensis* carries between ~7-19 genomes (13 genomes per cell on average).

## REFERENCES

1. Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019). doi:10.1111/mmi.14191
2. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7090–7095 (2000).
3. Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-I. & Kurumizaka, H. Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595–598 (2018).
4. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
5. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–8 (2008).
6. Xie, Y. & Reeve, J. N. Transcription by an Archaeal RNA Polymerase Is Slowed but Not Blocked by an Archaeal Nucleosome. *J. Bacteriol.* **186**, 3492–3498 (2004).
7. Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C. T., Munson, K., Core, L. J. & Lis, J. T. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-

- seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
8. Sandman, K., Soares, D. & Reeve, J. N. Molecular components of the archaeal nucleosome. *Biochimie* **83**, 277–81 (2001).
  9. Sandman, K. & Reeve, J. N. Chromosome packaging by archaeal histones. *Adv. Appl. Microbiol.* **50**, 75–99 (2001).
  10. Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.-P., Reeve, J. N., Santangelo, T. J. & Fondufe-Mittendorf, Y. N. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* **14**, 391 (2013).
  11. Chang, C. H. & Luse, D. S. The H3/H4 tetramer blocks transcript elongation by RNA polymerase II in vitro. *J. Biol. Chem.* **272**, 23427–34 (1997).
  12. Sanders, T. J., Marshall, C. J. & Santangelo, T. J. The Role of Archaeal Chromatin in Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.05.006
  13. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Publ. Gr.* **9**, 85–98 (2011).
  14. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.06.020
  15. Peeters, E., Driessen, R. P. C., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* **13**, 333–41 (2015).
  16. Hausner, W., Lange, U. & Musfeldt, M. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**, 12393–9

- (2000).
17. Lange, U. & Hausner, W. Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol. Microbiol.* **52**, 1133–1143 (2004).
  18. Fouqueau, T., Blombach, F., Hartman, R., Cheung, A. C. M., Young, M. J. & Werner, F. The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nat. Commun.* **8**, 1914 (2017).
  19. Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M. & Sekine, S. I. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science (80-. )*. **357**, 921–924 (2017).
  20. Farnung, L., Vos, S. M. & Cramer, P. Structure of transcribing RNA polymerase II-nucleosome complex. *Nat. Commun.* **9**, (2018).
  21. Schulz, S., Gietl, A., Smollett, K., Tinnefeld, P., Werner, F. & Grohmann, D. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl. Acad. Sci.* **113**, E1816–E1825 (2016).
  22. Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K. C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J. & Cramer, P. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* **66**, 38-49.e6 (2017).
  23. Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J. & Werner, F. The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol. Cell* **43**, 263–274 (2011).

24. Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P. & Werner, F. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
25. Gehring, A. M. & Santangelo, T. J. in *Methods Mol. Biol.* **1276**, 263–279 (2015).
26. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334 (2003).
27. Toulmé, F., Mosrin-Huaman, C., Sparkowski, J., Das, A., Leng, M. & Rachid Rahmouni, A. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* **19**, 6853–6859 (2000).
28. Ray-Soni, A., Mooney, R. A. & Landick, R. Trigger loop dynamics can explain stimulation of intrinsic termination by bacterial RNA polymerase without terminator hairpin contact. *Proc. Natl. Acad. Sci.* **114**, E9233–E9242 (2017).
29. Mattioli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., Byrne, K. R., Lee, T., Ahn, N. G., Santangelo, T. J., Reeve, J. N. & Luger, K. Structure of histone-based chromatin in Archaea. *Science (80-. ).* **357**, 609–612 (2017).
30. Luse, D. S., Spangler, L. C. & Újvári, A. Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors. *J. Biol. Chem.* **286**, 6040–8 (2011).
31. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription

- processivity. *EMBO J.* **30**, 1302–1310 (2011).
32. Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* **2**, 17021 (2017).
  33. Spaans, S. K., van der Oost, J. & Kengen, S. W. M. The chromosome copy number of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *Extremophiles* **19**, 741–750 (2015).
  34. Shin, J. H., Santangelo, T. J., Xie, Y., Reeve, J. N. & Kelman, Z. Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. *J. Biol. Chem.* **282**, 4908–4915 (2007).
  35. Chang, H. W., Pandey, M., Kulaeva, O. I., Patel, S. S. & Studitsky, V. M. Overcoming a nucleosomal barrier to replication. *Sci. Adv.* **2**, (2016).
  36. Kireeva, M. L., Hancock, B., Cremona, G. H., Walter, W., Studitsky, V. M. & Kashlev, M. Nature of the nucleosomal barrier to RNA polymerase II. *Mol. Cell* **18**, 97–108 (2005).
  37. Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y. Y., Lubkowska, L., Kashlev, M. & Bustamante, C. Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151**, 738–749 (2012).
  38. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
  39. Werner, F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* **16**, 247–250 (2008).

40. Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci.* **114**, E6767–E6773 (2017).

## CHAPTER 3: EXTENDED ARCHAEAL HISTONE-BASED CHROMATIN STRUCTURE REGULATES GENE EXPRESSION IN *THERMOCOCCUS KODAKARENSIS*

### 3.1 Summary

Histone proteins compact and organize DNA resulting in a dynamic chromatin architecture impacting DNA accessibility and ultimately gene expression<sup>1-11</sup>. Eukaryotic chromatin landscapes are structured through epigenetic marks, the activities of chromatin-remodeling complexes, and post-translational modification of histone proteins<sup>12-14</sup>. In most Archaea, histone-based chromatin structure is dominated by the helical polymerization of histone proteins wrapping DNA into a repetitive and closely-gyred configuration<sup>15,16</sup>. The formation of the archaeal-histone chromatin-superhelix is a regulatory force of adaptive gene expression and is likely critical for regulation of gene expression in all histone-encoding Archaea<sup>15,17</sup>. Single amino acid substitutions in archaeal histones that block formation of tightly packed chromatin structures have profound effects on cellular fitness, but the underlying gene expression changes resultant from an altered chromatin landscape have not been resolved. Using the model organism *Thermococcus kodakarensis*, we genetically alter the chromatin landscape and quantify the resultant changes in gene expression, including unanticipated and significant impacts on provirus transcription<sup>18-21</sup>. Global transcriptome changes reveal the regulatory importance of higher-order histone-based chromatin architectures.

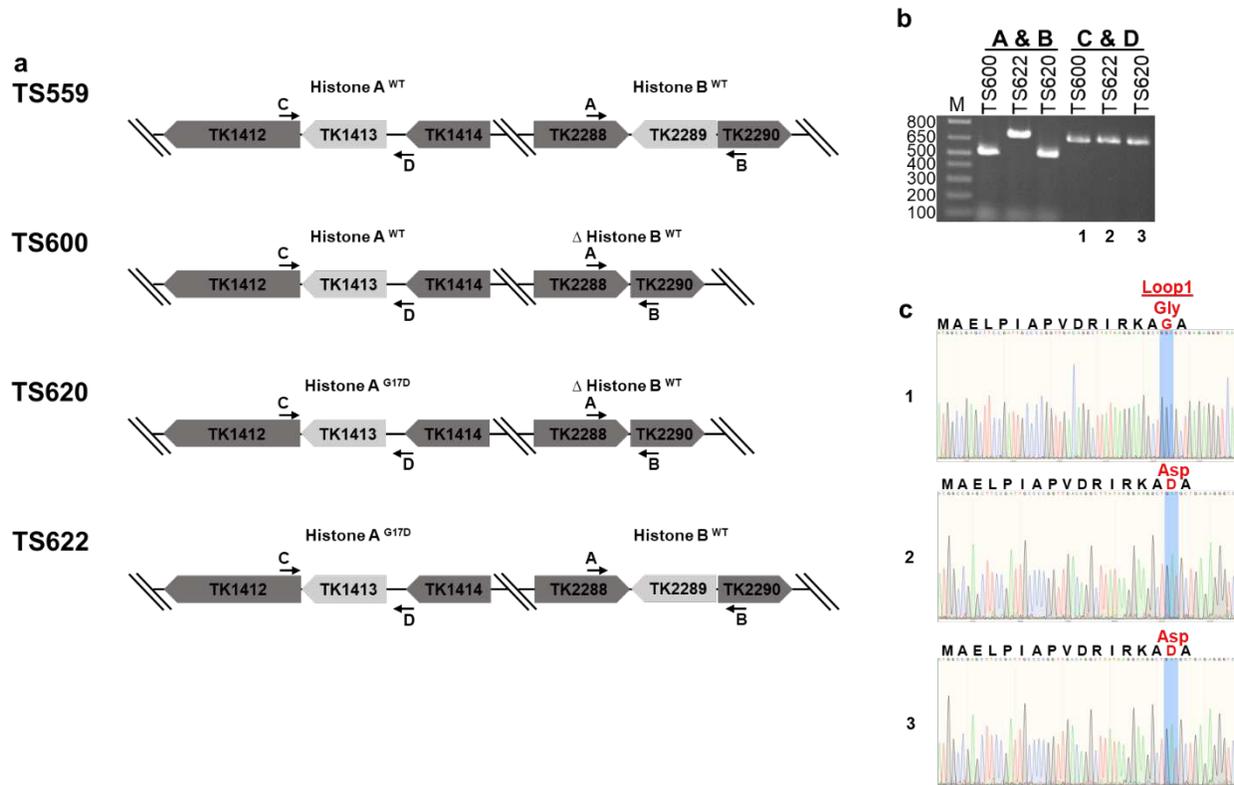
### 3.2 Results

#### ***A single histone protein is sufficient for superhelix formation.***

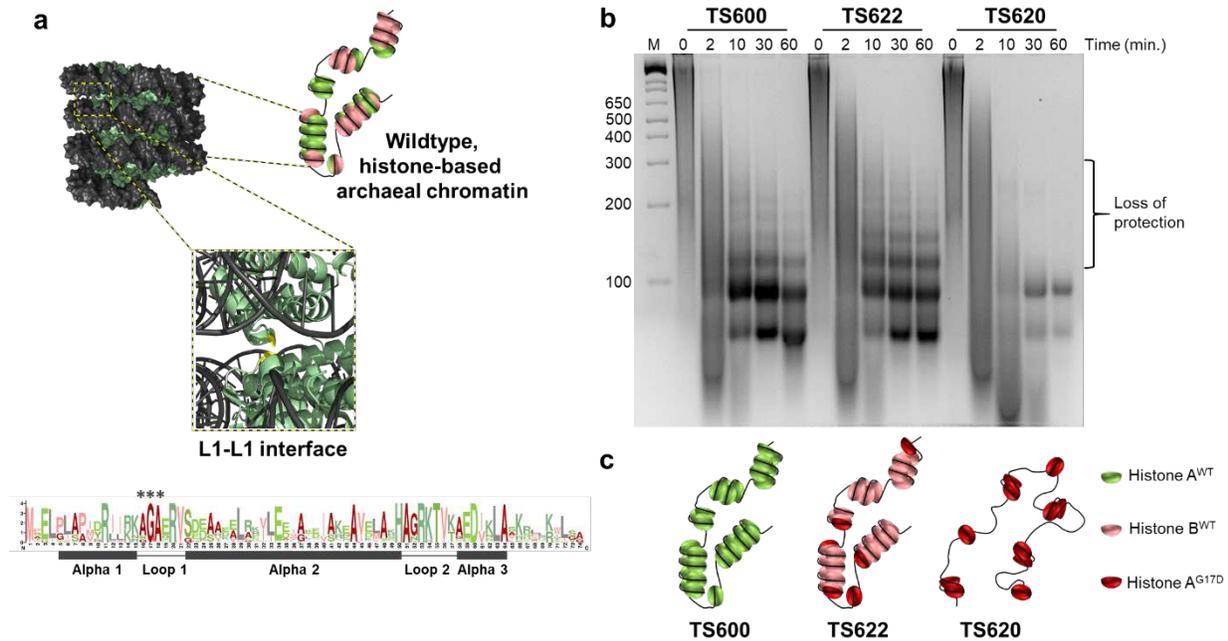
To assess the role of superhelix formation and histone isoforms within the superhelix, we generated *T. kodakarensis* strains that encode both or only a single

histone isoform in WT or variant form (Figure 3.1)<sup>22-25</sup>. Markerless modification or deletion of the genes encoding the endogenous histone proteins, HTkA (TK1413) and HTkB (TK2289) in their natural context, allowed for preservation of the native promoter elements and did not impact the surrounding loci. Strain TS559 served as the parental strain for construction of three strains with varied histone compositions. Deletion of TK2289 (encoding HTkB) resulted in strain TS600 (encoding only HTkA<sup>WT</sup>), whereas modification of TK1413 to generate a G17D variant of HTkA in an otherwise native background generated strain TS622. TS620 combines both genetic modifications, thereby generating a strain supported by only the synthesis of a single, G17D HTkA variant. Deletion of HTkB was confirmed in TS600 and TS620 by PCR amplification of the surrounding region and sequencing to confirm the exact desired endpoints of the genomic modification. (Figure 3.1b). HTkA<sup>G17D</sup> variants were confirmed in TS620 and TS622 by amplifying the entire HTkA coding and promoter sequences and subsequent Sanger sequencing to confirm retention of all native regulatory and coding sequences with the exception of the desired missense mutation (Figure 3.1b and 3.1c).

To assess the impacts of varied histone isoforms on chromatin superhelix formation, total chromatin purified from strains TS600, TS620 and TS622 was subjected to micrococcal nuclease (MNase) digestion. MNase digestion provides a rapid genome-wide mechanism to define minimal units of chromatin - such as the nucleosome in Eukarya<sup>4</sup>. Digestion of chromatin isolated from TS600 confirmed the previously observed wildtype archaeal histone-based protection patterning – a prominent 60 bp DNA fragment along with larger DNA fragments in increasing 30 bp increments (up to ~300 bp) – indicative of varied-length chromatin superstructures (Figure 3.2b and 3.2c).



**Figure 3.1 Only one histone protein is required in *T. kodakarensis*.** (a) Genomic loci diagrams for histone A (TK1413) and histone B (TK2289) and the modifications to each in strains TS600, TS620, and TS622 compared to parental strain TS559. (b) Amplification of TK1413 and TK2289 using the primers diagramed in (a). The open reading frame for TK2289 has been markerlessly deleted in both TS600 and TS620. Amplicons for TK1413 (lanes 1, 2, and 3) were sequenced and (c) confirmed to contain either a glycine codon at position 17 (TS600) or aspartic acid (TS620 and TS622).



**Figure 3.2 A single wild-type histone protein is sufficient for normal DNA protection.**

(a) Diagrammatic representation of wildtype chromatin modeled from the archaeal histone-based chromatin crystal structure: 9 polymerized histone B dimers (pale green) wrapped by DNA (grey) adapted from Mattioli et al.<sup>5</sup>. The central glycine in the AGA motif at the Loop1-Loop1 interface is colored in red. A Logo-plot highlights the conservation of this motif. Histone dimers may be heterogeneously composed. (b) DNA fragments resulting from micrococcal nuclease (MNase) digested chromatin demonstrate the state of chromatin structure in TS600, TS622, and TS620. Chromatin purified from TS600 (TK1413<sup>WT</sup>:histone A/ΔTK2289:Δhistone B) resists MNase digestion over time, resulting in a laddered DNA banding pattern. Prominent 60bp and 90bp bands in addition to higher molecular weight bands (increasing 30bp increments up to ~300bp) represent varying levels of histone dimerization and MNase protection. Chromatin purified from TS622 (TK1413<sup>WT</sup>:histone A/TK2289<sup>WT</sup>:histone B) exhibits an identical protection pattern to TS600 despite encoding a variant (G17D) histone A. This suggests a single WT histone is sufficient for normal chromatin structure formation. Chromatin purified from TS620 (TK1413<sup>G17D</sup>:histone A/ΔTK2289:Δhistone B) exhibits a markedly different protection pattern from TS600 and TS622. The presence of only a variant (G17D) histone A results in a loss of DNAs protected above 90bp, demonstrating the disruption of the L1-L1 interface interferes with continued histone dimer polymerization. (c) Diagrammatic representation of the potential chromatin structures in TS600, TS622, and TS620.

The persistence of the distinct  $60 + 30(n)$  ladder demonstrates that chromatin superhelix structures are varied in length but stable, and thus provide architectures that can be exploited to regulate DNA accessibility and gene expression. TS600 encodes only the HTkA isoform therefore all histone-based chromatin structure is composed of this single histone (Figure 3.2c).

In contrast, when HTkA was modified to place a larger and charged residue within L1 (G17D), the chromatin from strain TS620 displayed a dramatically different MNase protection pattern (Figure 3.2b). Discrete DNA fragments  $>90$  bp were absent, demonstrating histone:DNA interactions occurred but not tight-wrapping or continued polymerization of the archaeal histone-based chromatin superhelix. Further, this observed digestion pattern is consistent with previous digestions of HTkA<sup>G17</sup> variants suggesting histone dimers form tetramers, protecting 60bp of DNA, and an additional dimer interacts to form a hexamer, protecting 90bp of DNA, but larger three-dimensional associations of histone dimers are restricted. Thus, across the entire genome, the single HTkA<sup>G17D</sup> variant encoded in TS620 seemingly disrupts the L1-L1 interface within the superhelix, preventing continued polymerization of histone dimers that normally provides a route to superhelix formation and regulatory potential of such.

Surprisingly, digestion of total chromatin purified from TS622 resulted in an MNase protection pattern comparable to TS600 (Figure 3.2b) suggesting the presence of variant HTkA<sup>G17D</sup> did not interfere with superhelix formation of the HTkB isoform. Given that chromatin from strain TS620 demonstrates HTkA<sup>G17D</sup> alone cannot make up the superhelix, it is likely that all of the larger regions of DNA protection in strain TS622 results from superstructures formed entirely of HTkB. It is possible that at least some of

the smaller (60 – 90 bp) protected DNA fragments result from chromatin structures composed entirely of HTkA<sup>G17D</sup> or from HTkB/HTkA<sup>G17D</sup> heteromers; additionally, it is possible HTkA<sup>G17D</sup> may cap or terminate the superhelices composed of HTkB resulting in protection of >90 bp fragments (Figure 3.2b and 3.2c).

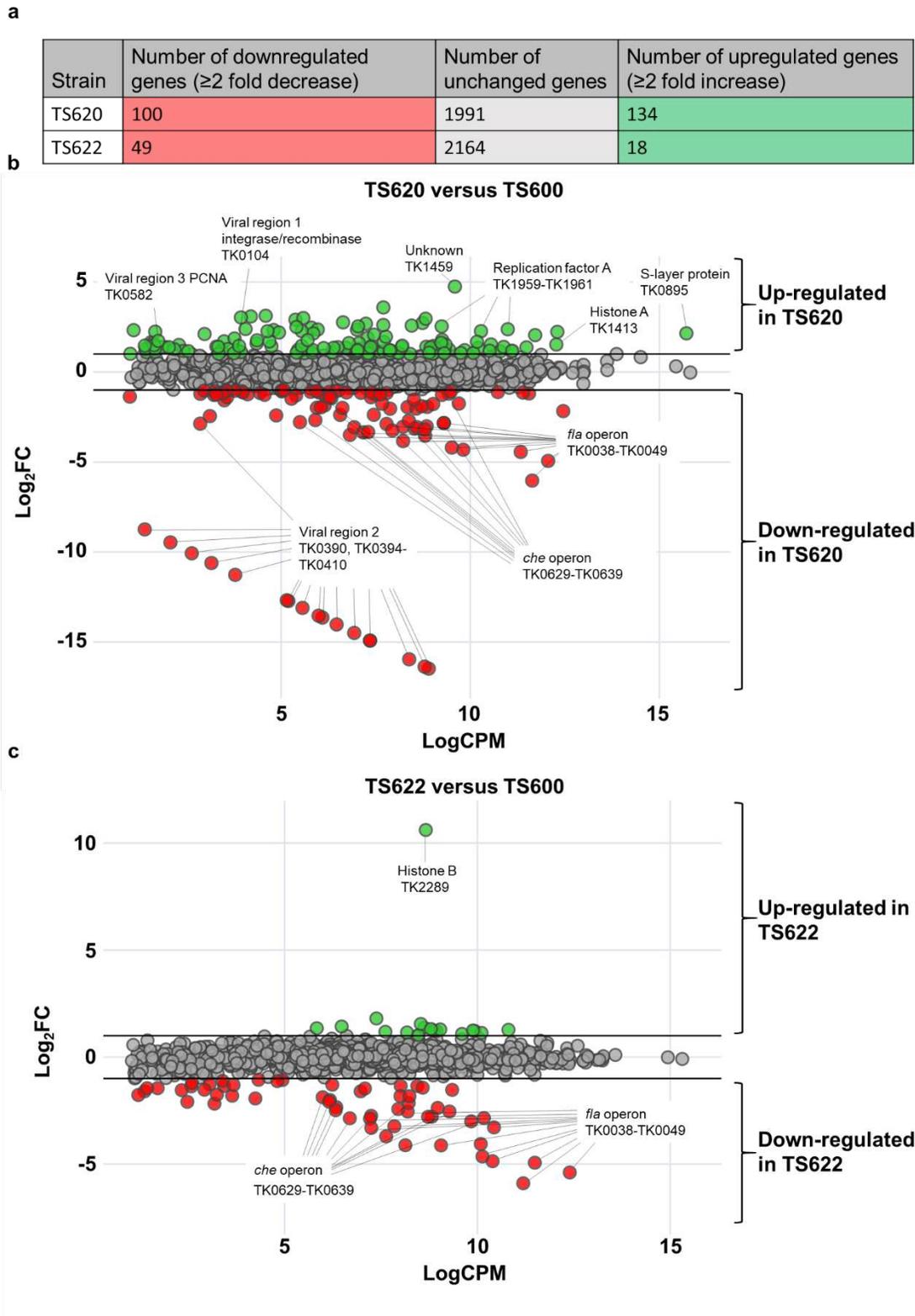
***Superhelical chromatin structures regulate genome-wide gene expression.***

With histone composition altered in strain TS622, and histone-composition and superhelix structural potential disrupted in strain TS620, we sought to quantify the transcriptomes of each strain in response to an environmental shift. Environmentally-cued changes to histone-based chromatin architecture are a known mechanism to regulate gene expression, and for the *Thermococcales*, one of the largest determinants of metabolism and gene expression profiles is the availability of different terminal electron acceptors<sup>17,26</sup>. Strains with variant histone- and chromatin-landscapes were grown to early exponential phase under conditions that permit elemental sulfur (S<sup>0</sup>) to serve as the terminal-electron acceptor, then rapidly transferred to conditions wherein S<sup>0</sup> was absent, signaling a necessary metabolic shift requiring substantial changes in gene expression for continued, rapid growth.

