DISSERTATION

ARCHAEAL TRANSCRIPTION AND REPLICATION: NEW INSIGHTS INTO TRANSCRIPTION-COUPLED DNA REPAIR AND ORIGIN-INDEPENDENT DNA REPLICATION

Submitted by

Alexandra Marie Gehring

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Thomas Santangelo

J. Lucas Argueso
Jennifer K. Nyborg
Olve B. Peersen

ABSTRACT

ARCHAEAL TRANSCRIPTION AND REPLICATION: NEW INSIGHTS INTO TRANSCRIPTION-COUPLED DNA REPAIR AND ORIGIN-INDEPENDENT DNA REPLICATION

The three Domains of extant life use similar mechanisms for information processing systems. Although many aspects of replication, transcription and translation are universally conserved, the evolutionary history of the enzymes involved is not always clear and domain-specific differences are known. The transcription apparatus, especially the multi-subunit RNA polymerase (RNAP), has a clear evolutionary conservation across all Domains. Elucidating the mechanisms of the transcription apparatus in Archaea will help further understanding of underlying transcription mechanisms and regulation of those mechanisms, not only in Archaea but also in Bacteria and Eukarya. Conversely, the DNA replication machinery, most notably the replicative DNA polymerases, are distinct for each Domain. Any demonstration of the activities of the replication proteins, and especially discovery of unique pathways and mechanisms underlying replication helps to improve the understanding of the larger evolutionary questions surrounding DNA replication.

The compact nature of archaeal genomes necessitates timely termination of transcription to prevent continued transcription of neighboring genes while ensuring complete transcription of the gene of interest. Transcription elongation is processive, and the transcription elongation complex is exceptionally stable. The disruption of this transcription elongation process, transcription termination, is an essential step in the transcription cycle. The presence of DNA lesions causes early termination of transcription in Bacteria and Eukarya. The results of this dissertation demonstrate this is also true in Archaea. Archaeal RNAP arrests transcription at DNA lesions and likely initiates transcription-coupled DNA repair (TCR) as will be soon demonstrated using *in vivo* techniques developed during this dissertation work.

DNA replication is a highly regulated cellular process, particularly initiation of DNA replication. The long-standing replicon hypothesis states a trans-acting replication initiation protein must recognize a cis-acting DNA element, the origin of replication. For the 50 years after the replicon hypothesis was first posited, the replication hypothesis was supported in phages, Bacteria, Archaea, and Eukarya. The work presented in this dissertation describes the non-essentiality of Cdc6 and the origin of replication, and further demonstrates that origin-independent DNA replication is the mechanism by which *Thermococcus kodakarensis* replicates its genome. The results of this study and others in the field brings forward questions about the evolutionary history of DNA replication in all three Domains of extant life.

ACKNOWLEDGEMENTS

I am not sure I will ever be able to find the words to properly thank all the people who have helped me over the last five years. Please know these brief acknowledgements are only a small portion of the gratitude I have.

To my advisor, Tom Santangelo – Thank you for helping me to find confidence I never knew I had. Your support, both professionally and personally, has been essential to my successes.

To Lucas Argueso, Jenny Nyborg, and Olve Peersen, my committee members – Thank you for asking the tough questions and helping me to move my projects forward.

To all the undergraduates who have helped me along the way, Anthony, Travis, Adam, Kyle, Oghene, and Brian – Thank you for teaching me how to be a mentor, bringing a smile to my face, and helping me complete some difficult science.

To Julie – I do not think I could have survived graduate school without you. You always know when I need to go for a walk around campus, to get a coffee, or to make one of our many trips to Lake Street.

To Brett, Hallie, and Travis and all the other members of the Santangelo lab – I will be forever grateful for our dance parties, long conversations, and friendships. Both me and my science would be worse without you.

To Mike Terns, Andy Gardner, and David Crowley – Thank you for your collaborations and always listening to ideas for new approaches to experimental problems. Your guidance has been crucial to my success.

To Nancy – Thank you for always making sure the equipment necessary for my experiments is ordered and delivered. You always seem to know how to brighten my day.

To Kristina Quynn – You are so much more than an internship advisor to me. You have become a confidant and somebody I regularly turn to for support. You have helped to improve my writing; however, most of all you have changed my relationship with writing. Because of you, I find my writing a joyful process.

To Matt and Nicole, my writing group partners – Thank you for helping me to meet my writing goals week after week and for always listening to my worries and stresses.

To my parents and family – Your support has been unwavering and for that I am forever grateful. Thank you for always listening me talk about lab, even when you didn't understand what I am saying.

To Laura, Annie, and Anna – Your friendship has been invaluable. Thank you for always asking how *Thermococcus* is doing and knowing when I just need a glass of wine.

To Noah – You always know when I need a bowl of mac n cheese and a hug after a long day. Thank you for understanding the long days in lab and the weekends when I just have to check on something quick because we both know quick is closer to 2 hours. Your support has been constant and essential.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION[1]

1.1 ARCHAEAL REGULATION STRATEGIES

Archaea encode a great diversity of metabolic pathways and unique physiologies. Although Archaea often resemble bacteria in size and shape, and occasionally share conserved metabolic pathways, the enzymology driving archaeal DNA replication, transcription, translation, recombination and repair shares homology with Eukarya. The eukaryotic homology of these critical pathways and enzymes, coupled with the component simplicity of the enzymology involved in archaeal central dogma processes, provides a unique opportunity to detail shared aspects – and compare unique strategies – of regulation in these Domains.

There is inherent value in determining the strategies employed in extant life to regulate gene expression and growth. The increase use of archaeal organisms for biofuel production and biotechnological applications also supports a more thorough discernment of archaeal physiology. The recalcitrance of many archaeal species to genetic investigations coupled with the very real expenses of maintaining many species in the laboratory has limited progress in understanding archaeal diversity, metabolism, regulation, and response to environmental stimuli. As such, significant gaps in our knowledge of archaeal information processing systems remain. This dissertation provides insights into mechanisms supporting initiation of genomic DNA replication, specialized processes of DNA repair, and mechanistic insights into response of

[1] A portion of this chapter was previously published as part of a review article, “Transcription Regulation in Archaea”, in June 2016. Parts of the original manuscript have been updated, expanded, or omitted where necessary and appropriate.

JEW and I conceived the content and co-wrote the manuscript with input from TJS.

Gehring, A. M.; Walker, J. E.; Santangelo, T. J. Transcription Regulation in Archaea. *Journal of bacteriology* **2016**, 198, 1906-17. PMID: 27137495.

the transcription apparatus during the elongation and termination phases of the archaeal transcription cycle.

1.2 TRANSCRIPTION

The multi-subunit RNA polymerase (RNAP) is essential for all life. Although RNA synthesis is carried out by RNAP, the activities of RNAP during each phase of transcription are subject to basal and regulatory transcription factors. Substantial differences in transcription regulatory strategies exist in the three Domains (Bacteria, Archaea, and Eukarya). Only a single transcription factor (NusG or Spt5) is universally conserved[1,2], and the roles of many archaeal-encoded factors have not been evaluated using either *in vivo* and *in vitro* techniques. Archaea are reliant on a transcription apparatus that is homologous to the eukaryotic transcription machinery; similarities include additional RNAP subunits that form a discrete subdomain of RNAP[3,4], as well as basal transcription factors that direct transcription initiation and elongation[5–8]. The shared homology of archaeal-eukaryotic transcription components aligns with the shared ancestry of Archaea and Eukarya, and this homology often is exclusive of Bacteria. Archaea are prokaryotic but the transcription apparatus of Bacteria differs significantly from that of Archaea and Eukarya.

The archaeal transcription apparatus is most commonly summarized as a simplified version of the eukaryotic machinery. In some respects, this is true, as homologs of only a few eukaryotic transcription factors are encoded in archaeal genomes and archaeal transcription *in vitro* can be supported by just a handful of transcription factors. However, much regulatory activity in eukaryotes is devoted to post-translational modifications of chromatin, RNAP, and transcription factors, and this complexity seemingly does not transfer to the archaea where few post-translational modifications or chromatin-imposed transcription regulation events have been described. The ostensible simplicity of archaeal transcription is under constant revision as more detailed examinations of archaeal-encoded factors become possible through increasingly sophisticated *in vivo* and *in vitro* techniques.