RNAs from each strain were purified from biological triplicate or duplicated cultures following the environmental shift, depleted of ribosomal RNAs and subjected to RNA-sequencing to quantify steady-state transcript abundance. Differential expression analyses defined meaningful changes in the transcriptomes of TS620 and TS622 compared to TS600 (Figure 3.3). Comparison of transcriptomes resultant from TS600 (HTkA<sup>WT</sup> only; typical chromatin superstructures of varied lengths) and TS620 (HTkA<sup>G17D</sup> only; abrogated superhelical chromatin structure) provides a quantitative

measure of the regulation normally afforded by superhelical chromatin structures in *T. kodakarensis* (Figures 3.1c and 3.2). Comparison of the transcriptomes of T600 and TS622 (HTkB<sup>WT</sup>, HTkA<sup>G17D</sup>) highlights the impact of histone variants on chromatin structures that can regularly form with HTkB (Figures 3.1c and 3.2). Transcriptome changes were quantified by comparing log<sub>2</sub>-average fold change (Log<sub>2</sub>FC) against log-average counts per million (LogCPM) (Figure 3.2b and 3.2c). *T. kodakarensis* encodes ~2,300 annotated open reading frames, with abundant antisense transcription and small transcripts. Our sequencing coverage was sufficient to detect nearly all transcripts with a LogCPM > 1.00: 2,225 transcripts were included in the TS620/TS600 comparison; 2,231 transcripts were included in the TS622/TS600 comparison.

The substitution of HTkA<sup>G17D</sup> (TS620) for HTkA<sup>WT</sup> (TS600) as the only histone in *T. kodakarensis* resulted in significant (>2-fold) changes in the steady-state abundance of 234 genes, representing ~11% of the entire transcriptome. The lack of stable archaeal histone-based chromatin superhelical structures in TS620 resulted in increased abundance of 100 genes and decreased abundance of another 134 genes (Figure 3.3a). Interestingly, genes with both ordinarily high or low expression were differentially expressed. Genes-encoding proteins involved in central metabolism, purine synthesis and metabolism, amino acid synthesis, transport and a number of hypothetical proteins were upregulated (Table 1a). No obvious chromosomal distribution was noted, with the obvious exception of coregulation of operons, suggesting the entire chromosome is normally subject to regulation imposed by stable extended archaeal chromatin superstructures. Further, hinting that expression of histone-encoding genes are regulated by archaeal chromatin structures, TK1413<sup>G17D</sup> transcripts were enriched



**Figure 3.3 Altering 3-dimensional chromatin structure dramatically alters gene expression.** (a) Table comparison of strain expression data. (b) Expression profile of TS620 vs TS600. (c) Expression profile of TS622 vs TS600.

**Table 1. Genes enriched in TS620 (a) or depleted in TS620 (b) compared to TS600.**

**a**

Transcript	Annotation	Process	Fold change
TK0038	archaeal flagellin B1 precursor	cell motility	-65.39
TK0039	archaeal flagellin B2 precursor	cell motility	-30.65
TK0040	archaeal flagellin B3 precursor	cell motility	-21.73
TK0042	archaeal flagellin B5 precursor	cell motility	-19.94
TK0041	archaeal flagellin B4 precursor	cell motility	-18.37
TK0043	archaeal flagella-related protein C	cell motility	-14.30
TK0044	archaeal flagella-related protein D, internal insertion	cell motility	-11.59
TK0045	archaeal flagella-related protein F	cell motility	-11.25
TK0046	archaeal flagella-related protein G	cell motility	-10.23
TK0632	chemotaxis response regulator, CheY	environmental information processing	-10.09
TK0631	chemotaxis protein methyltransferase, CheR	environmental information processing	-9.51
TK0634	chemotaxis histidine kinase, CheA	environmental information processing	-9.16
TK0633	chemotaxis protein-glutamate methyltransferase, containing receiver domain	environmental information processing	-8.81
TK0635	chemotaxis histidine kinase	environmental information processing	-8.39
TK0156	methyl-accepting chemotaxis protein	environmental information processing	-8.34
TK2147	methyl-accepting chemotaxis protein	environmental information processing	-8.25
TK0049	archaeal flagella-related membrane protein J	cell motility	-8.09
TK0047	archaeal flagella-related protein H	cell motility	-7.39
TK0630	methyl-accepting chemotaxis protein	environmental information processing	-7.14
TK0048	archaeal flagella-related protein I	cell motility	-7.13
TK0637	chemotaxis protein cheC	environmental information processing	-6.89
TK0050	hypothetical membrane protein	unknown	-6.51
TK0636	chemotaxis protein CheC	environmental information processing	-6.36
TK0168	predicted transcription regulator, Lrp/AsnC family	transcription	-5.47
TK0546	hypothetical protein	unknown	-5.32
TK0638	methyl-accepting chemotaxis protein	environmental information processing	-5.21
TK1139	ATPase, AAA superfamily	unknown	-5.19
TK1804	ABC-type dipeptide/oligopeptide transport system, probable periplasmic component	transport	-4.50
TK1606	methyl-accepting chemotaxis protein	environmental information processing	-4.11
TK1605	hydrolase, metallo-beta-lactamase superfamily	unknown	-4.09

**b**

Transcript	Annotation	Process	Fold Change
TK1459	hypothetical protein	unknown	22.52
TK1358	hypothetical protein	unknown	21.63
TK2061	Sodium/phosphate symporter	transport	12.84
TK0604	hypothetical protein	unknown, viral region 3	12.00
TK0605	hypothetical protein	unknown, viral region 3	9.75
TK0208	phosphoribosylformylglycinamide cyclo-ligase	purine metabolism	9.51
TK0202	phosphoribosylformylglycinamide synthase, PurS component	purine metabolism	9.18
TK0204	phosphoribosylamine-glycine ligase	purine metabolism	8.97
TK1392	NADH oxidase	metabolism	8.74
TK0203	phosphoribosylformylglycinamide cyclo-ligase	purine metabolism	8.43
TK1356	ATPase, AAA superfamily	unknown, viral region 4	8.41
TK0835	phosphoribosylaminoimidazole carboxylase, ATPase subunit	purine metabolism	7.60
TK1393	anaerobic glycerol 3-phosphate dehydrogenase	lipid metabolism	7.40
TK2060	distant homolog of phosphate transport system regulator PhoU	transport	7.20
TK1391	molybdopterin oxidoreductase, 4Fe-4S cluster-binding subunit	central metabolism	6.81
TK1023	hypothetical protein	unknown	6.64
TK0207	format-dependent phosphoribosylglycinamide formyltransferase	purine metabolism	6.42
TK1960	replication factor A complex, RPA14 subunit	replication/recombination/repair	6.42
TK0201	phosphoribosylformylglycinamide synthase I	purine metabolism	6.26
TK1464	hypothetical protein	unknown	6.05
TK0599	hypothetical protein	unknown, viral region 3	5.70
TK1959	replication factor A complex, RPA32 subunit	replication/recombination/repair	5.63
TK0389	hypothetical protein	unknown, viral region 2	5.57
TK0252	indole-3-glycerol phosphate synthase	amino acid synthesis	5.57
TK0836	phosphoribosylaminoimidazole carboxylase, catalytic subunit	purine metabolism	5.50
TK0601	ATPase, AAA superfamily	unknown, viral region 3	5.42
TK0580	hypothetical protein	unknown, viral region 3	5.39
TK0382	hypothetical protein	unknown, viral region 2	5.17
TK1961	replication factor A complex, RPA41 subunit	replication/recombination/repair	5.14
TK0600	hypothetical protein	unknown, viral region 3	5.13

~2.8 fold in TS620 compared to TK1413<sup>WT</sup> in TS600. Altered chromatin landscapes undoubtedly impact DNA replication, recombination and repair, and the loss of larger chromatin superhelices results in the large increase in abundance of all three replication factor A proteins (archaeal RPA is a heterotrimer composed in *T. kodakarensis* of the products of TK1959 (increased ~5.2 fold), TK1960 (increased ~5.8 fold) and TK1961 (increased ~4.8 fold) in strain TS620. The likely increased abundance of functional RPA proteins suggests inhibiting larger chromatin structures may permit DNA regions to locally unwind or melt, thereby requiring more RPA to protect the increased abundance of single-stranded DNA (Figure 3.3b).

While many genes classes showed increases in transcript abundance due to chromatin structural changes, a large percentage of genes, related to cell motility and environmental signal sensing, were downregulated due to the absence of superhelical chromatin structures in TS620 when compared to TS600. Localized extended chromatin structures are likely to regulate gene expression both positively and negatively, depending on the availability of DNA sequences critical for gene expression. Expression of operons encoding archaellum components (annotated as the *fla* operon) and chemotaxis proteins (*che* operon) appear particularly sensitive to chromatin structure and likely dependent on extended chromatin architectures for proper regulation (Figure 3.3b). Genes comprising the entire *T. kodakarensis* archaellum-encoding *fla* operon (TK0038-TK0049) and chemotaxis-encoding *che* operon (TK0629-TK0639) were ~7.1 – 65-fold and ~2.7 – 10-fold less abundant in strain TS620 compared to strain TS600 (Table 1b).

Given that superhelical chromatin structures can be generated in strain TS622 –

as assessed by MNase digestions (Figure 3.1) – we predicted a more minor impact on the total transcriptome of strain TS622 compared to TS600. Despite the presence of HTkA<sup>G17D</sup>, the added presence of HTkB in TS622 confers sufficient superhelical chromatin structure to reduce the number of aberrantly transcribed genes, when compared to TS620. When comparing the transcriptomes of TS622 (HTkA<sup>G17D</sup>, HTkB) and TS600 (HTkA<sup>WT</sup>, ΔHTkB) we noted only approximately half as many transcripts (49) of decreased abundance in TS622, while just 18 transcripts were enriched (Figure 3.3c and Table 2a). Similar decreases in the abundance of transcripts encoding cell motility and environmental signal sensing were observed in TS622, with transcripts from the *fla*- and *che*-operon decreased ~6 – 59-fold and ~2.6 – 9-fold, respectively (Figure 3.3c, Table 2b). The decreased abundance of *fla*- and *che*-operon transcripts suggests the HTkA isoform is critical for proper regulation of these loci or factors that control expression of such loci and that the presence of HTkB<sup>WT</sup> cannot compensate for the loss of HTkA<sup>WT</sup> in these limited scenarios.

***Superhelical chromatin structure is necessary for proviral region expression and retention.***

The *T. kodakarensis* genome contains 4 annotated, non-essential proviral regions, each ~20-25 Kbp in length: TKV1 (TK0073-TK0105), TKV2 (TK0381-TK0421), TKV3 (TK0575-TK0614) and TKV4 (TK1342-TK1378)<sup>18,21,27</sup>. Comparisons of the genomes of many *Thermococcales* suggests recombination events between proviral regions can rearrange the genome context in evolutionary timescales; however, the proviral regions of the *T. kodakarensis* appear genetically stable. These proviral regions represent ~100 Kbp in total (~5% of the 2.08 Mbp genome) and encode a plethora of

**Table 2. Genes enriched in TS622 (a) or depleted in TS622 (b) compared to TS600.**

**a**

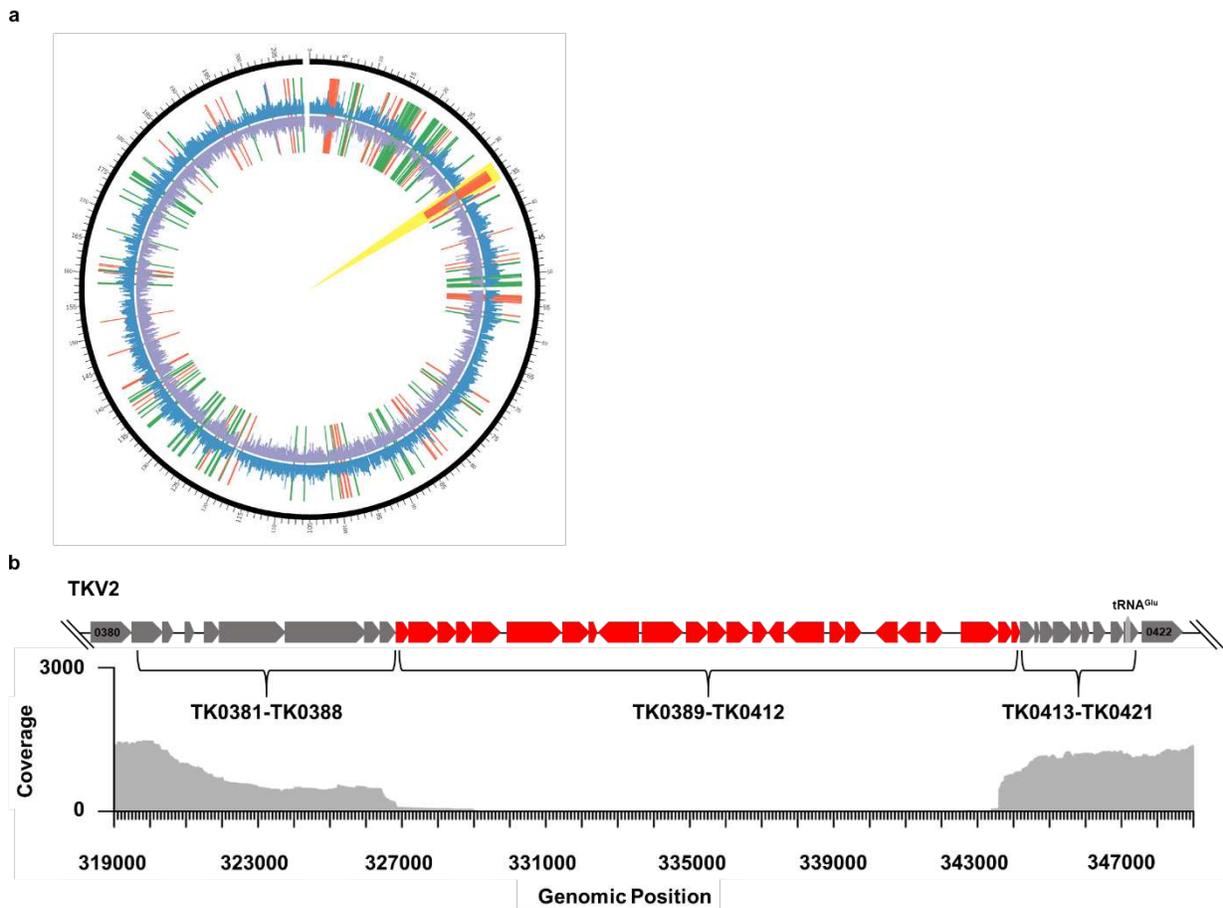
Transcript	Annotation	Process	Fold change
TK2289	Archaeal histone B	chromatin	1565.20
TK1020	hypothetical membrane protein	unknown	3.49
TK0717	molybdate transport system substrate-binding protein	transport	2.93
TK0162	hypothetical membrane protein	unknown	2.69
TK0718	molybdate transport system permease protein	transport	2.58
TK0720	hypothetical protein	unknown	2.55
TK0166	hypothetical protein	unknown	2.48
TK2070	sulphydrogenase subunit delta	energy metabolism	2.45
TK1862	hypothetical protein	unknown	2.45
TK0467	hypothetical protein	unknown	2.42
TK2071	sulphydrogenase subunit gamma (sulfur reductase)	energy metabolism	2.37
TK0719	molybdate transport system ATP-binding protein	transport	2.37
TK2072	sulphydrogenase subunit beta (sulfur reductase)	energy metabolism	2.33
TK0164	S-layer-like array protein	cell structure	2.33
TK0163	ABC-2 type transport system permease protein	transport	2.27
TK1463	hypothetical protein	unknown	2.22
TK2278	myo-inositol-1-phosphate synthase	metabolism	2.19
TK0765	glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	metabolism	2.12
TK2069	sulphydrogenase subunit alpha	energy metabolism	2.04

**b**

Transcript	Annotation	Process	Fold change
TK0038	archaeal flagellin B1 precursor	cell motility	-59.54
TK0039	archaeal flagellin B2 precursor	cell motility	-41.95
TK0040	archaeal flagellin B3 precursor	cell motility	-30.60
TK0042	archaeal flagellin B5 precursor	cell motility	-29.13
TK0041	archaeal flagellin B4 precursor	cell motility	-25.02
TK0043	archaeal flagella-related protein C	cell motility	-17.47
TK0812	adenylate kinase	purine metabolism	-17.38
TK0811	hypothetical protein	unknown	-16.68
TK0631	chemotaxis protein methyltransferase CheR	environmental information processing	-13.03
TK0046	archaeal flagella-related protein G	cell motility	-9.90
TK0044	archaeal flagella-related protein D, internal insertion	cell motility	-9.76
TK0045	archaeal flagella-related protein F	cell motility	-9.40
TK0156	methyl-accepting chemotaxis protein	environmental information processing	-8.01
TK0635	chemotaxis histidine kinase	environmental information processing	-7.36
TK0632	chemotaxis protein CheY	environmental information processing	-7.27
TK0630	methyl-accepting chemotaxis protein	environmental information processing	-7.20
TK0048	archaeal flagella-related protein I, predicted secretion ATPase	cell motility	-6.92
TK0633	chemotaxis protein-glutamate methylesterase, containing receiver domain	environmental information processing	-6.89
TK0047	archaeal flagella-related protein H, predicted ATPase	cell motility	-6.71
TK0049	archaeal flagella-related membrane protein J	cell motility	-5.85
TK0634	sensor kinase CheA	environmental information processing	-5.83
TK0636	chemotaxis protein CheC	environmental information processing	-5.59
TK0050	hypothetical protein	unknown	-5.39
TK2147	methyl-accepting chemotaxis protein	environmental information processing	-5.19
TK0637	chemotaxis protein CheC	environmental information processing	-5.07
TK0431	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase	purine metabolism	-4.51
TK0638	methyl-accepting chemotaxis protein	environmental information processing	-4.43
TK0432	phosphoribosylaminoimidazole-succinocarboxamide synthase	purine metabolism	-4.26
TK1139	ATPase, AAA superfamily	unknown	-4.23
TK0051	protein-L-isoaspartate(D-aspartate) O-methyltransferase	unknown	-4.01

genes with unknown function. Expression of the proviral regions has been observed in all transcriptome studies of *T. kodakarensis*<sup>17,21</sup>. An observed trend in the transcriptomic profiles obtained from strain TS620 was the depletion or enrichment of genes assigned to the *T. kodakarensis* viral regions (Figure 3.3c). Several predicted integrase genes (TK0104 and TK0381) were upregulated in TS620 as well as the nonessential PCNA<sup>27</sup> (TK0582) and a predicted AAA superfamily ATPase (TK0601). Most intriguing, transcripts aligning to a center portion of TKV2 (TK0390, TK0394-TK0410) were completely depleted from TS620. Among these genes were several predicted SpoVT, AbrB transcriptional regulators, and many hypothetical genes (TK0402, TK0405, TK0406, TK0409, and from TKV4, TK1372).