The Archaeal Transcription Cycle

Transcription is highly regulated and the transcription cycle is typically demarcated into three phases: initiation, elongation and termination[9–14] (Figure 1.1). An abbreviated and overall introduction to this cycle is presented first, with sections below detailing the activities of RNAP and associated factors during each stage of transcription. Briefly, archaeal transcription initiation requires that RNAP be directed to promoter-sequences defined by the binding of TATA binding protein (TBP) and transcription-factor B (TFB). TBP, TFB, and RNAP are sufficient to generate a single-stranded section of DNA (the transcription bubble) and feed the template strand into the bi-partite active center of RNAP[6,15]. RNAP can initiate transcript synthesis *de novo* and continued synthesis then competes with favorable promoter and initiation factor contacts until promoter escape can be achieved. Release of RNAP from the initiating factors classically defines the end of initiation, although in reality no clear boundary separates the latter stages of initiation from the early stages of elongation. Although TFB and TBP are necessary and sufficient to permit promoter-directed transcription initiation, a third conserved factor, transcription factor E (TFE), can also assist in transcription initiation and likely leaves the promoter with RNAP during the early stages of transcript elongation[16–19]. Transition to a stable, long-lived elongation complex is believed to involve internal rearrangements of RNAP. This transition involves the exchange of initiation factors for stably bound elongation factors that monitor RNA synthesis for accuracy, respond to regulatory DNA sequences, react to regulatory inputs of more transiently associated transcription factors, and influence processivity of RNAP. Elongation is, in general, very stable, but specific sequences can lower the overall energy of the transcription elongation complex permitting either spontaneous intrinsic or factor-assisted termination[20,21]. Transcription termination results in release of both the transcript and RNAP from the DNA template.

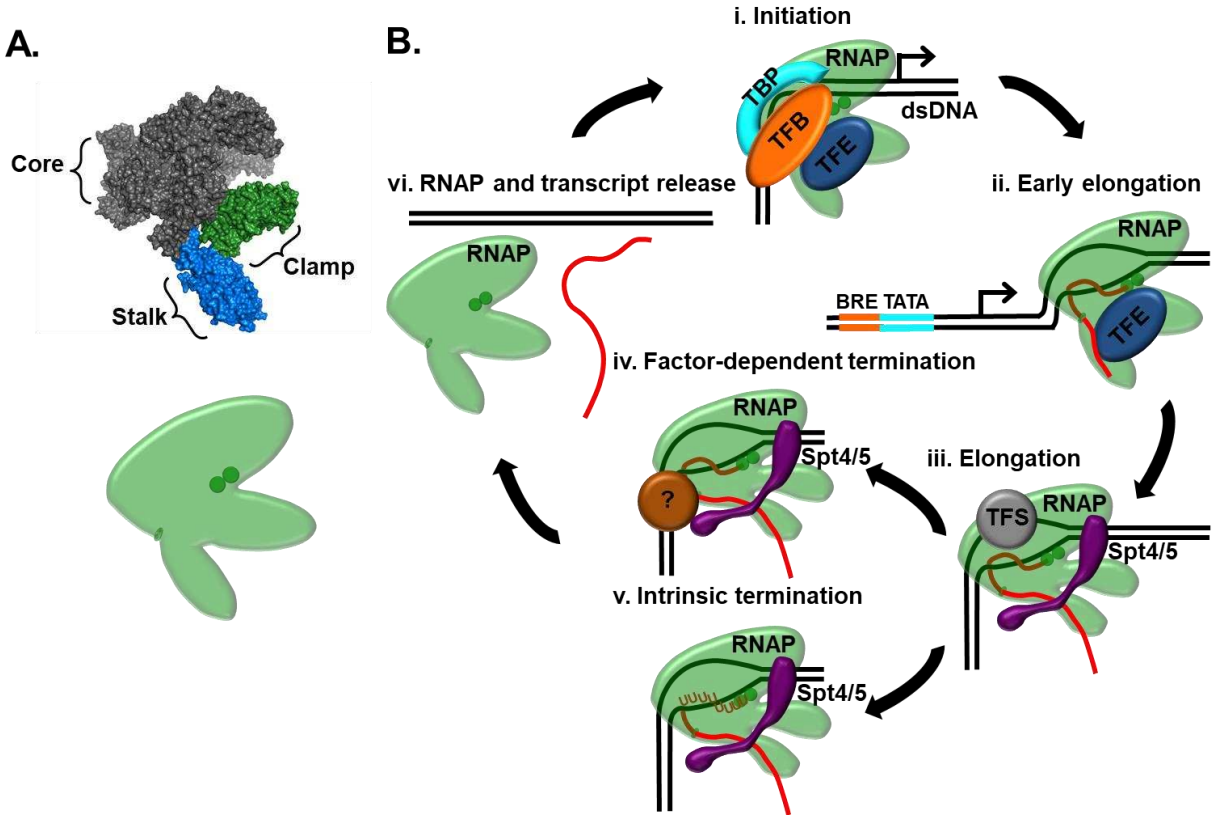

Figure 1.1. The archaeal transcription cycle. (A) The euryarchaeal RNA polymerase crystal structure from *T. kodakarensis* (PDB ID: 4QIW) is shown in a surface representation. The clamp (green) and stalk (blue) domains are highlighted. A simplified cartoon structure of RNA polymerase is shown below in green; the bipartite active site and RNA exit channel are highlighted in dark green. (B) i. RNAP is recruited to the promoter by transcription factors TFB, TFE and TBP during transcription initiation. ii. RNAP escapes the promoter and early elongation begins with TFE bound to RNAP. iii. TFE is replaced by elongation factor Spt5 during elongation. iv. Factor dependent termination is predicted to occur in archaea by an unknown factor. v. Intrinsic terminators are characterized by a run of T's on the non-template strand. vi. The transcript is released and RNAP is recycled for another round of transcription.

Regulated transcription initiation

Transcription initiation is tightly regulated by both transcription factors and DNA elements. The minimal, necessary proteins and DNA elements for archaeal transcription initiation are now well defined and characterized[22–29]. A recent excellent review[30] summarizes the actions of repressors and activators that function during initiation in archaeal species. The focus here is on the roles of new DNA elements and newly discovered strategies of basal initiation factors.

Basal Transcription Factors

TBP and TFB are the only transcription factors required for *in vitro* transcription under optimized conditions, and TFE has been shown to assist promoter-opening when conditions are sub-optimal[17]*. In vivo* studies have shown that archaea must retain at least one gene encoding TBP and one gene encoding TFB, although many archaeal species encode multiple TBP and TFB isoforms[5,22,31–36]. Some differences in promoter-sequence preferences and protein-pairing have been noted in TBP-TFB isoform pairs[37–42], but these minor differences are not on par with the clear but not always radical promoter-sequence differences noted for alternative σ-factors in bacterial transcription[40,43]. TFE also appears essential, and it is currently unclear if this essentiality is due to necessary activities during transcription initiation or some other role in the transcription cycle[27,44,45].

All three of the aforementioned transcription factors have close eukaryotic homologs: archaeal TBPs are nearly identical to eukaryotic TBPs[46]; archaeal TFB proteins are homologous to eukaryotic TFIIB proteins[47], with homology also seen with the Pol III initiation factor BRF1[48] and Pol I initiation factor Rrn7/TAF1B[49]; and archaeal TFE proteins are homologous to the N-terminal half of the eukaryotic TFIIEα, and recent evidence identified a separate homolog in some lineages to eukaryotic TFIIEβ[18]. TBP is needed to recognize the TATA box, bend the DNA, and recruit TFB[47]; its role has therefore been deemed equivalent to the role of eukaryotic TBPs. Recent, sophisticated TIRF-FRET measurements now detail differences in the activities

of archaeal and eukaryotic TBPs, despite the nearly identical 3-D folds of these factors[23]. In some cases, archaeal TBPs require the co-binding of TFB to stably bind and bend the promoter DNA [5,23,50,51]. It is tempting to speculate that different promoter sequences may be regulated by different TFB-TBP pairs based on the interdependence, or lack thereof of cooperative DNA bending for establishing a stable platform for RNAP recruitment. Recent studies suggest that select isoforms of TFB and TBP can result in differences in transcription output, but further studies will be needed to determine if these effects on such preliminary steps of transcription initiation are a direct mode of regulation resulting in phenotypic differences[38,52].