The observed complete absence of these TKV2 transcripts prompted further evaluation of the genome of TS620 (Figure 3.4). Whole-genome sequencing (WGS) of TS620 revealed a relatively large (~15 Kbp) central region of TKV2 was spontaneously excised from the genome (Figure 3.3b). Excision of most, but not all of TKV2 was confirmed by PCR amplifications of loci within and flanking TKV2 sequences in the genome of TS620. This missing region of TKV2 within strain TS620 genomic sequences aligns closely with the observed decreased/lost abundance of TKV2 transcripts (Figure 3.3a). Despite differential expression of portions of other viral regions, the genomic loci for TKVR1, TKVR3, and TKVR4 remain intact. The excision of TKVR2 in only TS620 suggests not only is superhelical structure necessary for regulated gene expression, but also plays a role in genome stability and recombination, perhaps related to viral region retention or repression.



**Figure 3.4 Disruption of 3-dimensional chromatin structure results in genome instability.** (a) A circos plot comparing TS620 to TS600. The outermost black circle represents genomic position. The outer coverage plot (blue) represents Fragments Per Kilobase of transcript per Million mapped reads (FPKM) for TS620. The inner coverage plot (purple) represents FPKM for TS600. Notably, nearly zero reads mapped to TKVR2 in TS620 (highlighted in yellow). Red lines represent fragments enriched in TS600 while green lines represent fragments enriched in TS620. (b) A loci diagram of the annotated *T. kodakarensis* viral region 2 (TKVR2: TK0381-TK0421) that highlights the observed region of excision (~TK0389 - ~TK0412) superimposed over a genome alignment plot derived from PacBio long read sequencing of TS620.

### 3.3 Discussion

Histone proteins encoded in most Archaea and all Eukarya are the primary proteins responsible for genome organization<sup>4,15,28</sup>. Despite geometric and structural similarities, the semi-continuous helical polymerization of archaeal histones is distinct from the eukaryotic nucleosome<sup>15,16,29</sup>. The formation of the archaeal histone-based chromatin superstructure is a major regulatory event in adaptive gene expression in *T. kodakarensis* and the regulation afforded by superhelical histone-based chromatin structures likely extends to most histone-encoding Archaea<sup>1,2,30,31</sup>.

The transcriptomes of archaeal strains that can (TS600 and TS622) and cannot generate histone-based chromatin superstructures (TS620) reveal significant differences most likely associated directly with altered chromatin landscapes. The impacts are often dramatic, and the regulation is not associated with discrete, linear chromosomal regions but instead is noted genome wide. In strain TS620, where formation of stable and detectable (via MNase digestions) extended chromatin structure was inhibited, >11% of the transcriptome was significantly altered. Approximately equal numbers of transcripts were more or less abundant due to the loss of histone-based chromatin superstructures, suggesting that extended chromatin structures can both positively and negatively impact gene expression, depending on the loci under study.

Expression of the *che-* and *fla-*operons, as well as many proviral regions were most dramatically altered by changes in the superhelical chromatin structure. In support of superhelical structures, and not histone abundance influencing gene expression in these operons/proviral regions, expression profiles of *T. kodakarensis* strains encoding WT versions of a single histone isoform, found no significant differences in the

expression of these same operons. The formation of superhelices composed entirely of HTkA or HTkB have been previously shown<sup>22,30</sup> to be sufficient for normal expression suggesting the presence of HTkA<sup>G17D</sup> disrupts any superhelical formation at these loci thereby repressing transcription in both TS620 and TS622. It is also plausible disrupted chromatin structures limit or increase expression of transcription factors that regulated select operon expression, but our transcriptomics data does not identify any obvious candidates.

Perhaps, the most striking dysregulation observed in strains incapable of chromatin superhelix formation (TS620) was the loss of a portion of TKV2 from the genome. Like many proviral integrations in archaeal genomes, all four proviral regions in the *T. kodakarensis* genome overlap with tRNA encoding loci and large rearrangements noted in the genomes of *T. kodakarensis* and related *Thermococcales* often begin and end internal to the proviral regions<sup>19,21,32</sup>. The combined impacts of disrupted chromatin architecture on replication and recombination, the increased abundance of predicted viral integrase transcripts, and the potential for more ssDNA due to DNA melting in strains with altered chromatin landscapes provides a plausible explanation for the loss of TKV2 sequences. The inability to detect DNA sequences encompassing TK0389-0412 or viral particles containing these genes suggests this DNA fragment is degraded and thus is unstable once excised from the genome. Although dispensable, the obvious growth defects of *T. kodakarensis* strains lacking proviral regions suggest their incorporation and proper regulation within archaeal genomes confers an evolutionary advantage.

The retention of histone proteins in most archaeal clades suggest histone-based chromatin structures provide beneficial regulatory roles that are exploited to provide a level of regulation on gene expression. The varied lengths of archaeal histone-based chromatin superstructures, the presence of multiple histone isoforms in many Archaea, and the known changes in histone isoform expression in response to environmental changes all present routes to activate, repress and fine-tune gene expression to maximize growth in changing environments. The mechanisms controlling the formation of chromatin superhelices with varying lengths at different loci and in different sequences context likely plays an additional role in transcription regulation and may be exploited by yet to be discovered archaeal chromatin remodeling complexes or histone isoforms that promote or inhibit formation of archaeal chromatin superstructures.

### **3.4 Materials and Methods**

#### ***Strain construction and growth conditions.***

*T. kodakarensis* strains were constructed as previously described<sup>23–25</sup>. *T. kodakarensis* strain TS600 was constructed from parental strain TS559 by marker-less deletion of TK2289 (HTkB). TS622 was constructed from TS600 by allelic substitution of TK1413 (HTkA) for TK1413<sup>G17D</sup> (HTkA<sup>G17D</sup>). TS620 was constructed by marker-less deletion of TK2289 (HTkB) from TS622. Allelic substitution of TK1413 for TK1413<sup>G17D</sup> was confirmed by PCR amplification of the TK1413 loci and subsequent sequencing while TK2289 deletion was confirmed via PCR. Cultures were grown at 85°C in artificial seawater (ASW) supplemented with 0.5% (w/v) tryptone, 0.5% (w/v) yeast extract (ASW-YT), trace mineral solution and vitamin mixture (nutrient rich medium). ASW contains, per l, 20 g NaCl, 3 g MgCl<sub>2</sub>·6H<sub>2</sub>O, 6 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 1 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 200

mg NaHCO<sub>3</sub>, 300 mg CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.5 g KCl, 420 mg KH<sub>2</sub>PO<sub>4</sub>, 50 mg NaBr, 20 mg SrCl<sub>2</sub>·6H<sub>2</sub>O, and 10 mg Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O. The trace mineral solution (1000x) contains, per l, 0.5 g MnSO<sub>4</sub>·6H<sub>2</sub>O, 0.1 g CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.1 g ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.01 g CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.01 g AlK(SO<sub>4</sub>)<sub>2</sub>·12H<sub>2</sub>O, 0.01 g H<sub>3</sub>BO<sub>3</sub>, 0.01 g Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O. The vitamin mixture (200x) contains, per l, 0.2 g niacin, 0.08 g biotin, 0.2 g pantothenate, 0.2 g lipoic acid, 0.08 g folic acid, 0.2 g p-aminobenzoic acid, 0.2 g thiamine, 0.2 g riboflavin, 0.2 g pyridoxine, 0.2 g cobalamin. When present, sodium pyruvate was added at 5 g per l, agmatine sulfate to 1 mM and 6- methyl purine (6MP) to 100 µM (ASW-YT-Pyr). Elemental sulfur (S<sup>0</sup>) was added at 2 g per l in liquid media (ASW-YT-), but was replaced by polysulfide in solid media. Polysulfide solution (500x) contained, per l, 66.7 g sodium sulfide (Na<sub>2</sub>S·9H<sub>2</sub>O) and 3 g sulfur. Gelrite was added to 1% (w/v) to solidify media.

### ***Chromatin isolation and micrococcal nuclease digestion.***

Chromatin isolation and micrococcal nuclease digestions were adapted from Mattioli et. al. TS600, TS620 and TS622 were grown to an O.D.<sub>600</sub> of ~0.5 in liquid ASW-YT-S and used to inoculate (1:100) 200 ml of ASW-YT-Pyr per strain and allowed to grow to an O.D.<sub>600</sub> of ~0.5 to encourage chromatin reprogramming. Cultures were pelleted at 10,000 rpm and immediately frozen at -80° C. Cell pellets were resuspended in 1.0 mL of MNase buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl and 1 mM CaCl<sub>2</sub>) per 0.2 g of cell mass and ground to homogeneity with a mortar and pestle. Homogenized cells were mechanically lysed by repeated liquid nitrogen freezing and subsequent grinding 5 times. Whole cell lysate was gently clarified at 1,700 g for 5 minutes and the chromatin containing clarified lysate was RNase A digested (Sigma, 4,000 U) for 1 hour

at 37° C. ~1,500 U of micrococcal nuclease (New England Biolabs) was added to chromatin and aliquots (~100 µl) of digested DNAs were extracted by the addition of 300 µl of 10 mM Tris-HCl pH 8.0 and 400 µl of phenol/chloroform/isoamyl alcohol (25:24:1). Following thorough emulsion and centrifugation at 15,000 rpm for 5 minutes, ~200 uL of the DNA-containing aqueous layer was precipitated by the addition of an equal volume of 1 M Tris-HCl pH 8.0 and 2.6X volumes of 100% EtOH proceeding a 1-hour incubation at -80° C. DNAs were pelleted in a 4° C centrifuge at 15,000 rpm for 30 minutes and subsequently resolved in a 4% agarose gel.

### ***RNA isolation.***

TS600 and TS622 were grown in triplicate to an O.D.<sub>600</sub> of ~0.5 in liquid ASW-YT-S and used to inoculate (1:100) 300 ml of ASW-YT-Pyr per strain and allowed to grow to an O.D.<sub>600</sub> of ~0.05 to encourage chromatin reprogramming. Cultures were rapidly chilled and pelleted at 10,000 rpm for 5 min and then resuspended in 1.0 mL of Trizol (Invitrogen) with a 10-minute incubation at room temperature. 200 µl of chloroform was added followed by centrifugation at 15,000 rpm at 4 °C for 15 minutes yielding an RNA-containing aqueous layer which was added to 500 µl of isopropanol and incubated at room temperature for 10 minutes. Centrifugation at 15,000 rpm for 15 minutes at 4° C yielded an RNA pellet that was washed with 1 ml of 75% EtOH and subsequently resuspended in 88 µl of RNase-free H<sub>2</sub>O, 10 µl of DNaseI buffer and 1 µl of DNaseI (New England Biolabs) to digest residual DNA (37° C for 30 minutes). A second triplicate of TS600 and a triplicate of TS620 were prepared identically.

### ***TS600 and TS622 library construction and sequencing.***

Using ~120ng of RNA that was ribosomally depleted following the NEBNext® rRNA Depletion Kit (E6310), cDNA libraries were constructed with New England Biolabs NEBNext® Ultra™ Directional RNA Library Prep kit for Illumina® (E7420s) and NEBNext® Multiplex Oligos for Illumina® (Index primer set 1, E7335) according to the manufacture's procedure. The multiplexed libraries were sequenced at Cofactor Genomics using one high output NextSeq Illumina® run for single-end reads with a minimum read length of 75 bp and with a requested > 40 million reads per sample.

### ***TS600 and TS620 library construction and sequencing.***

1.5 µg of RNA was processed at Novogene for Prokaryotic RNA-seq, specifically for rRNA depletion (Ribo-Zero™ Magnetic Kit), library construction (NEBNext® Ultra™ RNA Library Prep kit) and 150 bp paired-end sequencing on a HiSeq Illumina® platform.

## **3.5 Data Preprocessing**

### ***TS600 and TS622.***

RNA-Seq reads were first analyzed for quality control using FastQC<sup>33</sup>. To remove adapter sequences and other artifacts, fastx-trimmer was used to trim the first 11 positions in each read<sup>34</sup>. After filtering, the reads were aligned to the *T. kodakarensis* (KOD1) reference genome using bowtie<sup>35</sup> with the following parameter: *-m 1*, this ensured suppressing all multiple-aligned reads. Finally, the bowtie output was converted to BAM format, sorted, and indexed using samtools<sup>36</sup>.

### ***TS600 and TS620.***

RNA-Seq reads were first analyzed for quality control using FastQC. To remove adapter sequences and other artifacts, fastx-trimmer was used to trim the first 19

positions in each read. After filtering, the reads were aligned to the *T. kodakarensis* (KOD1) reference genome using bowtie default parameters. Next, in each library, reads that aligned to multiple locations were filtered out. Finally, the filtered output was converted to BAM format, sorted, and indexed using samtools.

### ***Differential Gene Expression Analysis.***

To identify differentially expressed genes we used EdgeR<sup>37</sup>. To generate read counts for every gene in Ensembl based annotations of the species, we developed a custom python script that used SpliceGrapher and pysam<sup>38,39</sup>. The EdgeR p-values were adjusted for multiple comparisons using the Benjamini-Hochberg method<sup>40</sup>. Finally, we used a cutoff of 1.00 on the transcript abundance (log CPM) and fold change (log FC).

### ***DNA Sequencing of TS620.***

Genomic DNA was purified from strain TS620 using the Monarch Genomic DNA Purification kit (New England Biolabs). Pacific Biosciences (PacBio) libraries were constructed following the Pacific Biosciences Template Preparation and Sequencing Protocol. The library was then sequenced on a PacBio Sequel instrument using Polymerase 3.0 Chemistry and diffusion loading for 600 minutes. The data was then analyzed using PacBio SMRT Analysis tools.

## REFERENCES

1. Peeters, E., Driessen, R. P. C., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* **13**, 333–41 (2015).
2. Sanders, T. J., Marshall, C. J. & Santangelo, T. J. The Role of Archaeal Chromatin in Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.05.006
3. Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019). doi:10.1111/mmi.14191
4. Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.-P., Reeve, J. N., Santangelo, T. J. & Fondufe-Mittendorf, Y. N. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* **14**, 391 (2013).
5. Sandman, K. & Reeve, J. N. Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.* **173**, 165–9 (2000).
6. Chang, C. H. & Luse, D. S. The H3/H4 tetramer blocks transcript elongation by RNA polymerase II in vitro. *J. Biol. Chem.* **272**, 23427–34 (1997).
7. Farnung, L., Vos, S. M. & Cramer, P. Structure of transcribing RNA polymerase II-nucleosome complex. *Nat. Commun.* **9**, (2018).
8. Decanniere, K., Babu, A. M., Sandman, K., Reeve, J. N. & Heinemann, U. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* **303**, 35–47 (2000).

9. Kireeva, M. L., Hancock, B., Cremona, G. H., Walter, W., Studitsky, V. M. & Kashlev, M. Nature of the nucleosomal barrier to RNA polymerase II. *Mol. Cell* **18**, 97–108 (2005).
10. Crickard, J. B., Lee, J., Lee, T.-H. & Reese, J. C. The elongation factor Spt4/5 regulates RNA polymerase II transcription through the nucleosome. *Nucleic Acids Res.* **45**, 6362–6374 (2017).
11. Gaffney, D. J., McVicker, G., Pai, A. A., Fondufe-Mittendorf, Y. N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y. & Pritchard, J. K. Controls of Nucleosome Positioning in the Human Genome. *PLoS Genet.* **8**, (2012).
12. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
13. Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y. Y., Lubkowska, L., Kashlev, M. & Bustamante, C. Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151**, 738–749 (2012).
14. Xie, Y. & Reeve, J. N. Transcription by an Archaeal RNA Polymerase Is Slowed but Not Blocked by an Archaeal Nucleosome. *J. Bacteriol.* **186**, 3492–3498 (2004).
15. Mattioli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., Byrne, K. R., Lee, T., Ahn, N. G., Santangelo, T. J., Reeve, J. N. & Luger, K. Structure of histone-based chromatin in Archaea. *Science (80-. )*. **357**, 609–612 (2017).
16. Bhattacharyya, S., Mattioli, F. & Luger, K. Archaeal DNA on the histone merry-

- go-round. *FEBS J.* **285**, 3168–3174 (2018).
17. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15**, 684 (2014).
  18. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**, 352–363 (2005).
  19. Gorlas, A., Koonin, E. V., Bienvenu, N., Prieur, D. & Geslin, C. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. *Environ. Microbiol.* **14**, 503–516 (2012).
  20. Mochizuki, T., Sako, Y. & Prangishvili, D. Provirus induction in hyperthermophilic archaea: characterization of *Aeropyrum pernix* spindle-shaped virus 1 and *Aeropyrum pernix* ovoid virus 1. *J. Bacteriol.* **193**, 5412–5419 (2011).
  21. Tagashira, K., Fukuda, W., Matsubara, M., Kanai, T., Atomi, H. & Imanaka, T. Genetic studies on the virus-like regions in the genome of hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Extremophiles* **17**, 153–60 (2013).
  22. Cubonova, L., Katano, M., Kanai, T., Atomi, H., Reeve, J. N. & Santangelo, T. J. An Archaeal Histone Is Required for Transformation of *Thermococcus kodakarensis*. *J. Bacteriol.* **194**, 6864–6874 (2012).
  23. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *BIO-PROTOCOL* **7**, (2017).

24. Hileman, T. H. & Santangelo, T. J. Genetics techniques for *Thermococcus kodakarensis*. *Front. Microbiol.* **3**, 195 (2012).
25. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic Techniques for the Archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).
26. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* **81**, 897–911 (2011).
27. Pan, M., Santangelo, T. J., Čuboňová, L., Li, Z., Metangmo, H., Ladner, J., Hurwitz, J., Reeve, J. N. & Kelman, Z. *Thermococcus kodakarensis* has two functional PCNA homologs but only one is required for viability. *Extremophiles* **17**, 453–461 (2013).
28. Richmond, R. K., Sargent, D. F., Richmond, T. J., Luger, K. & Ma, A. W. Crystal structure of the nucleosome ° resolution core particle at 2 . 8 Å. **7**, 251–260 (1997).
29. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
30. Dulmage, K. A., Todor, H. & Schmid, A. K. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *MBio* **6**, e00649-15 (2015).
31. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.06.020
32. Gonnet, M., Erauso, G., Prieur, D. & Le Romancer, M. PAMT11, a novel plasmid

- isolated from a *Thermococcus* sp. strain closely related to the virus-like integrated element TKV1 of the *Thermococcus kodakaraensis* genome. *Res. Microbiol.* **162**, 132–143 (2011).
33. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
  34. Gordon, A., (unpublished, G. H.-F. short-reads preprocessing tools & 2010, undefined. Fastx-toolkit.
  35. Langmead, B. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* **Chapter 11**, Unit 11.7 (2010).
  36. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
  37. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
  38. Heger, A., Belgrad, T., Goodson, M., mapping, K. J.-B. sequence alignment and & 2014, undefined. pysam: Python interface for the SAM.
  39. Rogers, M. F., Thomas, J., Reddy, A. S. N. & Ben-Hur, A. SpliceGrapher: Detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biol.* **13**, (2012).
  40. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

## CHAPTER 4: FTTA IS A CPSF73 HOMOLOGUE THAT TERMINATES TRANSCRIPTION IN ARCHAEA<sup>2</sup>

### 4.1 Summary

Regulated gene expression is achieved in large part by controlling the activities of essential, multi-subunit RNA polymerase transcription elongation complexes (TECs)<sup>1–3</sup>. The extreme stability required of TECs to processively transcribe large genomic regions necessitates robust mechanisms to terminate transcription<sup>4–6</sup>. Efficient transcription termination is particularly critical for gene-dense bacterial and archaeal genomes wherein continued transcription would necessarily transcribe immediately adjacent genes, result in conflicts between the transcription and replication apparatuses and the coupling of transcription and translation would permit loading of ribosomes onto aberrant transcripts<sup>1,5,7–9</sup>. Only select sequences or transcription termination factors can disrupt the otherwise extremely stable TEC and we demonstrate that one of the last universally conserved archaeal proteins with unknown biological function is the Factor that terminates transcription in Archaea (FttA)<sup>9–14</sup>. FttA resolves the dichotomy of a prokaryotic gene structure (operons and polarity) and eukaryotic molecular homology (general transcription apparatus) observed in Archaea. This missing-link between prokaryotic and eukaryotic transcription regulation provides the most parsimonious link to the evolution of the processing activities involved in RNA 3'-end formation in Eukarya.

Transcription termination, driven by DNA sequence and encoded RNA structures

---

<sup>2</sup> This chapter was published in February 2020 under the same title.

Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E. & Santangelo, T. J. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* 5, 545–553 (2020).

(e.g. intrinsic termination) or by protein factors (e.g. factor-dependent termination) ensures rapid dissociation of RNA polymerase (RNAP) from the DNA template to recycle RNAP and generate RNA 3' ends<sup>7-11,15,16</sup>. While often prevalent within prokaryotic genomes, intrinsic termination sequences are typically neither sufficiently abundant nor efficient to mediate all termination events<sup>7,11,15</sup>. Transcription termination factors must then efficiently recognize TECs that are not intrinsically terminated and compete with continued elongation to mediate release of the nascent transcript. While the identification of Eta provided evidence of factor-dependent archaeal termination<sup>16</sup>, no kinetically-efficient mechanism of factor-dependent archaeal transcription termination had been described<sup>1,5</sup>. The retention of operon-organized archaeal genomes and the sensitivity of the archaeal transcription apparatus to bacterial rho-mediated termination *in vitro* – combined with the normal coupling of transcription and translation, the resultant polar suppression of downstream expression in the absence of such coupling in archaeal cells and the conservation of Spt5/NusG in all genomes – implied the existence of a kinetically-relevant archaeal transcription termination activity that might function akin to the bacterial rho protein<sup>17-22</sup>. Rho homologues are, however, restricted to Bacteria, arguing instead that conserved archaeal-eukaryotic or unique archaeal factors may drive factor-dependent archaeal transcription termination.

Only a core set of ~200 gene families (more properly, archaeal clusters of orhologous genes; arCOGs) are conserved in most archaeal genomes, just ~129 arCOGs are strictly ubiquitous and one is an obvious orthologue of a subunit of the cleavance and polyadenylation specificity factor (CPSF) complex<sup>23</sup>. The homology of most archaeal transcription components to eukaryotic factors argued that the archaeal

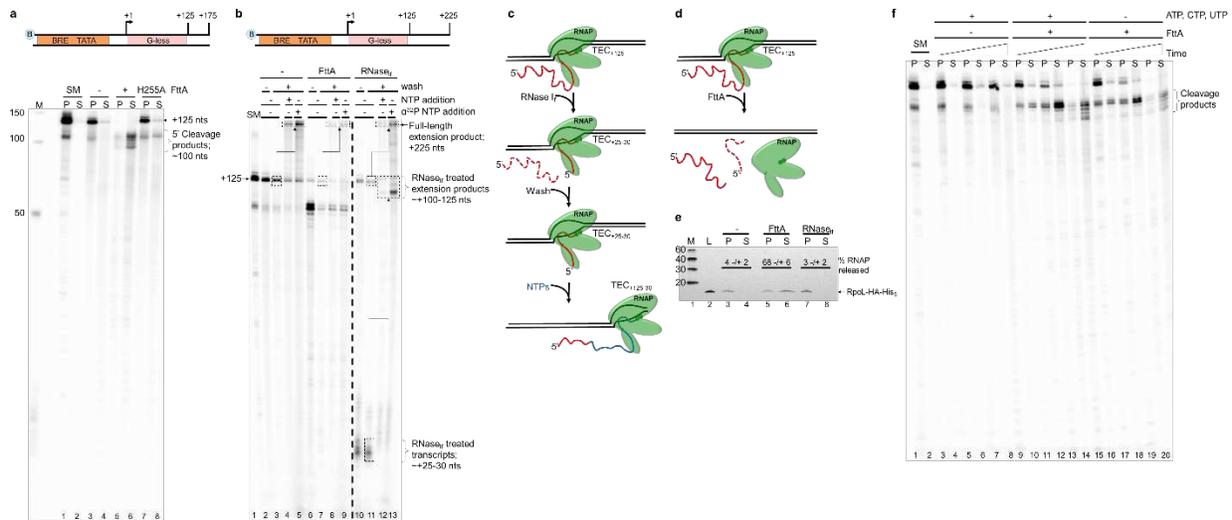
homologue of eukaryotic CPSF73 might function as the Factor that terminates transcription in Archaea (FttA)<sup>6,9,12–14,24,25</sup>.

## 4.2 Results

### ***FttA is a bona fide archaeal transcription termination factor.***

We challenged promoter-initiated TECs – generated with an RNAP variant with a His<sub>6</sub>/HA-epitope-tagged RpoL subunit and containing a radiolabeled nascent transcript – with FttA and monitored transcription termination by quantifying release of transcripts from TECs (Figure 4.1). TECs stalled by nucleotide deprivation with +125 nucleotide (nt) nascent transcripts (TECs<sub>+125</sub>) remain stably associated in the absence of FttA (Figure 4.1a, lanes 1-4). Addition of FttA to stalled TECs results in cleavage and release of ~100 nts of the nascent transcript (Figure 4.1a, lanes 5-6). However, despite repeated and exhaustive efforts to monitor FttA-mediated transcript cleavage within seconds of FttA-addition, we never observed a ~25nt 3'-transcript fragment. We were thus initially hesitant to assume that FttA-mediated transcript cleavage was coupled to *bona fide* transcription termination, as a ~25 nt transcript is sufficient to stabilize an archaeal TEC.

To fully validate that the cleavage and termination activity of FttA is distinct from that of a general RNase, we challenged TECs with either FttA or RNase I<sub>r</sub> in parallel. If TECs remain intact following FttA-mediated cleavage of the nascent transcript then i) radiolabeled 3'-nascent transcripts should remain associated with TECs, ii) intact TECs should survive washes designed to remove transcripts not associated with TECs, iii) NTP addition should permit continued elongation of active TECs, allowing extension of the nascent transcript, and iv) RNAP should remain within TECs. In contrast, if FttA-



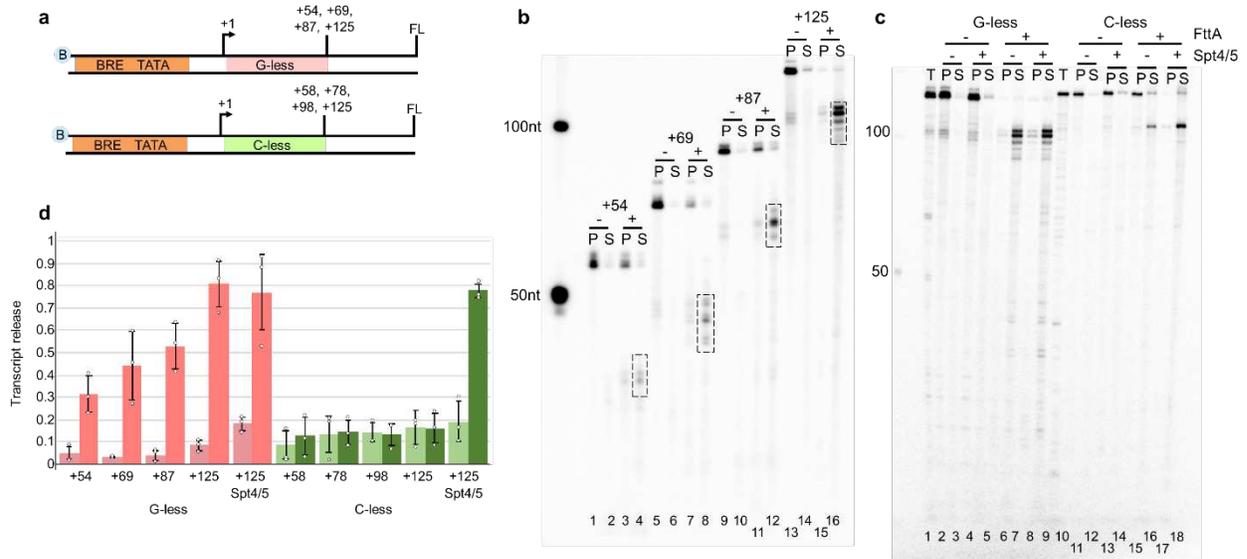
**Figure 4.1 FttA is a bona-fide termination factor.** (a) FttA directs transcript cleavage and transcription termination. Transcripts associated with intact TECs are retained in pellet (P) fractions whereas transcripts released from terminated complexes partition into the supernatant (S). Addition of ATP, CTP, UTP and <sup>32</sup>P-a-UTP allows RNAP to elongate to the end of a +125 nt G-less cassette and generate a uniformly radiolabeled RNA transcript. Radiolabeled transcripts within starting material (SM) TECs<sub>+125</sub> and mock treated TECs<sub>+125</sub> are retained in pellet fractions (lanes 1-4), whereas FttA<sup>WT</sup> addition results in cleavage of nascent transcripts and termination of most TECs (lanes 5-6). Addition of a catalytically deficient FttA variant (FttA<sup>H255A</sup>) abrogates cleavage and RNA release (lanes 7-8). Lane M contains <sup>32</sup>P-labeled ssDNA markers. (b) FttA-mediated termination is distinct from RNase treatment of intact TECs. TECs<sub>+125</sub> (SM, lane 1) are resistant to repeated high-salt buffer washes and readily resume elongation upon NTP addition to generate +225 nt full-length transcripts (lanes 2-5). Dashed boxes and arrows denote +125 transcripts that are elongated to +225 nt transcripts; the specific activity of +225 transcripts can be increased by addition of additional <sup>32</sup>P-a-UTP during elongation from +125 to +225. RNase I<sub>f</sub> digestion of nascent transcripts associated with washed TECs<sub>+125</sub> results in degradation of the nascent transcript to just ~20-30 nts, but TECs with shortened transcripts remain associated with the DNA and thus survive repeated washing (lanes 10-11). TECs<sub>+125</sub> resultant from RNase I<sub>f</sub> treatment of TECs<sub>+125</sub> readily resume elongation upon NTP addition to generate ~+125 nt full-length transcripts (lanes 12-13). Dashed boxes and arrows denote ~+25 transcripts that are elongated to ~+125 nt transcripts; the specific activity of ~+125 transcripts can be increased by addition of additional <sup>32</sup>P-a-UTP during elongation from ~+25 to ~+125. FttA addition to TECs<sub>+125</sub> results in the disruption of most TECs with nascent transcript cleavage (lanes 6-9). FttA-mediated termination results in release of most TECs from the template and cleaved transcripts cannot be extended by NTP addition (lanes 8-9). (c) and (d) Schematic diagrams of the fate of TECs<sub>+125</sub> follow RNase I<sub>f</sub> and FttA treatment, respectively. (e) FttA, but not RNase I<sub>f</sub> treatment releases RNAP from the DNA template into solution confirming dissociation of the TEC and *bona fide* FttA-mediated transcription termination. (f) FttA is not reliant on NTP hydrolysis to quickly inactivate TECs, cleave nascent transcripts and

mediated cleavage of the transcripts inactivates and terminates transcription, RNAP should be released to the supernatant and resumed elongation following NTP supplementation will not be possible. Treatment of TECs<sub>+125</sub> with RNase I<sub>f</sub> (Figure 4.1c) fulfills all of the expectations of transcript processing that are not linked to transcription termination: stable TECs<sub>-+25</sub> are observed (Figure 4.1b, lanes 10-11), addition of unlabeled NTPs (Figure 4.1b, lane 12) or radiolabeled NTPs (Figure 4.1b, lane 13) permits all TECs<sub>-+25</sub> to resume elongation, and RNAP partitioning confirms essentially all TECs remain intact (Figure 4.1e, lanes 7-8). Treatment of identically prepared TECs<sub>+125</sub> with FttA (Figure 4.1b-e) are, in contrast, fully supportive of FttA-mediated termination: the bulk of FttA-treated TECs<sub>+125</sub> do not survive washes and FttA-activity releases ~70% of RNAP to solution (Figure 4.1e, lanes 5-6). FttA is thus the second archaeal-encoded factor that can mediate transcription termination.

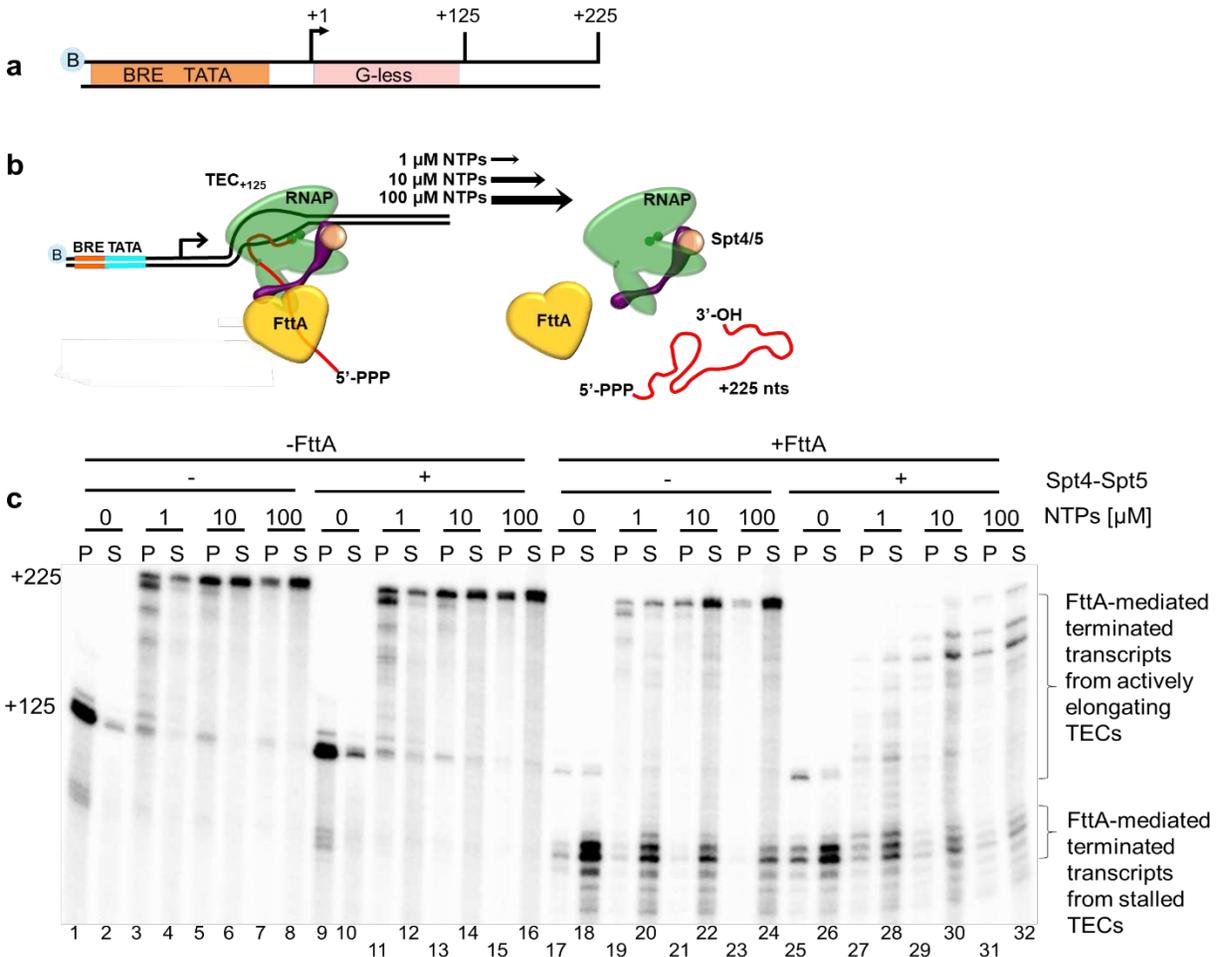
Addition of an FttA variant - FttA<sup>H255A</sup> - reduced but did not eliminate FttA-mediated termination (Figure 4.1a, lanes 7-8). Termination activity is thus linked to FttA-mediated RNA cleavage, rather than FttA-mediated stimulation of the intrinsic cleavage activity of RNAP. FttA-mediated cleavage of the nascent RNA to yield a ~100 nt 5'-transcript is consistent with FttA stimulating RNA cleavage at the first solvent accessible phosphodiester linkage and the ~25 nt of nascent transcript protection are consistent with the results of previous digestions of intact archaeal and eukaryotic TECs with RNA exonucleases<sup>9,16</sup>. In contrast to other prokaryotic transcription termination factors, FttA-mediated termination is not energy-dependent (Figure 4.1f).

***FttA-mediated termination shares mechanistic requirements of Rho-mediated bacterial transcription termination.***

FttA recognizes TECs through binding to nascent transcripts (Figure 4.2). TECs stalled on G-less cassettes, and thus with A, U and C-rich RNAs, revealed a near-linear relationship between transcript length and FttA-mediated termination (Figure 4.2b-d). While FttA-mediated termination is possible with only short segments of solvent accessible nascent transcript sequences – a notable discriminating feature between rho- and FttA-mediated termination – the efficiency and rate of FttA-mediated termination are modest in such instances. In contrast, TECs stalled on C-less cassettes, and thus with A, U and G-rich RNAs effectively abolish FttA-activity (Figures 4.2). FttA-mediated termination is thus stimulated by C-rich RNAs – as is the case for bacterial rho-mediated termination – or is inhibited by transcripts that are particularly G-rich. Rho-activity can be stimulated at suboptimal *rut*-sites by NusG, and the archaeal-eukaryotic homologue of NusG, Spt5, together with its common binding partner Spt4, can likewise stimulate FttA when transcript sequences limit FttA-recognition or FttA-activity (Figure 4.2c-d). As such, Spt4-Spt5 temper the nucleotide requirements of FttA. FttA is a known endo- and 5'-3' exonuclease and cleavage of nascent transcripts is stimulated by interactions with the archaeal TEC, but not RNAP. FttA-mediated cleavage of TEC-associated nascent transcripts is complete within ~1-2 minutes, while incubations of FttA with purified RNA under identical conditions require ~30-times longer to generate even mild cleavage patterns, consistent with previous results. FttA-mediated endonucleolytic cleavage of free RNA at CA and CC dinucleotide sequences is consistent with FttA activity on C-rich transcripts. The consistently observed cleavage



**Figure 4.2 FttA-mediated termination shares mechanistic requirements of rho-mediated bacterial transcription termination.** (a) Promoter-directed transcription of biotinylated templates encoding G-less or C-less cassettes permits formation of TECs with increasing length A-, C-, and U-rich, or A-, G-, and U-rich nascent transcripts, respectively. FL = full-length; all templates permit elongation for 100 nts beyond the G- or C-less cassette. (b) TECs remain stably associated and transcripts are primarily recovered in the pellet (P) fraction in the absence (-) of FttA. When FttA is present (+), transcripts are cleaved and primarily recovered in the supernatant (S) fraction. Cleavage releases ~20 – 30 nt shorter transcripts (boxed). The left-most lane contains  $^{32}$ P-labeled ssDNA markers. (c) Addition of Spt4-Spt5 largely abrogates the RNA sequence-requirements of FttA-mediated transcription termination. T= total reaction = P+S. The left-most lane contains  $^{32}$ P-labeled ssDNA markers. (d) Transcript release was quantified with and without FttA addition for TECs with increasing length transcripts on G-less (pink/salmon) and C-less cassettes (mint/green), with and without Spt4-Spt5 addition for TECs+125 formed on G- and C-less cassettes. Error bars were calculated as standard deviation from the mean ( $n \geq 3$  replicates).



**Figure 4.3 FttA-mediated transcription termination is competitive with transcription elongation.** (a and b) Washed, NTP-deprived TECs+125 were assembled on biotinylated templates with a +125 nt G-less cassette. Resumed elongation upon differential [NTP] addition permits transcription to generate +225 nt transcripts, albeit at different rates. (c) FttA readily terminates stalled or slowly elongating TECs (lanes 17-24) and FttA-mediated termination becomes competitive with transcription elongation even at high [NTP] in the presence of Spt4-Spt5 (lanes 25-32).