In contrast to eukaryotic transcription, archaeal promoter opening is not an energy-dependent process[6]. Therefore, TBP and TFB alone are capable of assisting RNAP in formation of the transcription bubble. In all archaea, TFB is responsible for stabilizing the TBP bound DNA complex and together this bi-partite protein platform recruits RNAP[53], but how these molecular interactions melt the DNA is still unresolved. Reconstructions and analyses of open complexes using archaeal components reveal an overall architecture of the open promoter complex and provide the first placement of the non-template strand within the complex[53]. TBP and TFB are located closer to RNAP than would be the case for eukaryotic promoters and this proximity may provide more intimate contacts that collectively provide the energy to open the promoter DNA. The tight network of interactions in the archaeal open complex may torsionally strain the DNA and melting is likely to reveal this strain and result in open complex formation.

Several new insights into TFE activity and evolution have been described. Archaeal TFE had previously been characterized as a monomer and as a homologue of the alpha-subunit of eukaryotic TFIIE, termed TFIIEα[17,19,54]. Eukaryotic TFIIE is a heterodimeric complex of TFIIEα and TFIIEβ, but archaeal genomes had only been shown to encode a homologue of the alpha-subunit[55,56]. Eukaryotic RNAPs differ in their requirements for initiation, with RNAP III incorporating homologues of several RNAP II initiation factors as core components of RNAP III[57–59]. Comparisons of the RNAP III subunit hRCP39 revealed a well-conserved archaeal

homolog (termed TFEβ) that directly and extensively interacts with archaeal TFE (now named TFEα)[18]. Although TFEβ is not conserved in all archaea, TFEβ is essential for some crenarchaea, and when employed *in vitro*, TFEα/TFEβ complexes are effective in binding RNAP, stabilizing open complex formation, and stimulating total transcriptional output[18].

The mechanism of TFE recruitment to the initiation complex and its activities during initiation have been partially resolved. TFEα simultaneously binds TBP, RNAP and downstream DNA, and has been shown to stimulate transcription at non-canonical promoter sequences and at reduced temperatures *in vitro*[17,19,60]. Several studies have identified critical interactions between TFE and the pre-initiation complex that have furthered our understanding of TFE function during initiation[2,16,27,60,61]. TFEα consists of two domains: a winged helix (WH) domain and a zinc-ribbon (ZR) domain[62,63]; TFEβ contains a conserved WH domain and a FeS domain[18]. The winged-helix domain of TFEα contacts the upstream, non-template strand of DNA and helps form the open-promoter complex through an unknown mechanism[16,53]. Several studies have shown that the presence of the RNAP stalk domain, unique to archaeo-eukaryotic RNAPs and comprised of two subunits – RpoE and RpoF in archaea and Rpo4 and Rpo7 in eukaryotes – is essential for the full activity of TFEα[60,64,65]. The predicted interaction between TFEα and the stalk domain was bolstered by co-purification of TFEα with intact RNAP but the loss of TFEα from RNAP preparations wherein the stalk domain was missing[45]. A recent structure-function study identified critical interactions between TFEα and RpoE of the stalk domain[27]. TFE may have an essential role in modulating intramolecular movements of RNAP during the transcription cycle, most notably movements of the clamp domain. Interaction of TFEα with both the stalk and clamp domains of RNAP during transcription initiation may retain the clamp domain in an open conformation necessary for initiation and early elongation. Replacement of TFE by Spt4/5 during early elongation may alter clamp positioning and further stabilize the elongation complex[2].

DNA elements

Transcription initiation is regulated by DNA elements that are recognized by basal transcription factors and that influence subsequent steps in promoter opening. There are four DNA elements currently known to regulate archaeal transcription initiation; i) the TATA box located approximately 25 bp upstream of the site of transcription initiation[66–68], ii) the TFB Recognition Element (BRE) located immediately upstream of the TATA box[5], iii) the initiator element (INR) located within the initially transcribed region, and iv) promoter proximal element (PPE) located between the TATA box and the site of transcription initiation[69–71]. Of these four, only the TATA box and BRE are required for transcription initiation, although alterations to all four elements can influence total output of a promoter.