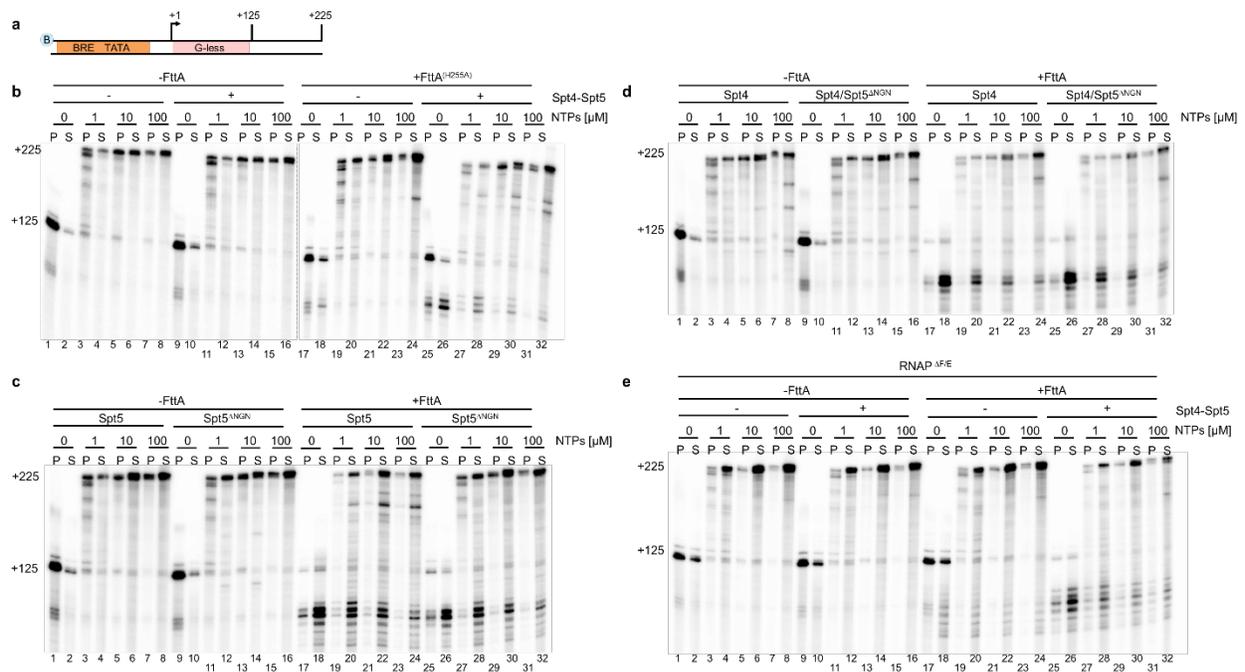
pattern on various substrates (reduced transcript length by ~20-30 nts) supports FttA-mediated cleavage and termination being dictated and positioned by RNAP-RNA interactions near the stalk-domain and RNA exit channel and is further enhanced by Spt4-Spt5.

***FttA-mediated transcription termination is competitive with transcription elongation.***

Archaeal transcription units are typically separated by only short (< 100 bp) intergenic regions<sup>11,15</sup>, thus for FttA-mediated termination to be effective mechanism of gene regulation *in vivo*, FttA must quickly recognize and disrupt TECs before transcription continues significantly into downstream genes or operons. To establish if FttA-mediated termination was competitive with transcription elongation, stalled TECs<sub>+125</sub> were permitted to resume elongation with different [NTPs]. Differential elongation rates resultant from varying [NTPs] provide a relative measure of the efficiency of FttA-mediated transcription termination in competition with transcription elongation (Figure 4.3). At low [NTPs], TECs elongate slowly and many TECs are still transcribing after several minutes of incubation as evidenced by a mixture of nascent transcripts between 125-225 nts (Figure 4.3b, lanes 3-4). At increasingly higher [NTP], elongation rates increased until TECs are elongating at rates comparable to normal elongation rates *in vivo* (Figure 4.3b, lanes 5-8). Addition of FttA to stalled TECs (Figure 4.3b, lanes 17-18) resulted in near complete termination, but as the rate of elongation increased with increasing [NTP], FttA-mediated termination decreased. Although not an obligate subcomplex of archaeal RNAP, Spt4-Spt5 engages RNAP *in vivo* early during elongation and remains associated with TECs throughout long genes. The ability of

Spt4-Spt5 to temper the transcript requirements for FttA-mediated termination (Figure 4.2) suggested that addition of Spt4-Spt5 may accelerate FttA-recognition of or action towards TECs. In support of this hypothesis, addition of Spt4-Spt5 greatly increased the termination efficiency of FttA, demonstrated by the release of transcripts  $> +125$  nt but  $< +225$  nt (Figure 4.3b, lanes 27-32). The results demonstrate that FttA is kinetically coupled to RNAP via elongation factors Spt4-Spt5, a striking analogy to the stimulation of the unrelated bacterial rho protein by NusG and to the observed stimulation of Pol II termination by CPSF73/Xrn2<sup>6,24,26-28</sup>. To ensure that FttA mediates termination when combined with Spt4-Spt5 – and that termination observed in the presence of all three factors was not a new activity of Spt4-Spt5 – we employed a variant of FttA (FttA<sup>H255A</sup>) that retains only partial activity (Figure 4.4a and b)<sup>25</sup>.

Interactions between Rho and the C-terminal KOW domain of NusG stimulate rho-mediated termination<sup>29</sup>. The NusG-KOW domain is normally engaged with the ribosome and only becomes available when transcription becomes uncoupled from translation. Archaeal transcription and translation are coupled, and we asked whether the isolated KOW domain of Spt5 would suffice to stimulate FttA-mediated termination. Addition of the Spt5-KOW-domain (Spt5<sup>ΔNGN</sup> which remains thermostable) alone does not influence the activities of FttA or RNAP *in vitro* (Figure 4.4c). Spt5 is often in a heterodimeric partnership with Spt4, and this partnership is critical to kinetically couple FttA activity to RNAPs (Figure 4.4d). Like the nuclear eukaryotic RNAPs, the archaeal RNAP contains a stalk domain. The stalk provides binding surfaces for conserved initiation and elongation factors and the nascent transcript. Purified stalk-less RNAP (RNAP<sup>ΔE/F</sup>), when combined with TBP and TFB, is competent for transcription initiation,

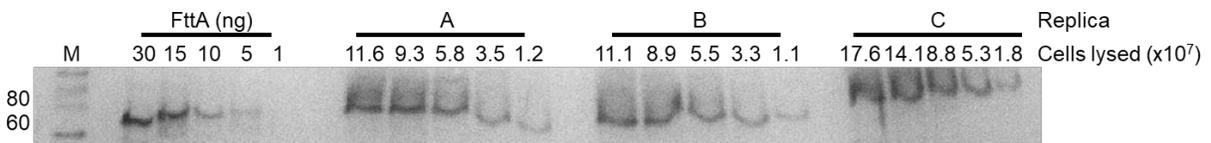


**Figure 4.4 FttA-mediated transcription termination requires both spt4 and spt5, the endonuclease activity of FttA and the stalk domain of RNAP.** (a) TECs<sub>+125</sub> were generated to assay various components of the kinetically competitive termination complex. (b) FttA nuclease activity is required for kinetically competitive termination. (c) Neither Spt5 alone or (d) Spt4 alone was sufficient to facilitate competitive transcription termination. Additionally, the NGN domain of Spt5 is required. (e) The stalk domain of RNAP is required to for competitive transcription termination by FttA.

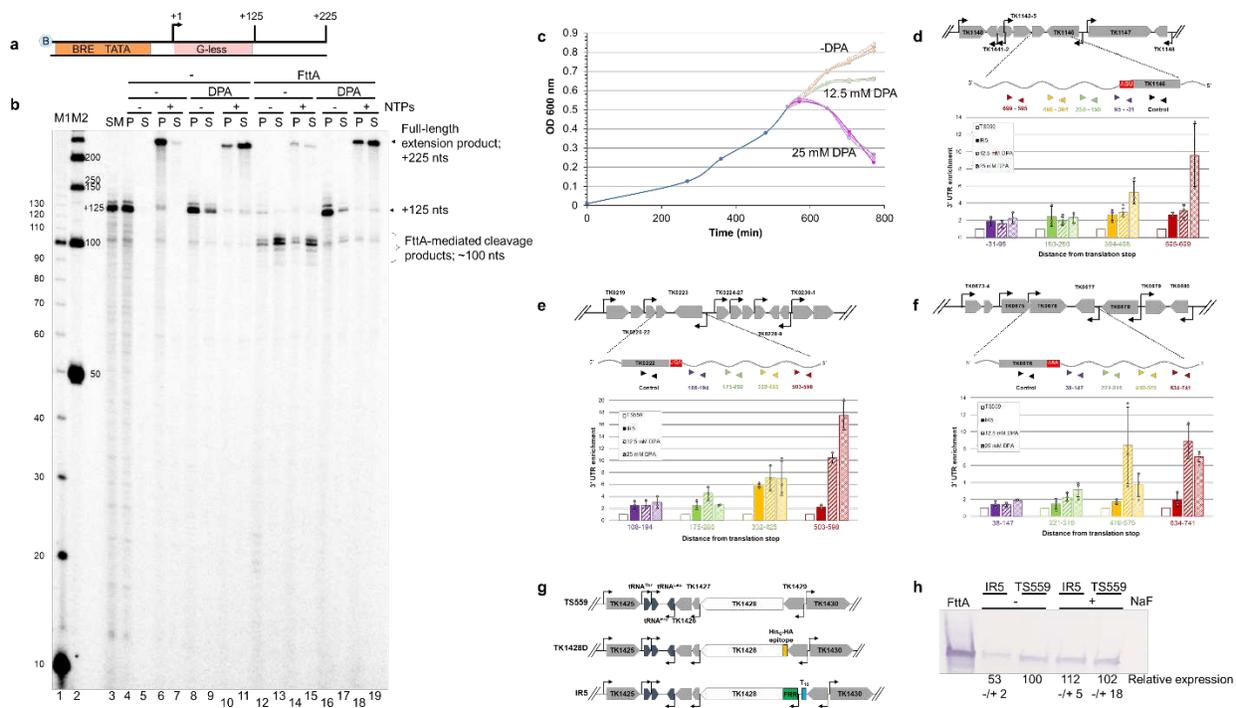
elongation, and intrinsic termination, but fails to respond correctly to FttA-mediated termination (Figure 4.4e). Even when continued elongation was prohibited, the termination activities of FttA were stunted by the loss of the RNAP-stalk domain (Figure 4.4e). Addition of Spt4-Spt5 to TECs assembled with RNAP $\Delta E/F$  does not stimulate FttA-mediated termination to rates that are competitive with continued elongation even at low [NTP] (Figure 4.4e, lanes 27-32). The results suggest roles for both the stalk domain and Spt5-Spt4 in accelerating FttA-mediated termination to permit kinetically competitive termination *in vitro*.

***Inhibition of FttA activity abolishes transcription termination in vitro and reduced FttA expression or activity alters steady-state RNA 3' termini.***

FttA is likely sufficiently abundant (~2,100 +/- 500 molecules/cell; Figure 4.5) to monitor global transcription (RNAP is estimated at ~3,000 molecules/cell). FttA is a metallo-beta lactamase-fold protein containing a beta-CASP domain and we predicted that general inhibitors of metallo-beta-lactamase proteins may impair FttA activity. 2,6-pyridine dicarboxylic acid (dipicoloinic acid, or DPA) nearly-completely inhibited FttA-mediated termination *in vitro* (Figure 4.6a-b)<sup>30</sup>. *T. kodakarensis* cultures challenged with DPA demonstrated a reduced, then complete inhibition of growth (Figure 4.6c). While DPA may impact several factors *in vivo*, we rationalized that monitoring RNA 3'-ends following DPA addition may reveal changes due to inactivation of FttA. Quantitative, reverse-transcription-PCR (q-RT-PCR) analyses revealed significant changes to 3'-ends at several loci *in vivo* (Figure 4.6d-f) following DPA addition. The fold-changes in extended 3' ends of each transcript generally increase in magnitude compared to untreated cultures both with respect to the distance from the translation stop codon and



**Figure 4.5 FttA is an abundant protein likely responsible for 3'-end formation in archaeal cells.** Quantitative Western blots employing anti-FttA antibodies, purified recombinant FttA, and total cellular lysates derived from known numbers of lysed *T. kodakarensis* cells reveal that FttA is present at ~2,100 +/- 500 copies per cell. Cell counts and protein calculations were performed as described.



**Figure 4.6 Inhibition of FttA activity abolishes transcription termination in vitro and reduced FttA-expression or activity alters steady-state RNA 3'-termini in vivo.** (a and b) TECs+125 (lane 3) assembled on biotinylated templates with G-less cassettes are stable and readily resume elongation upon NTP addition to generate +225 nt transcripts in the presence or absence of 25 mM DPA. FttA addition results in transcript cleavage, termination, and release of most TECs to the supernatant in the absence of DPA. Incubation of FttA with 25 mM DPA prior to addition to reactions inhibits FttA-mediated termination and RNA cleavage, permitting intact TECs+125 to resume elongation and generate +225 nt transcripts. Lanes M1 and M2 contain 32P-labeled 10- and 50-nt ssDNA markers, respectively. (c) Inhibition of metallo beta-lactamase/beta-CASP protein activity by addition of DPA impairs growth of *T. kodakarensis*. An actively growing culture of *T. kodakarensis* strain TS559 was split after reaching an optical density of 0.5 to nine cultures, with three biological replicates of cultures exposed to 0 mM DPA (peach series), 12.5 mM DPA (green series) or 25 mM DPA (purple series). (d, e and f) RNAs recovered one-hour post DPA-addition to cultures of TS559, or from cultures of IR5 grown in the absence of NaF display altered 3'-termini. Trizol extracted RNAs were reverse transcribed with oligonucleotide primers complementary to nascent transcript sequences of TK1146, TK0222 and TK0676 to generate cDNAs that were quantified and normalized to internal controls. RNA abundance in untreated TS559 cultures (open bars) was set to 1.0, and fold changes in the abundance of amplicons reflecting RNA transcripts with extended 3'-sequences at increasing distances from the translation stop site (purple, green, orange and red) are shown for strain IR5 (solid bars), strain TS559 treated with 12.5 mM DPA (wide stripes) and strain TS559 treated with 25 mM DPA (narrow stripes). Errors are calculated from minimally triplicate technical replicates of at least three biological replicates. (g) Genome maps of the TK1428 locus in parental (TS559), N-terminally tagged (TK1428D) and riboswitch-regulated expression (IR5) strains of *T. kodakarensis*. (h) Western blot demonstrating the reduction in steady-state FttA protein levels in strain IR5 upon removal of NaF from the culture medium;  $n \geq 6$  independent replicates.

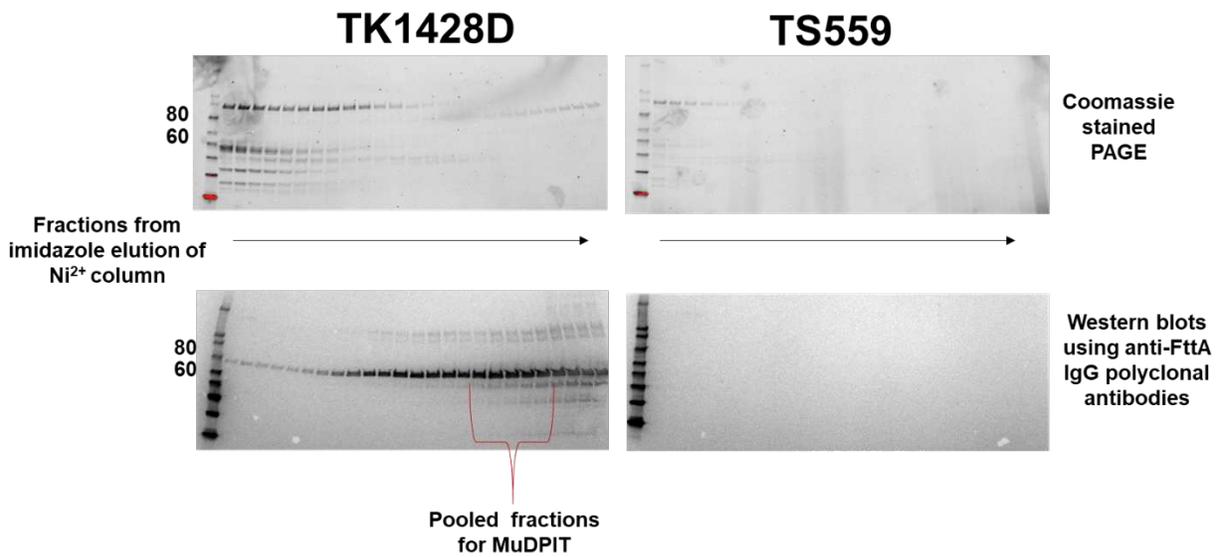
with increasing [DPA]. Altered 3'-termini stemming from exposure to a general metallo-beta-lactamase inhibitor cannot be definitively attributed to direct inhibition of FttA activity *in vivo*. Given our desire to directly demonstrate that reduced FttA activity impacts termination *in vivo*, coupled with our inability to generate *T. kodakarensis* strains encoding enzymatically-impaired FttA variants, we next reduced FttA activity by limiting FttA expression and altering steady-state FttA<sup>WT</sup> protein levels. To ensure that introduced and regulated expression of TK1428 did not impact TK1429 expression, we separated TK1428 expression from TK1429 by introduction of a new promoter and intrinsic termination sequence, then placed the TK1428 coding sequences downstream of sequences encoding an archaeal fluoride-responsive riboswitch, thereby generating strain IR5 (Figure 4.6g)<sup>31</sup>. Construction of IR5 was only possible when cultures were continuously provided with fluoride even though fluoride impairs growth of *T. kodakarensis*, supporting that very limited expression of TK1428 was not compatible with life. Steady-state FttA levels in IR5 strains grown in the absence and presence of fluoride reveals a modest ~2-fold change in FttA levels *in vivo* when fluoride is removed from cultures (Figure 4.6h), yet even this modest alteration significantly and reproducibly impacts transcription termination *in vivo* (Figure 4.6d-f). The increased abundance of RNA with extended 3'-UTRs in strains with reduced FttA protein abundance is supportive of FttA normally directing transcription termination *in vivo*.

FttA is conserved in all archaeal genomes, including the severely reduced genomes of symbiotic Nanoarchaeota, and it was perhaps not surprising that exhaustive attempts to delete or generate variants that radically impair activity of FttA in *T. kodakarensis* were unsuccessful; our failures were supported by the essentiality of

FttA in other archaea<sup>23,32</sup>. We were able to generate a strain (termed TK1428D) encoding an His<sub>6</sub>-affinity and HA-epitope tagged FttA (Figure 4.6g). Strain TK1428D growth was indistinguishable from the parental strain<sup>33,34</sup> and N-terminally tagged FttA was recovered directly from TK1428D cell lysates in large abundance (Figure 4.7). Proteins co-purifying with FttA from TK1428D were identified by MuDPIT<sup>35,36</sup>, returning only a small number of proteins (Figure 4.7) that have minimal inferred activity related to transcription and gene expression. No obvious stoichiometric FttA interaction partners were recovered, supportive of our *in vitro* demonstration that FttA alone can disrupt archaeal TECs. Affinity purification of FttA does not return RNAP subunits nor Spt4-Spt5, suggesting FttA transiently encounters and disrupts TECs rather than forming stable interactions with TEC components or Spt4-Spt5.

### 4.3 Discussion

The essentiality of FttA in *T. kodakarensis* and other archaea<sup>32,37</sup>, the complete conservation of FttA in Archaea<sup>23</sup>, the demonstrated *in vitro* ability of FttA-mediated termination to compete with productive elongation (Figure 4.3) and the demonstrated changes to RNA 3'-ends in strains wherein FttA activity is reduced by two independent mechanisms (Figure 4.6) suggests that FttA is likely responsible for 3'-end formation of transcripts that are not directed by intrinsic termination, and further that FttA-mediated termination is likely responsible for polarity in archaeal cells<sup>10</sup>. By establishing the requirements for FttA-mediated transcription termination we complete the archaeal transcription cycle and describe an additional mechanism of 3' end formation (Figure 4.8). It is important to note that the described activities of FttA suggest that the steady-state 3'-termini of *in vivo* transcripts terminated by FttA do not reflect that actual position



Gene	Annotation	Mascot Score	Number of Unique Peptides
TK 1428	Cleavage and Polyadenylation specificity factor homologue	6388	47
TK 2215	tRNA splicing endonuclease	633	12
TK 0011	Uncharacterized protein	585	7
TK 1557	Predicted dehydrogenase	537	18
TK 1165	Predicted AP endonuclease	468	18
TK 2250	Serine/Threonine protein kinase	421	17
TK 0976	Putative snRNP Sm-like protein	351	8
TK 1509	Probable tRNA pseudouridine synthase	202	13
TK 0528	Serine hydroxyethyltransferase	168	13
TK 0211	Amidophosphoribosyltransferase	147	9
TK 1305	Probable translation initiation factor IF-2	133	11

**Figure 4.7 Western blot analysis of TK1428D versus TS559 and the proteins identified as co-eluting partners of FttA from lysates of strain TK1428D.**

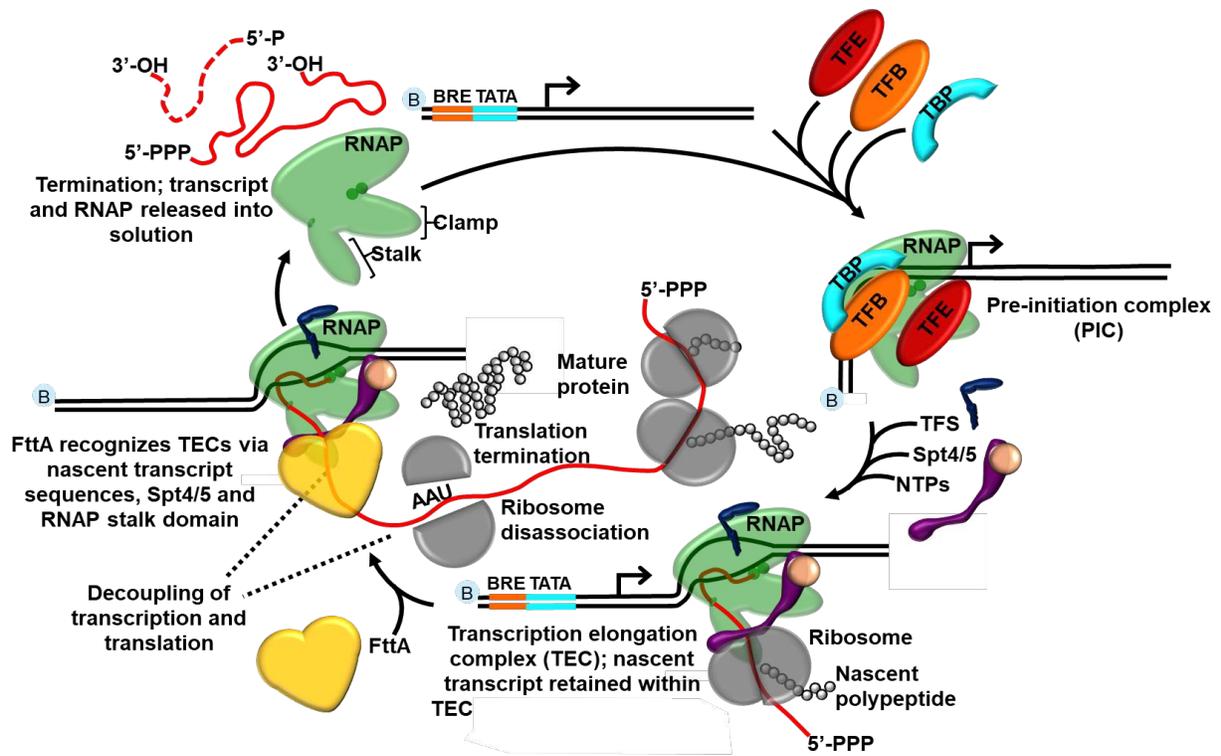


Figure 4.8 FttA completes the archaeal transcription cycle.

of termination of the archaeal RNAP. Thus, consensus termination sequences derived from next-generation sequencing and Term-seq data should be re-evaluated given that FttA-terminated transcripts are likely to be lacking minimally ~20-30 nts from the 3'-terminus; additional RNA processing events are likely to further complicate attempts to map the 3'-termini of transcripts that reflect the true position of TEC dissociation<sup>7,38</sup>. Even transcripts derived from loci encoding putative intrinsic termination sequences should be reevaluated, as FttA-activity may influence the efficiency of intrinsic termination or serve as a backup mechanism of transcription termination for genes/operons with less-efficient intrinsic termination signals<sup>14</sup>.

The requirements for FttA-mediated termination suggest that long 5' UTRs observed for some archaeal transcripts may serve as points of regulation for premature termination upstream of the coding sequences. How transcription of stable RNAs, including rRNAs is protected from FttA-mediated termination will be of interest to determine<sup>39</sup>. Exclusion of Spt4-Spt5 from TECs transcribing stable RNAs, or structures within the nascent transcript may suffice to hinder FttA-loading or FttA-mediated termination of archaeal TECs; a delayed mechanism of Spt5 recruitment to rRNA and CRISPR loci has been identified. It is interesting that full-length FttA homologues are found in the genomes of several bacterial species, suggesting that FttA may function as a termination factor in multiple domains. It will be of immediate interest to determine if the bacterial FttA proteins can direct transcription termination, and if they can, whether they cooperate with or can substitute for rho. It will be similarly interesting to determine if FttA activity can disrupt eukaryotic TECs formed with Pol I, II and III, given the structure of FttA is nearly identical to the CPSF73 subunit of the eukaryotic CPSF

complex. The combined activities of CPSF and Xrn2 are necessary for normal termination patterns in Eukarya<sup>6,24,26,28,40-43</sup>. FttA retains all the necessary activities within a single protein: FttA can bind TECs, mediate cleavage and release of the nascent transcript and use 5'-3' exonuclease activities to degrade the 3'-transcript<sup>12</sup>. We propose that the eukaryotic CPSF complex, which minimally contains 4 homologous but non-identical subunits, arose from archaeal FttA. The ability of the CPSF complex to directly terminate transcription was likely lost during the specialization and partnership with factors that direct RNA 3'-maturation in Eukarya.

#### **4.4 Materials and Methods**

##### ***T. kodakarensis* culturing conditions.**

*T. kodakarensis* strain TS559 and derivatives of such were grown at 85°C under anaerobic conditions as previously described. DPA (Sigma) was added at neutral pH to either 12.5 or 25 mM. NaF was added to 4 mM when necessary.

##### ***Protein purifications.***

Archaeal RNA polymerases (WT and  $\Delta$ E/F variant) containing His<sub>6</sub>-HA-epitope-tagged-RpoL subunits, TBP, and TFB were purified as previously described. *T. kodakarensis* Spt5 and His<sub>6</sub>-Spt4 were purified as previously described<sup>34,44,45</sup>. Spt5 <sup>$\Delta$ NGN</sup> was purified as was full-length Spt5. WT and an H255A variant of FttA were purified from Rosetta2 *E. coli* cells carrying pQE-80L (Qiagen) expression vectors carrying the wildtype or variant TK1428 coding sequence. Cells were grown in LB medium at 37°C with shaking (~220 rpm) with 30  $\mu$ g/ml chloramphenicol and 100  $\mu$ g/ml ampicillin to an optical density at 600 nm of 0.5 before expression was induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside. Cultures were grown for an additional 3 h at 37°C with

shaking before biomass was harvested via centrifugation (~8,000 x g, 20 min, 4°C), resuspended and lysed via sonication (3 ml/g of biomass) in 20 mM Tris-HCl pH 8.0, 5 mM 2-mercaptoethanol, 10 mM MgCl<sub>2</sub>, 100 mM NaCl. Cellular lysates were clarified by centrifugation (~20,000 x g, 20 min, 4°C), heated to 85°C for 30 min to denature most host proteins, and clarified again by centrifugation (~20,000 x g, 20 min, 4°C). Heat-treated clarified cell lysates were resolved through a 5 ml HiTrap-heparin column (GE Healthcare) with a linear gradient from 0.1 – 1.0 M NaCl dissolved in 20 mM Tris-HCl pH 8.0, 5 mM 2-mercaptoethanol, 10 mM MgCl<sub>2</sub>. Fractions containing > 95% pure FttA were identified by SDS-PAGE, pooled, and dialyzed into 25 mM Tris-HCl pH 8.0, 100 mM KCl, 10 mM 2-mercaptoethanol, 50% glycerol before storage at -80°C. All protein concentrations were quantified using a Bradford Assay.

#### ***DNA templates.***

Double-stranded DNA templates used in all transcription reactions were PCR amplified from plasmids and gel purified as previously described<sup>16,46,47</sup>. All transcription templates contain a non-template 5'-strand biotin-TEG moiety to provide attachment to streptavidin-coated paramagnetic beads (Promega).

#### ***In vitro transcription assays.***

Assembly of preinitiation complexes (PICs) and elongation via NTP deprivation was carried out as described previously<sup>46</sup>. To obtain stalled TECs on G-less cassette templates, PICs were assembled using 10 nM template, 20 nM RNAP, 40 nM TBP, 40 nM TFB in a 20 µl total volume of transcription buffer (20 mM Tris-HCl pH 8.0, 250 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT) with 75 µM ApC for 3 min at 85°C before addition of 200 µM ATP, 200 µM CTP, 10 µM UTP and 10 µCi [α-<sup>32</sup>P]-UTP for 3 additional min at 85°C,

then chilled to 4°C. To obtain stalled TECs on C-less cassette templates, reactions were identical to those above, with the substitution of 200 μM GTP for 200 μM CTP. RNAP bound templates were captured with HisPur™ Ni-NTA magnetic particles (Thermo Fisher Scientific) and washed three times with 100 μl 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM KCl.

For Figure 4.1a, washed TECs were resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, with 10 μM each of ATP, CTP, and UTP before addition of 1 μM FttA or FttA<sup>H255A</sup> for 5 min at 85°C. Reactions were chilled to 4°C followed by separation of pellet and supernatant fractions by addition of streptavidin coated paramagnetic particles (Promega). Similar results were observed in 4 independent experiments. For Figure 4.1b, washed TECs (lane 1) were resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT with 10 μM each of ATP, CTP, and UTP and were incubated at 85°C for 7 min (lane 2) before being chilled on ice, bound to streptavidin-coated paramagnetic beads and washed with 100 μl 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM KCl (lane 3). Washed TECs were incubated at 85°C for 1 minute before addition of 100 μM NTPs (lane 4) or 100 μM ATP, CTP and GTP, 10μM UTP containing 1μC <sup>32</sup>P-α-UTP (lane 5) and continued incubation at 85°C for 3 minutes. Washed TECs were exposed to 1μM FttA (lane 6) at 85°C for 7 minutes before being chilled on ice, bound to streptavidin-coated paramagnetic beads and washed with 100 μl 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM KCl (lane 7). FttA-treated, washed TECs were incubated at 85°C for 1 minute before addition of 100 μM NTPs (lane 8) or 100 μM ATP, CTP and GTP, 10μM UTP containing 1μC <sup>32</sup>P-α-UTP (lane 9) and continued incubation at 85°C for 3 minutes. Washed TECs were exposed

to 50U RNase I<sub>f</sub> (lane 10) at 37°C for 7 minutes before being chilled on ice, bound to streptavidin-coated paramagnetic beads and washed with 100 µl 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM KCl (lane 11). RNase I<sub>f</sub>-treated, washed TECs were incubated at 85°C for 1 minute before addition of 100 µM NTPs (lane 12) or 100 µM ATP, CTP and GTP, 10µM UTP containing 1µC <sup>32</sup>P-α-UTP (lane 13) and continued incubation at 85°C for 3 minutes. Similar results were observed in 4 independent experiments. For Figure 4.2, panel f, washed TECs were resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, with or without 10 µM each of ATP, CTP, and UTP before addition of 1 µM FttA at 85°C. Reaction aliquots were removed after 1, 2, or 5 minutes, chilled to 4°C, then pellet and supernatant fractions were separated by addition of streptavidin coated paramagnetic particles (Promega). Similar results were observed in 4 independent experiments.

For Figures 4.2, panels b and c, washed TECs were assembled as above on G-less or C-less cassettes of various lengths. Washed TECs were resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, with or without 1 µM FttA at 85°C for 3 minutes. Reactions were chilled to 4°C, then pellet and supernatant fractions were separated by addition of streptavidin coated paramagnetic particles (Promega).

For Figure 4.2b-c similar results were observed in 3 independent experiments.

For Figures 4.3 and 4.4, washed TECs were assembled with WT or ΔE/F RNAP as above on a +125 G-less cassette template. Continued elongation was permitted by the addition of 0, 1, 10 or 100 µM NTPs in the presence or absence of combinations of 6 µM Spt4, 6 µM Spt5, 6 µM Spt5<sup>ΔNGN</sup>, 1 µM FttA or FttA<sup>H255A</sup>. After 5 min at 85°C, reactions were chilled to 4°C followed by separation of pellet and supernatant fractions

by addition of streptavidin coated paramagnetic particles (Promega). For Figure 4.3c, similar results were observed in 3 independent experiments. Figure 4.4 was performed once.

For Figure 4.6b, stalled TECs on a G-less cassette template assembled using 10 nM template, 20 nM RNAP, 40 nM TBP, 40 nM TFB in a 20  $\mu$ l total volume of transcription buffer (20 mM Tris-HCl pH 8.0, 250 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT) with 75  $\mu$ M ApC for 3 min at 85°C before addition of 200  $\mu$ M ATP, 200  $\mu$ M CTP, 10  $\mu$ M UTP and 10  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]-UTP and incubation for 3 additional min at 85°C, then chilled to 4°C. RNAP bound templates were captured with HisPur™ Ni-NTA magnetic particles (Thermo Fisher Scientific) and washed three times with 100  $\mu$ l 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM KCl. Washed TECs were resuspended in 10 mM Tris-HCl pH 8.0, 125 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, with 10  $\mu$ M each of ATP, CTP, and UTP before addition of reaction buffer (10 mM Tris-HCl pH 8.0, 120 mM KCl, 8 mM DTT, and 1.25 mM MgCl<sub>2</sub>) +/- 25 mM DPA and/or +/- 1  $\mu$ M FttA at 85°C. Reaction aliquots were incubated for 3 minutes, and then chased with 250  $\mu$ M ATP, CTP, UTP, and GTP for 2 minutes to allow for elongation to +225. Reactions were chilled to 4°C, then pellet and supernatant fractions were separated by addition of streptavidin coated paramagnetic particles (Promega). Similar results were observed in 5 independent experiments.

Radiolabeled transcripts from transcription experiments were recovered by addition of 5 volumes STOP buffer (600 mM Tris-HCl pH 8.0, 30 mM EDTA) containing 7  $\mu$ g of tRNA (total), equal volume phenol/chloroform/isoamyl alcohol (25:24:1, V/V) extractions, and precipitations of the aqueous phase with 2.6 volumes 100% ethanol. Precipitated transcripts were resuspended in 95% formamide, 1X TBE, heated to 99°C

for 2 minutes, rapidly chilled on ice, loaded and resolved in 10-20% polyacrylamide, 8M urea, 1X TBE denaturing gels. Radiolabeled RNA was detected using phosphorimaging (GE Healthcare). Gel images were analyzed using GE Imagequant 5.2 software.

***Western blot analysis of RNAP release to solution.***

Anti-HA antibodies (BioLegend 901513) were employed as previously described to quantify RpoL-levels in P and S fractions.

***Western blot analysis of FttA protein.***

Purified, recombinant full-length FttA was used as an antigen to prepare polyclonal antibodies in mice (Cocalico Biologicals). Known quantities of purified FttA were resolved as comparative quantification standards in adjacent lanes to clarified cell lysates derived from known quantities of cells. Proteins were separated via SDS-PAGE, transferred to PVDF membranes, and probed with primary anti-FttA antibodies. Blots were developed by addition of an IgG-AP conjugated anti-mouse secondary antibody allowed for detection by NBT/BCIP (Roche). A linear regression FttA signal intensity to FttA amount in ng was generated.

***Construction of strains TK1428D and IR5.***

*T. kodakarensis* strains used here were constructed from the parental strains TS559 as previously described<sup>31,34</sup>.

***Purification of FttA directly from lysates of strain TK1428D and MuDPIT analysis.***

Purification procedures and MuDPIT analysis was performed largely as previously described<sup>35,36</sup>. Briefly, 2 L early-exponential phase cultures ( $OD_{600\text{ nm}} = \sim 0.3$ ) of *T. kodakarensis* strains TK1428D and TS559 grown in ASW-YT-S° medium at 85°C were rapidly chilled to 4°C and harvested by centrifugation (20,000 x g). All subsequent

procedures were carried out at 4°C and completed as rapidly as possible to retain native *in vivo* protein-protein interactions. Cells from each strain were individually resuspended in 3 ml of 25 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol per gram of wet biomass and lysed by repetitive sonication. The resulting lysates were clarified via centrifugation (20,000 x g) before passage through 5-ml HiTRAP chelating columns (GE Healthcare) pre-charged with NiSO<sub>4</sub> and equilibrated in 25 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol. The lysate components that did not bind the columns (e.g. flowthrough) were discarded and the columns were washed with ~20 column volumes 25 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol until no additional proteins eluted. Bound proteins were eluted using a linear gradient from (initial) 25 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol to (final) 25 mM Tris-HCl pH 8, 100 mM NaCl, 150 mM imidazole, 10% glycerol. Fractions that contained the tagged TK1428 protein were identified by Western blotting, pooled and were twice dialyzed in 3 Kda molecular weight cut-off dialysis tubing against 1 L 25 mM Tris-HCl pH 8, 500 mM NaCl, 0.5 mM EDTA, 2 mM dithiothreitol. FttA was not identified by Western blots within fractions resultant from processing TS559 biomass, but identical fractions were collected and processed to identify native *T. kodakarensis* present in such fractions that spontaneously bind the chelating resin. Protein concentrations were determined by Bradford assays, and thirty-microgram aliquots of the proteins present in solution were precipitated by adding trichloroacetic acid (TCA; 15% final concentration). The TCA-precipitated proteins were identified by multidimensional protein identification technology at the Ohio State University mass spectrometry facility (<https://www.ccic.osu.edu/MSP>) using the MASCOT search engine. We required a minimum of two unique peptide fragments to

positively identify a protein. TS559 control samples identified several proteins that bound and eluted from the Ni<sup>2+</sup>-charged matrix in the absence of a His<sub>6</sub>-tagged protein. All of the proteins identified in the experimental samples that had MASCOT scores of >100 and were not also present in the control samples are shown in Figure 4.8.

***TRizol-based RNA purifications from *T. kodakarensis* cultures and q-RT-PCR.***

RNA extractions were performed essentially as previously described from strain TS559 prior to, or 1 hour after DPA addition<sup>48</sup>. RNA extractions from IR5 and TS559, in the absence and presence of 4 mM NaF were performed as previously described. qRT-PCR reactions were performed as previously described, except that 500 ng of total RNA was used during cDNA synthesis.