The INR is not a required DNA element for transcription initiation; however, it is a regulatory element that can increase the strength of the promoter in a TATA- and BRE-dependent manner. The INR is a core promoter element located in the 5′ untranslated region (UTR), and has sequence similarity to the TATA box. The INR has been shown to be targeted by some transcriptional activators, and its high A/T content may facilitate promoter-opening in some instances. Given that many archaeal transcripts are leaderless, the INR is not consistently identifiable, and the regulatory influence of INR sequences does not appear to extend to RNA half-life or to the altering of translational capacity[72]. PPEs, centered at approximately ten bps upstream of the site of initiation, have been shown to increase transcription output through recruitment of TFB[69,70]. Additionally, permanganate footprinting data of the preinitiation complex, demonstrated that the border of the transcription bubble is at the PPE and that this region is important for the activity of TFEα/β[18].

Regulation of transcription elongation

As transcription transitions from initiation to elongation, RNAP undergoes a conformational change accompanied by the replacement of initiation factors with elongation factors[2,12,73–76]. It is plausible that the emerging nascent transcript stimulates the swap of

regulatory factors and initiates the intramolecular movements that result in stable elongation complex formation[64,77]. Very few transcription elongation factors have been bioinformatically identified within archaeal genomes, and it is probable that archaeal-specific factors await discovery. Archaeal genomes do not appear to encode any co-activator complexes nor mega-complexes for chromatin-modification or rearrangements. There does not appear to be machinery for regulated post-translational modifications of the archaeal transcription apparatus nor chromatin, with the exception of acetylation/deacetylation of the small chromatin-associated protein Alba[78–81]. Furthermore, archaeal transcripts are not capped, do not require nuclear export and with the exception of self-splicing introns, are intron-less; thus factors responsible for these activities are similarly lacking from archaeal genomes[82–84].

Transcription elongation factors have various roles including increasing processivity and fidelity of RNAP and/or increasing genome stability. Only two archaeal elongation factors have been experimentally studied: the aforementioned universally conserved elongation factor Spt5, often with a conserved binding partner Spt4 (Spt4/5)[2,85,86] and transcription factor S (TFS)[87,88]. Several recent studies have shed light onto the roles of Spt5 during elongation[64,74,89,90]. TFS, with homology to the C-terminal domain of eukaryotic TFIIS and functionally analogous to GreA/GreB in Bacteria[8,91–93] can stimulate endonucleolytic cleavage of the RNA from backtracked RNAP complexes[87,93–95]. The finding of multiple TFS homologs in some archaeal lineages offers the possibility of unique regulatory roles of specific isoforms.

Transcription factor Spt5

Archaeal Spt5, homologous to bacterial-encoded NusG, consists of two domains: the NusG N-terminal (NGN) domain and a single, C-terminal Kyrpides-Ouzounis Woese (KOW) domain with affinity for single stranded RNA[85,86,89]; eukaryotic Spt5 typically contain three to six repeats of the C-terminal KOW domain[95–97]. Critical, direct molecular interactions between Spt5 and RNAP have been identified in both Bacteria and Archaea[85,86,89,90,95,98–100], and the conservation of RNAP and Spt5 infers that these same interactions are used in Eukarya. Briefly,

a hydrophobic depression on the NGN domain interacts with the mobile clamp domain of RNAP, with additional interactions between the NGN domain and RNAP jaw domain likely fixing the location of the clamp domain in a closed configuration[11,98]. Spt5 interaction with RNAP is not necessary for productive and processive elongation *in vitro*, but the interaction does increase total output of transcription systems[1]. It is plausible that Spt5 increases elongation rates and processivity, as *E. coli* NusG does, and it is further possible that the increased efficiency of transcription results from the stabilization of the clamp domain that in turn stabilizes the DNA:RNA hybrid in place during transcription elongation[89,101–103]. The NGN domain also contacts the upstream strands of DNA offering protection from backtracking and, by inference, may reduce pausing of the transcription elongation complex[89,90,104,105]. It is of importance to note that NusG/Spt5 can have a positive and/or negative effect on elongation rates and pause events of RNAP. In *Thermus thermophilus*, NusG slows down RNA elongation rather than increases elongation rates[98]. In *Bacillus subtilis*, sequence-specific interactions of the NGN and non-template DNA strand within the paused transcription bubble stabilized the pause event in the *trp* operon[104,106]. Furthermore, evidence in *Saccharomyces cerevisiae* has shown that Spt4/5 induces pauses during early elongation of RNAP I but promotes elongation downstream[107]. Although NusG can elicit opposite roles on transcription elongation, the NusG:RNAP binding sites remain well-conserved across various species. Archaeal and eukaryotic-genomes often encode an additional elongation factor, Spt4 (annotated as RpoE''/RpoE2 in archaea), that forms a complex with Spt5 and stabilizes the Spt5-RNAP interaction[1,86,95]. Spt4 does not appear to be essential; however, the affinity of Spt5 for RNAP decreases in the absence of Spt4 *in vitro*[1].