## REFERENCES

1. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription Regulation in Archaea. *J. Bacteriol.* **198**, 1906–1917 (2016).
2. Grohmann, D., Hirtreiter, A. & Werner, F. Molecular mechanisms of archaeal RNA polymerase. *Biochem. Soc. Trans.* **37**, 12–17 (2009).
3. Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J. & Werner, F. The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol. Cell* **43**, 263–274 (2011).
4. Svetlov, V. & Nudler, E. Towards the unified principles of transcription termination. *EMBO J.* **39**, (2020).
5. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.06.020
6. Eaton, J. D., Francis, L., Davidson, L. & West, S. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145 (2020).
7. Dar, D., Prasse, D., Schmitz, R. A. & Sorek, R. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat. Microbiol.* **1**, 16143 (2016).
8. Berkemer, S. J., Maier, L. K., Amman, F., Bernhart, S. H., Wörtz, J., Märkle, P., Pfeiffer, F., Stadler, P. F. & Marchfelder, A. Identification of RNA 3' ends and

- termination sites in *Haloferax volcanii*. *RNA Biol.* **17**, 663–676 (2020).
9. Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E. & Santangelo, T. J. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* **5**, 545–553 (2020).
  10. Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T. & Reeve, J. N. Polarity in Archaeal Operon Transcription in *Thermococcus kodakaraensis*. *J. Bacteriol.* **190**, 2244–2248 (2008).
  11. Santangelo, T. J., Cubonova, L., Skinner, K. M. & Reeve, J. N. Archaeal Intrinsic Transcription Termination In Vivo. *J. Bacteriol.* **191**, 7102–7108 (2009).
  12. Phung, D. K., Rinaldi, D., Langendijk-Genevaux, P. S., Quentin, Y., Carpousis, A. J. & Clouet-d'Orval, B. Archaeal  $\beta$ -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res.* **41**, 1091–1103 (2013).
  13. Silva, A. P. G., Chechik, M., Byrne, R. T., Waterman, D. G., Ng, C. L., Dodson, E. J., Koonin, E. V., Antson, A. A. & Smits, C. Structure and activity of a novel archaeal  $\beta$ -CASP protein with N-terminal KH domains. *Structure* **19**, 622–632 (2011).
  14. Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., Li, L., Zheng, X. & Dong, X. The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. *Nucleic Acids Res.* **48**, (2020).
  15. Maier, L.-K. & Marchfelder, A. It's all about the T: transcription termination in archaea. *Biochem. Soc. Trans.* **47**, 461–468 (2019).
  16. Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal

- transcription termination. *Proc. Natl. Acad. Sci.* **114**, E6767–E6773 (2017).
17. Strauß, M., Vitiello, C., Schweimer, K., Gottesman, M., Rösch, P. & Knauer, S. H. Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res.* **44**, 5971–5982 (2016).
  18. Roberts, J. W. Mechanisms of Bacterial Transcription Termination. *J. Mol. Biol.* **431**, 4030–4039 (2019).
  19. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85**, 319–47 (2016).
  20. Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**, 935–8 (2008).
  21. Santangelo, T. J. & Roberts, J. W. Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol. Cell* **14**, 117–26 (2004).
  22. Hart, C. M. & Roberts, J. W. Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *J. Biol. Chem.* **266**, 24140–8 (1991).
  23. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46 (2012).
  24. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K. & Bentley, D. L. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo”

- Mechanism. *Mol. Cell* **76**, 896-908.e4 (2019).
25. Kolev, N. G., Yario, T. A., Benson, E. & Steitz, J. A. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO Rep.* **9**, 1013–1018 (2008).
  26. Luo, W. & Bentley, D. A Ribonucleolytic Rat Torpedoes RNA Polymerase II. *Cell* **119**, 911–914 (2004).
  27. Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K. C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J. & Cramer, P. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* **66**, 38-49.e6 (2017).
  28. Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., Nguyen, T., Karp, S. & Bentley, D. L. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell* **60**, 256–267 (2015).
  29. Lawson, M. R. & Berger, J. M. Tuning the sequence specificity of a transcription terminator. *Curr. Genet.* **65**, 729–733 (2019).
  30. Horsfall, L. E., Garau, G., Liénard, B. M. R., Dideberg, O., Schofield, C. J., Frère, J. M. & Galleni, M. Competitive inhibitors of the CphA metallo- $\beta$ -lactamase from *Aeromonas hydrophila*. *Antimicrob. Agents Chemother.* **51**, 2136–2142 (2007).
  31. Speed, M. C., Burkhart, B. W., Picking, J. W. & Santangelo, T. J. An Archaeal Fluoride-Responsive Riboswitch Provides an Inducible Expression System for Hyperthermophiles. *Appl. Environ. Microbiol.* **84**, (2018).
  32. Sarmiento, F., Mrazek, J. & Whitman, W. B. Genome-scale analysis of gene

- function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl. Acad. Sci.* **110**, 4726–4731 (2013).
33. Santangelo, T. J., Cubonová, L. & Reeve, J. N. *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* **76**, 1044–52 (2010).
  34. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *BIO-PROTOCOL* **7**, (2017).
  35. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* **81**, 897–911 (2011).
  36. Burkhart, B. W., Febvre, H. P. & Santangelo, T. J. Distinct Physiological Roles of the Three Ferredoxins Encoded in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *MBio* **10**, (2019).
  37. Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J. & Whitaker, R. J. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.* **9**, 4908 (2018).
  38. Berkemer, S. J., Maier, L. K., Amman, F., Bernhart, S. H., Wörtz, J., Märkle, P., Pfeiffer, F., Stadler, P. F. & Marchfelder, A. Identification of RNA 3' ends and termination sites in *Haloferax volcanii*. *RNA Biol.* **17**, 663–676 (2020).
  39. Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA

- polymerase. *Nat. Microbiol.* **2**, 17021 (2017).
40. Pettinati, I., Grzechnik, P., Ribeiro de Almeida, C., Brem, J., McDonough, M. A., Dhir, S., Proudfoot, N. J. & Schofield, C. J. Biosynthesis of histone messenger RNA employs a specific 3' end endonuclease. *Elife* **7**, (2018).
  41. Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).
  42. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
  43. Cardiello, J. F., Goodrich, J. A. & Kugel, J. F. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. *Mol. Cell. Biol.* **38**, (2018).
  44. Hileman, T. H. & Santangelo, T. J. Genetics techniques for *Thermococcus kodakarensis*. *Front. Microbiol.* **3**, 195 (2012).
  45. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic Techniques for the Archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).
  46. Gehring, A. M. & Santangelo, T. J. in *Methods Mol. Biol.* **1276**, 263–279 (2015).
  47. Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019). doi:10.1111/mmi.14191
  48. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N.

Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15**, 684 (2014).

## CHAPTER 5: DEVELOPING ARCHAEAL NASCENT ELONGATING TRANSCRIPT SEQUENCING

### 5.1 Summary

Total transcriptome profiling by way of 'Next Generation Sequencing' or NGS RNA sequencing (RNA-seq) has revolutionized the study of gene expression<sup>1</sup>. However, bulk RNA-seq experiments only give a steady-state view of the transcriptome. RNAs are subject to processing and degradation after being transcribed/translated<sup>2-10</sup>. This processing and degradation results in the loss of information, particularly regulatory points of transcription initiation, pausing, elongation and termination. Newer techniques such as precision nuclear run-on sequencing (PRO-seq) and native or nascent elongating transcript sequencing (NET-seq) overcome this problem by facilitating the affinity capture of active transcription complexes and subsequent RNA-seq<sup>5,6,11-13</sup>. PRO-seq and NET-seq additionally give strand specific, single nucleotide resolution positional information on the active site of RNA polymerase (RNAP). While these techniques are prevalent in eukaryotic and bacterial systems, no such technique has been developed for the Archaea. Given the ease of epitope tagging RNAP in the model archaeal organism *Thermococcus kodakarensis*, the affinity capture of active transcription elongation complexes (TECs) for use in NET-seq was an obvious choice for the advancement of transcriptome profiling in the Archaea<sup>14,15</sup>. Further development of archaeal NET-seq can be used for determining points of RNAP regulation in the context Transcription Factor S (TFS) induced backtracking, histones and chromatin structure, factor and sequence dependent transcription termination, the coupling of transcription and translation and environmental shifts. By tagging RNAP subunit RpoL

(encoded by TK1167) with a 6xHIS-HA affinity/epitope, capture and confirmation of active TECs is possible<sup>16</sup>. Subsequent purification of the associated RNA allows for high-throughput sequencing and transcriptome profiling. Archaeal NET-seq and important validating questions are described below.

## 5.2 Results

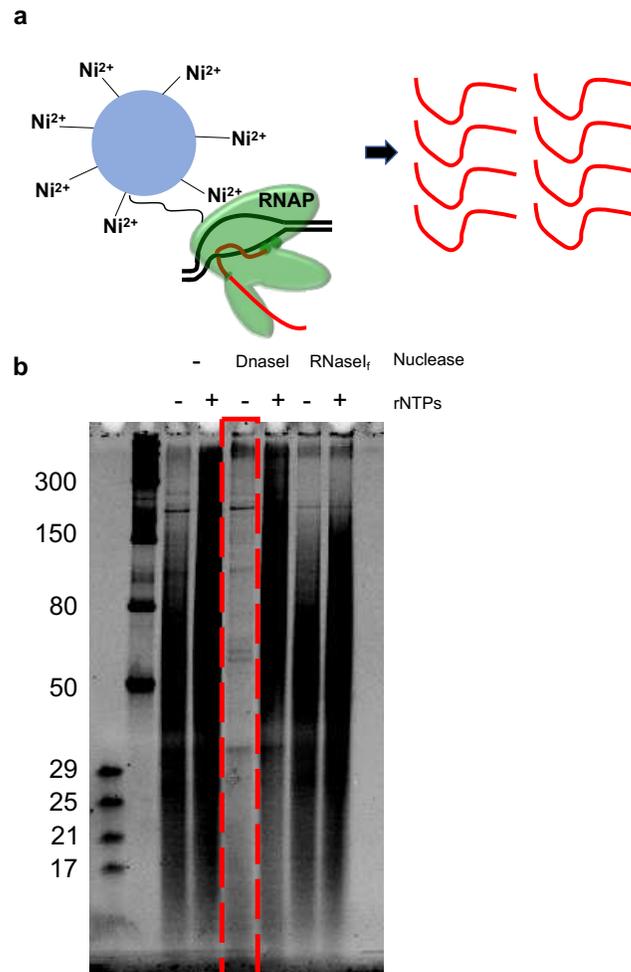
### ***Affinity capture and validation of active transcription elongation complexes for *T. kodakarensis*.***

NET-seq relies on the affinity capture of RNAP from cell lysates and the digestion of unprotected DNA and RNA, followed by the purification of RNAP footprinted RNA and sequencing to map global RNAP distribution. To capture RNAP, TK1167 encoding the RpoL subunit of RNAP was modified to encode a c-terminal 6xHis affinity tag preceded by an HA epitope tag (as previously described) thus generating strain TK1167C<sup>15,17,18</sup>. Due to the native quality of NET-seq (no chemical crosslinking of proteins to DNA before affinity capture) the rapid collection and freezing of cells is essential to maintain an accurate 'snapshot' of RNAP distribution across the genome. To ensure captured TECs reflect in vivo RNAP distribution, mid, log-phase TK1167C cultures were rapidly pelleted at 4° C and then flash frozen in liquid nitrogen. The gentle lysis of cells and the maintenance of the collected supernatant at 4° C sufficiently dilutes NTP concentration and prevents elongation of TECs or their degradation by other cellular components. Treating the TEC containing supernatant with DNaseI and RNaseA ensured rapid breakdown of unprotected DNA, RNA and rRNA. TECs were then captured by the addition of Ni<sup>2+</sup>-conjugated agarose particles and thoroughly washed. Captured and washed TECs could then be further footprinted by the addition of RNaseI<sub>f</sub>, ensuring the

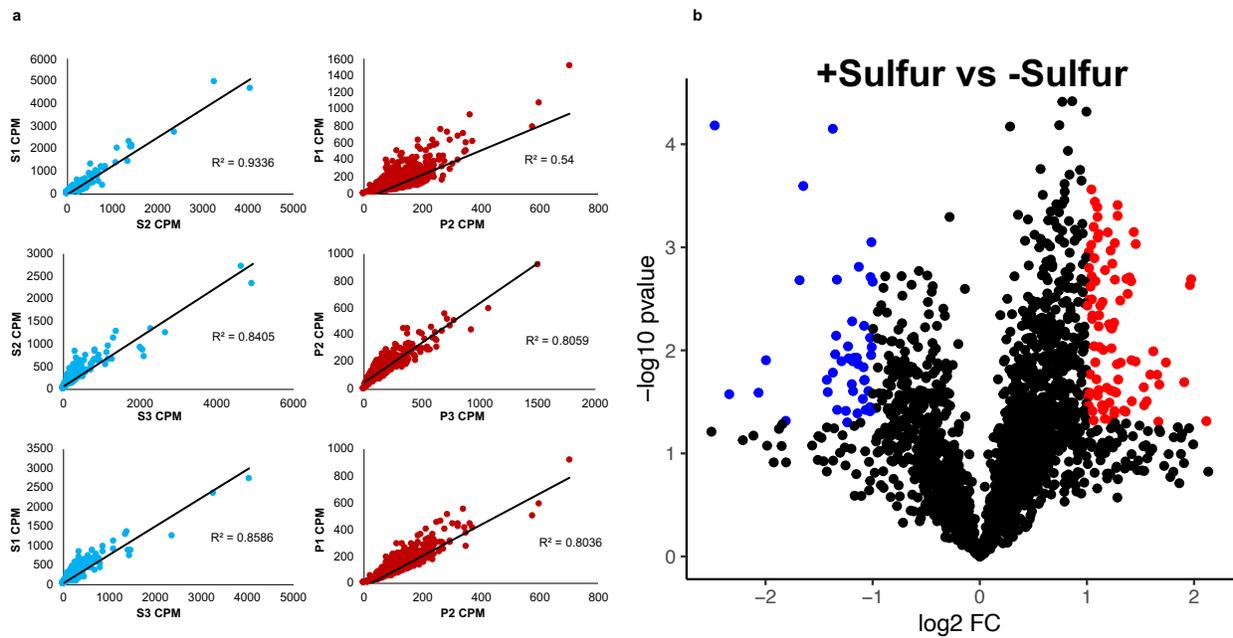
RNA population contained only RNAs which were encapsulated by RNAP (Figure 5.1a). To further validate the RNAs returned from TEC affinity capture, RNAP activity was assayed (Figure 5.1b). Captured, washed TECs incubated with NTPs resulted in the elongation of a non-homogenous RNA population. When the purified nucleic acids from this reaction were treated with DNaseI, a shift in population size (i.e. shorter to longer, representative of elongation) RNAs were still observed. Additionally, when treated with RNaseA the purified nucleic acids showed no shift in size, confirming active elongation of RNA by RNAP on a DNA template.

***A method for testing differential RNAP positioning across the *T. kodakarensis* genome.***

Growth of *T. kodakarensis* in media containing or lacking sulfur results in a distinct and characterized expression profile<sup>19,20</sup>. Upregulation and downregulation of the membrane bound sulfur-reductase (MBS) complex is seen in the presence and absence of sulfur, respectively. To test the viability of NET-seq to identify RNAP distribution across the genome, TK1167C was grown in triplicate in artificial seawater yeast and tryptone media (ASW-YT) containing sulfur (S) or lacking sulfur (P). Cultures were pelleted as described above and RNA from active TECs was purified and sent to Novogene for RNA-seq library construction and sequencing on an Illumina Hi-seq platform. The resulting reads were trimmed of adaptor sequences, aligned to the *T. kodakarensis* genome and visualized at loci-specific coordinates. To first assess the reproducibility of NET-seq in *T. kodakarensis*, each replicate's whole transcriptome (+ or -sulfur, normalized cpm) was compared using a linear regression (Figure 5.2a). Next, the expression profiles of + or -sulfur were compared (Figure 5.2b). Notably, this



**Figure 5.1 NET-seq allows for the capture of RNA and DNA foot-printed TECs. (a)** NET-seq utilizes footprinted RNAs from affinity captured TECs. **(b)** Digestion of total nucleic acids attained from TEC affinity capture demonstrates that only the RNA is elongated and thus the captured complex is an active TEC.



**Figure 5.2** The data generated from NET-seq is highly reproducible. (a) Strains grown in sulfur (S1-S3) are shown in blue while strains grown in the absence of sulfur (P1-P3) are shown in red. (b) There is little differential regulation between strains grown in the presence or absence of sulfur. Genes significantly upregulated in the presence of sulfur are shown in blue while those shown in red represent genes upregulated in the absence of sulfur.

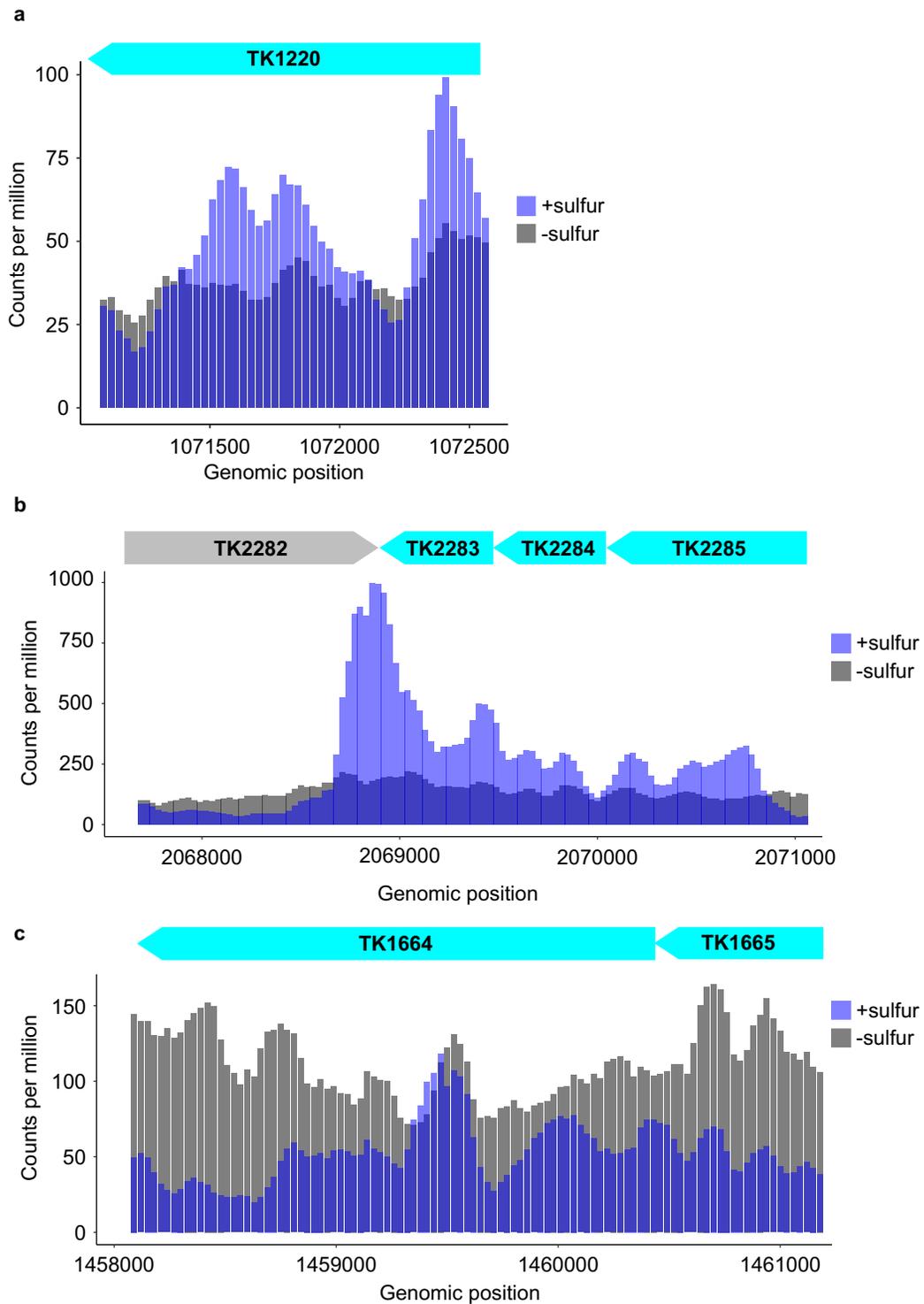
data was reproducible and showed little variation in expression between + and -sulfur cultures.

The main expression difference expected in +/-sulfur was predicted to be in genes related specifically to the MBS complex. A notable increase in RNAP distribution was observed at the start of the MBS operon when cultures were grown in the presence of sulfur (Figure 5.3a). The 'spikiness' of this data is a quality typically observed in NET-seq<sup>12</sup>. Additionally, several loci corresponding to operons of unknown function showed differential RNAP distribution (Figure 5.3b), suggesting additional regions of interest controlling sulfur-based metabolism in *T. kodakarensis*.

### **5.3 Discussion**

#### ***Further validation of NET-seq is required.***

The value of NET-seq stems from being able to place the active site of RNAP globally across the genome at single nucleotide resolution. Given the current coverage of NET-seq (~20X coverage), the RNAP active site is easily determined. However, two concerns remain. First, there is a large portion of reads which align to the antisense strand. While a high level of antisense transcription is predicted in archaeal systems, roughly 50% of all aligned reads align to the antisense strand<sup>7,8,19</sup>. This problem may have stemmed from the library construction at Novogene which was specially modified to accommodate the potentially tiny RNA fragments (<20nt) returned from purified and digested TECs. To overcome this issue, long read sequencing of non-RNaseI<sub>f</sub> digested RNA purified from TECs is necessary. The advent of Nanopore sequencing as a quick and localized NGS technology will allow for the rapid and long-read sequencing of these RNAs. The retention or absence of antisense transcript alignment will either validate or



**Figure 5.3 NET-seq has revealed unique points of regulation.** (a) The MBS operon shows differential RNAP density in the presence and absence of sulfur. (b and c) Unique regulatory patterns of RNAP densities are observed in the presence and absence of sulfur at a variety of loci.

refute the current protocol.

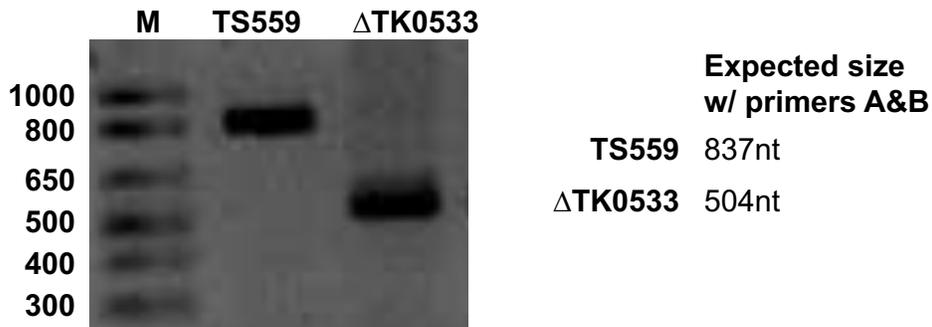
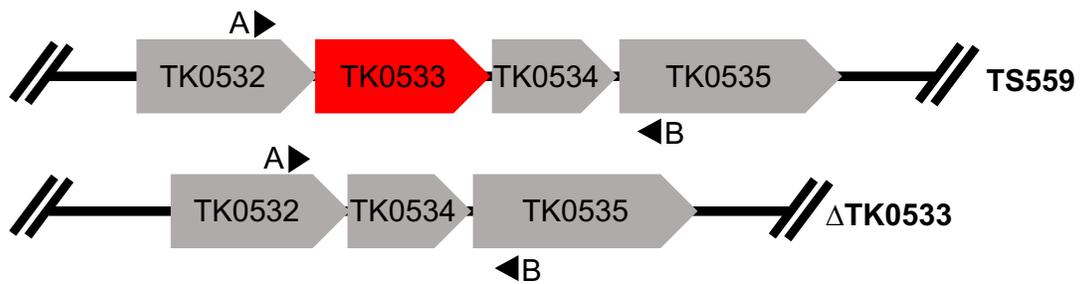
Secondly, the method used to capture TECs may not be quick enough. A short time (<5 minutes) to go from actively growing culture to frozen cell pellet is important to maintain an accurate snapshot of RNAP distribution. Recent work on ribosome profiling in Archaea has shown the importance of rapidly freezing biomass to maintain the 3-nucleotide periodicity exhibited by the ribosome. It may be necessary to establish a faster way to pellet *T. kodakarensis* cells to achieve the most accurate snapshot of RNAP distribution.

***Using NET-seq to characterize coupled transcription translation, the influence of histones and regulation of pausing and backtracking during transcription.***

While testing cultures in the presence or absence of sulfur offers valuable metabolic information, NET-seq should be leveraged to characterize regulatory components which directly influence RNAP. The observation of both polarity and coupled transcription-translation in archaeal and bacterial systems suggests the translating ribosome plays a role in the regulation of transcription<sup>21–23</sup>. NET-seq protocols are prevalent in bacterial systems and ribosome foot-printing protocols are present in both bacterial and archaeal systems<sup>24,25</sup>. However, the coupled complex has yet to be globally profiled. The addition of an affinity tag such as a 3xFLAG or GST tag to the small ribosomal protein encoded by TK1695, in tandem with current RNAP affinity tags would allow the purification of a coupled complex. The RNAs purified from this complex could be globally compared to those from NET-seq and ribosome foot-printing experiments giving a global view of RNAs that are transcribed, RNAs that are translated

and RNAs that transcribed and translated in tandem. Unique regulatory mechanisms will likely be revealed through this experimentation.

Beyond the ribosome, both Transcription Factor S (TFS) and histone proteins greatly influence transcription of RNAP<sup>26-28</sup>. *T. kodakarensis* offers a genetic system that allows for the iterative modification of the genome. By constructing strains in which chromatin structure is disrupted and RNAP is tagged, the influence of 3-dimensional chromatin structure on transcription may be assayed. More recently, TFS has been deleted from the *T. kodakarensis* genome to our surprise (Figure 5.4). The subsequent tagging of RNAP in this strain will allow for the study of backtracking and RNA cleavage in archaeal systems. Combinations of these systems as well as others will provide an extensive and global picture of RNAP regulation.



**Figure 5.4 TFS, encoded by TK0533 is nonessential.** By tagging the RpoL subunit of RNAP in  $\Delta$ TK0533, the role of TFS in global RNAP backtracking and endonucleolytic cleavage.

## 5.4 Materials and Methods

### ***Strain construction and growth of *T. kodakarensis*.***

Strains of *T. kodakarensis* were constructed as previously described<sup>15,17,18</sup>. For isolation of TECs, Strain TK1167C was grown in 1 l of ASW-YT-S or 1 l of ASW-YT-P for 8 hours ( $OD_{600} \sim 0.6$ ) as previously described. Cells were pelleted at 10,000 rpm for 3 minutes at 4° C then rapidly frozen by submerging pellets in liquid nitrogen. Pellets could then be stored at -80° C until further use.

### ***Capture and purification of TEC-associated RNAs.***

The ~1.0 g pellets were resuspended in 3.25 ml 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl, 0.1% TritonX 100, and 0.2 mM PMSF on ice. To rapidly lyse cells and degrade ribosomal RNA, 400 kU of ReadyLyse lysozyme and 50 µg of RNase A was added and allowed 5 minutes at room temperature. Following lysis, 900 µg of heparin was added to capture unbound RNAP while 100 U of DNaseI with 10X DNaseI buffer (100 mM MnCl<sub>2</sub> and 100 mM Tris-HCl pH 7.5) to a final concentration of 1X was added to digest unprotected chromosomal DNA for 10 minutes at room temperature. The reaction was then centrifuged at 20,000 xg for 3 minutes and the resulting supernatant was kept on ice. 0.5 ml of Ni-NTA agarose particles (Qiagen) were washed 3 times in 1 ml of binding buffer (0.5 M NaCl, 5 mM imidazole, 5% glycerol, 20 mM Tris-HCl pH 8.0 and 1 mM beta-mercapto ethanol). Washed particles were resuspended in 1 mL of binding buffer and subsequently split into 5 equal aliquots (200 µl each) and brought up to 800 µl with binding buffer. 800 µl of supernatant was added to 800 µl of washed particles in a 1.7 mL Eppendorf tube and incubated at 4° C for 15 minutes with gentle nutation to capture TECs. Working in a cold room, particles were pelleted (30

second spin at 2,500 xg) and washed 4 times with wash buffer (1 M NaCl, 15 mM imidazole, 5% glycerol, 20 mM Tris-HCl pH 8.0 and 1 mM beta-mercapto ethanol). Particles were then washed 2 times in nuclease buffer (40 mM KCl, 15 mM imidazole, 20 mM Tris-HCl pH 8.0, 0.3 mM MgCl<sub>2</sub>, 5% glycerol and 1 mM beta-mercapto ethanol). Resuspension and incubation of particles in 1 ml of nuclease buffer and the addition of 80 U of DNaseI, 6 U of RNaseI and MgCl<sub>2</sub> to a final concentration of 8 mM at 37° C for 7 minutes allowed the foot-printing of the captured TECs. Particles were washed 4 times in nuclease buffer lacking MgCl<sub>2</sub> and then TECs were eluted with the addition of 600 µl of elution buffer (40 mM KCl, 100 mM imidazole, 5% glycerol, 20 mM Tris-HCl pH 8.0 and 1 mM beta-mercapto ethanol). Total nucleic acids associated with TECs were emulsion extracted with an equal volume of 25/24/1 phenol/chloroform/isoamyl alcohol. The nucleic acids in the aqueous layer following clarification at 15,000 rpm was added to 2.6X volumes of 100 % EtOH and incubated at -80° C for 1 hour. Precipitated nucleic acids were pelleted at 15,000 rpm for 30 minutes at 4° C. The pellet was treated with 10 U of DNaseI and sent to Novogene for sequencing following an additional P/C/I clean-up.

#### ***Confirmation of TEC activity.***

To confirm the activity of captured and foot-printed TECs, the elution was added to transcription components (1X transcription buffer, 100 mM KCl, 6 mM MgCl<sub>2</sub> and 250 µM ATP/GTP/CTP/UTP – all final concentrations) and allowed elongation at 85° C for 5 minutes. Nucleic acids were extracted as described previously<sup>26,29,30</sup> and subsequently treated with either 10 U of DNaseI, RNase A or no nuclease followed by equal-volume P/C/I extraction and ethanol precipitation as described above. Pellets were

resuspended in formamide loading buffer and resolved on 15% UREA-PAGE and stained in 2X SYBR Gold stain.

### ***Data analysis.***

Fastq files containing raw sequencing reads from Novogene were trimmed of adaptors and barcode sequences using the fastp command with the -i option. Trimmed reads were aligned to the *T. kodakarensis* TS559 genome using bowtie2. The output SAM files were then converted to BAM files using the samtools. BAM files were converted into bedfiles using bedtools with options -s or -S to specifically obtain either the sense or antisense strand coverage. Coverage bins of 30nt were chosen as previous foot printing of the RNA within RNAP demonstrated ~30nt of RNA protection.

### ***Example pipeline***

*Trim adaptors using fastp:*

```
fastp -i P1_1.fq.gz -l P1_2.fq.gz -o P1_1_trimmed.fq.gz -O P1_2_trimmed.fq.gz
```

*Align trimmed, forward and reverse read to TS559 genome. Need to be in reference genome drive:*

```
bowtie2 -p 12 -x TS559 -1
```

```
/projects/evo\@colostate.edu/NET_seq/Net_seq_data/usftp21.novogene.com/raw_data/
```

```
P1/P1_1_trimmed.fq.gz -2
```

```
/projects/evo\@colostate.edu/NET_seq/Net_seq_data/usftp21.novogene.com/raw_data/
```

```
P1/P1_2_trimmed.fq.gz -S P1_paired.sam
```

*Convert sam to bam file, sort and index:*

```
samtools view -b -S P1_paired.sam > P1_paired.bam
```

```
samtools sort P1_paired.bam -o P1_paired.sort
```

```
samtools index P1_paired.sort
```

*Generate bam file for specific loci of interest (MBS here):*

```
samtools view -b -h P1_paired.sort "TS559_Genomic_Sequence.seq:1064000-1074500" > P1_MBS.bam
```

*Generate 30nt bin TS559 reference file*

```
python3 makebedfile.py 2 2080000 --name TS559_Genomic_Sequence.seq --size 30 --step 30 > TK_genome_30step
```

*Generate bedfile and force certain strand coverage (either sense or antisense):*

```
bedtools coverage -s -a TK_genome_30step -b P1_MBS.bam >
```

```
P1_MBS_30StepSense.bed
```

```
bedtools coverage -S -a TK_genome_30step -b P1_MBS.bam >
```

```
P1_MBS_30StepAnti.bed
```

### **Data visualization in R**

With all files of interest generated, R was used to visualize loci specific coverage.

The following custom R script was used:

```
library(tidyverse)
```

```
library(reshape2)
```

```
#OPEN PYR SENSE AND ANTI-SENSE MBS FILES, SLICE THEM FOR THE  
OPERON, SELECT DESIRED COLUMNS
```

```
p1_sense_MBS <-
```

```
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P1_
```

```

MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P1S','x','y','z')) %>%

  select(genome, start, stop, strand, P1S) %>%

  slice(35461:35822)

p1_anti_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P1_
MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P1A','x','y','z')) %>%

  select(genome, start, stop, strand, P1A) %>%

  slice(35461:35822)

p2_sense_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P2_
MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P2S','x','y','z')) %>%

  select(genome, start, stop, strand, P2S) %>%

  slice(35461:35822)

p2_anti_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P2_
MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P2A','x','y','z')) %>%

  select(genome, start, stop, strand, P2A) %>%

  slice(35461:35822)

```

```

p3_sense_MBS <-
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P3_
MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P3S','x','y','z')) %>%
  select(genome, start, stop, strand, P3S) %>%
  slice(35461:35822)

p3_anti_MBS <-
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/P3_
MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'P3A','x','y','z')) %>%
  select(genome, start, stop, strand, P3A) %>%
  slice(35461:35822)

```

#OPEN SULF SENSE AND ANTI-SENSE MBS FILES, SLICE THEM FOR THE  
OPERON, SELECT DESIRED COLUMNS

```

s1_sense_MBS <-
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S1_
MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S1S','x','y','z')) %>%
  select(genome, start, stop, strand, S1S) %>%
  slice(35461:35822)

s1_anti_MBS <-
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S1_

```

```

MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S1A','x','y','z')) %>%

  select(genome, start, stop, strand, S1A) %>%

  slice(35461:35822)

s2_sense_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S2_
MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S2S','x','y','z')) %>%

  select(genome, start, stop, strand, S2S) %>%

  slice(35461:35822)

s2_anti_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S2_
MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S2A','x','y','z')) %>%

  select(genome, start, stop, strand, S2A) %>%

  slice(35461:35822)

s3_sense_MBS <-

read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S3_
MBS_30StepSense.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S3S','x','y','z')) %>%

  select(genome, start, stop, strand, S3S) %>%

  slice(35461:35822)

```

```

s3_anti_MBS <-
read_delim('/Users/travis/Desktop/NET_seq/Net_seq_data/MBH_and_MBS_strand/S3_
MBS_30StepAnti.bed', delim = '\t', col_names = c('genome', 'start', 'stop', 'bin', 'zero',
'strand', 'S3A','x','y','z')) %>%
  select(genome, start, stop, strand, S3A) %>%
  slice(35461:35822)

```

#MERGE PYR DATA TO MAKE SENSE AND ANTISENSE MBS DF

```

pyr_MBS_df <- p1_sense_MBS %>% merge(y= p2_sense_MBS, by=
c('genome','start','stop','strand')) %>%
  merge(y= p3_sense_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= p1_anti_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= p2_anti_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= p3_anti_MBS, by= c('genome','start','stop','strand'))

```

#MERGE SULF DATA TO MAKE SENSE AND ANTISENSE MBS DF

```

sulf_MBS_df <- s1_sense_MBS %>% merge(y= s2_sense_MBS, by=
c('genome','start','stop','strand')) %>%
  merge(y= s3_sense_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= s1_anti_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= s2_anti_MBS, by= c('genome','start','stop','strand')) %>%
  merge(y= s3_anti_MBS, by= c('genome','start','stop','strand'))

```

```
head(sulf_MBS_df)
```

```
#CPM
```

```
pyr1_mapped_reads <- 38023339/1000000
```

```
pyr2_mapped_reads <- 41955275/1000000
```

```
pyr3_mapped_reads <- 36251480/1000000
```

```
sulf1_mapped_reads <- 31192263/1000000
```

```
sulf2_mapped_reads <- 39891004/1000000
```

```
sulf3_mapped_reads <- 46512732/1000000
```

```
#NORMALIZE PYRUVATE MBS DATA VIA CPM AND SELECT CPM COLUMNS AND  
CALCULATE MEAN
```

```
pyr_norm_MBS_df <- pyr_MBS_df %>% mutate('P1S_CPM' =
```

```
P1S/pyr1_mapped_reads) %>%
```

```
  mutate('P2S_CPM' = P2S/pyr2_mapped_reads) %>%
```

```
  mutate('P3S_CPM' = P3S/pyr3_mapped_reads) %>%
```

```
  mutate('P1A_CPM' = P1A/pyr1_mapped_reads) %>%
```

```
  mutate('P2A_CPM' = P2A/pyr2_mapped_reads) %>%
```

```
  mutate('P3A_CPM' = P3A/pyr3_mapped_reads) %>%
```

```
  select(genome,start,stop,P1S_CPM, P2S_CPM, P3S_CPM, P1A_CPM, P2A_CPM,  
P3A_CPM) %>%
```

```

rowwise() %>%
mutate('PS_average' = mean(c(P1S_CPM, P2S_CPM, P3S_CPM))) %>%
rowwise() %>%
mutate('PA_average' = mean(c(P1A_CPM, P2A_CPM, P3A_CPM))) %>%
#MUTATE PA_AVERAGE TO BE NEGATIVE FOR THE GRAPH
mutate('PA_average'= PA_average*-1) %>%
select(start, PS_average, PA_average)

```

```

#NORMALIZE SULF MBS DATA VIA CPM AND SELECT CPM COLUMNS AND
CALCULATE MEAN

```

```

sulf_norm_MBS_df <- sulf_MBS_df %>% mutate('S1S_CPM' =
S1S/sulf1_mapped_reads) %>%
mutate('S2S_CPM' = S2S/sulf2_mapped_reads) %>%
mutate('S3S_CPM' = S3S/sulf3_mapped_reads) %>%
mutate('S1A_CPM' = S1A/sulf1_mapped_reads) %>%
mutate('S2A_CPM' = S2A/sulf2_mapped_reads) %>%
mutate('S3A_CPM' = S3A/sulf3_mapped_reads) %>%
select(genome,start,stop,S1S_CPM, S2S_CPM, S3S_CPM, S1A_CPM, S2A_CPM,
S3A_CPM) %>%
rowwise() %>%
mutate('SS_average' = mean(c(S1S_CPM, S2S_CPM, S3S_CPM))) %>%
rowwise() %>%

```

```

mutate('SA_average' = mean(c(S1A_CPM, S2A_CPM, S3A_CPM))) %>%
#MUTATE SA_AVERAGE TO BE NEGATIVE FOR THE GRAPH
mutate('SA_average'= SA_average*-1) %>%
select(start, SS_average, SA_average)

combined_MBS_df <-sulf_norm_MBS_df %>% merge(y=pyr_norm_MBS_df, by='start')
head(combined_MBS_df)

#MELT SENSE AND ANTISENSE TO THE SAME COLUMN
combinedMBS_melt_df <- data.frame(combined_MBS_df) %>%
  melt(measure.vars= c('PS_average', 'PA_average','SS_average','SA_average'))
head(combinedMBS_melt_df)

ggplot(combinedMBS_melt_df, aes(x=start, y=value, fill=variable)) +
  geom_bar(stat= 'identity', position= 'identity', alpha= .5) +
  scale_fill_manual(values=c('black', 'blue','red','grey')) +
  theme_classic()

```

### ***Data storage***

All data is stored on the Santangelo Backup #2 SSD located in MRB 301.

## REFERENCES

1. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
2. Pettinati, I., Grzechnik, P., Ribeiro de Almeida, C., Brem, J., McDonough, M. A., Dhir, S., Proudfoot, N. J. & Schofield, C. J. Biosynthesis of histone messenger RNA employs a specific 3' end endonuclease. *Elife* **7**, (2018).
3. Maier, L.-K. & Marchfelder, A. It's all about the T: transcription termination in archaea. *Biochem. Soc. Trans.* **47**, 461–468 (2019).
4. Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L. & Tong, L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956 (2006).
5. Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).
6. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
7. Berkemer, S. J., Maier, L. K., Amman, F., Bernhart, S. H., Wörtz, J., Märkle, P., Pfeiffer, F., Stadler, P. F. & Marchfelder, A. Identification of RNA 3' ends and termination sites in *Haloferax volcanii*. *RNA Biol.* **17**, 663–676 (2020).
8. Dar, D., Prasse, D., Schmitz, R. A. & Sorek, R. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat.*

- Microbiol.* **1**, 16143 (2016).
9. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.06.020
  10. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription Regulation in Archaea. *J. Bacteriol.* **198**, 1906–1917 (2016).
  11. Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C. T., Munson, K., Core, L. J. & Lis, J. T. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
  12. Imashimizu, M., Takahashi, H., Oshima, T., McIntosh, C., Bubunenko, M., Court, D. L. & Kashlev, M. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.* **16**, (2015).
  13. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K. & Bentley, D. L. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Mol. Cell* **76**, 896-908.e4 (2019).
  14. Gehring, A. M. & Santangelo, T. J. in *Methods Mol. Biol.* **1276**, 263–279 (2015).
  15. Gehring, A., Sanders, T. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *BIO-PROTOCOL* **7**, (2017).
  16. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus*

- kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Res.* **15**, 352–363 (2005).
17. Hileman, T. H. & Santangelo, T. J. Genetics techniques for *Thermococcus kodakarensis*. *Front. Microbiol.* **3**, 195 (2012).
  18. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic Techniques for the Archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).
  19. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15**, 684 (2014).
  20. Santangelo, T. J., Čuboňová, L. & Reeve, J. N. Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* **81**, 897–911 (2011).
  21. Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T. & Reeve, J. N. Polarity in Archaeal Operon Transcription in *Thermococcus kodakaraensis*. *J. Bacteriol.* **190**, 2244–2248 (2008).
  22. Proshkin, S., Rachid Rahmouni, A., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science (80-. )*. **328**, 504–508 (2010).
  23. Stevenson-Jones, F., Woodgate, J., Castro-Roa, D. & Zenkin, N. Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8462–8467 (2020).
  24. Gelsinger, D. R., Dallon, E., Reddy, R., Mohammad, F., Buskirk, A. R. & DiRuggiero, J. Ribosome profiling in archaea reveals leaderless translation, novel

- translational initiation sites, and ribosome pausing at single codon resolution. *Nucleic Acids Res.* **48**, 5201–5216 (2020).
25. Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R. J., Typas, A., Gross, C. A., Kramer, G., Weissman, J. S. & Bukau, B. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).
  26. Sanders, T. J., Lammers, M., Marshall, C. J., Walker, J. E., Lynch, E. R. & Santangelo, T. J. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol. Microbiol.* (2019). doi:10.1111/mmi.14191
  27. Chang, C. H. & Luse, D. S. The H3/H4 tetramer blocks transcript elongation by RNA polymerase II in vitro. *J. Biol. Chem.* **272**, 23427–34 (1997).
  28. Kim, J., Guermah, M. & Roeder, R. G. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
  29. Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E. & Santangelo, T. J. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* **5**, 545–553 (2020).
  30. Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci.* **114**, E6767–E6773 (2017).