The primary interacting partners (e.g. RNAP and Spt4) of the Spt5-NGN domain have been established in molecular detail; however, no specific interacting partners of the KOW domain have been identified in archaea. It is possible that the affinity of the KOW domain for RNA leads to non-specific interactions with the emerging transcript, however, it is tempting to

speculate greater involvement of the KOW domain based on the known activities of the C-terminus of bacterial NusG[108]. Bacterial NusG can facilitate elongation or termination depending on its binding partner[99,101,102,109–111]. The bacterial NusG KOW domain can interact with the S10 ribosomal subunit (NusE) during elongation, thereby linking the leading ribosome with the transcription apparatus[110,111]. When not bound to a trailing ribosome, the bacterial NusG-KOW domain can be bound by and stimulate the activity of the transcription termination factor Rho[109,112,113]. Archaeal transcription and translation are similarly coupled[114,115], and it is reasonable to venture that archaeal Spt5 can also link the archaeal transcription and translation apparatuses, and also potentially interact with termination factors.

Intramolecular rearrangements of RNAP may increase processivity

The archaeal and three eukaryotic RNAPs can be reduced in complexity to three large domains: the core, the mobile clamp, and the stalk[4,75,116]. The archaeo-eukaryotic stalk, absent from bacterial RNAP, is used by a host of archaeal and eukaryotic transcription factors to bind and regulate the activities of RNAP. Increasing evidence from biochemical, biophysical, and *in vivo* approaches indicate that transcription factor binding often stimulates intramolecular movements of RNAP that appear necessary for transitions between phases of the transcription cycle[4,27,90,100,117,118].

Hinge-like movement of the mobile clamp domain has been demonstrated for the bacterial RNAP[75]. The movements of the mobile clamp are sufficiently large enough to open the main channel of RNAP such that double stranded DNA can easily enter and exit when the clamp is open, whereas double stranded DNA – or the RNA/DNA hybrid – would be trapped inside RNAP when the clamp is closed. The bacterial RNAP clamp is open during initiation but remains closed during processive elongation[73], leading to a simple model of encapsulation of the nucleic acids to explain the dramatic stability of the elongation complex. It was then logical to propose mechanistic actions of transcription factors that may modulate the clamp positioning with respect to the core and stalk domains of RNAP and thus alter the stability and transitions of

RNAP throughout the transcription cycle. TFE is predicted to make contacts with both the clamp and stalk domain of RNAP thereby fixing the clamp into the open conformation critical for initiation[27,60,118–120]. As transcription transitions into the elongation phase, RNA emerges from the enzyme and interacts with the stalk domain[64,77], where a predicted steric clash occurs between the RNA and TFE, likely driving TFE to disengage from RNAP. The disengagement of TFE allows for Spt5 to bind to the clamp and core domains of RNAP and lock the clamp in the closed position, thus ensuring processivity during elongation[89].

RNAP clamp movement is predicted to be universal; however, both the archaeal and eukaryotic RNAP contain additional subunits, including the stalk domain[2,75,116,119,120], and previous structural data predicted the stalk domain would sterically limit or abolish major movements of the clamp domain. Recent crystallographic evidence of the complete euryarchaeal RNAP demonstrated that the clamp is able to open without a steric clash with the stalk domain through a coordinated swing and rotation movement of both the clamp and stalk domains[75]. This evidence supports the bacterial mechanism of the clamp opening and closing during initiation/termination or elongation, respectively, thus supporting a universal model of clamp movement.

Transcription Termination

Transcription termination occurs when the transcription elongation complex becomes sufficiently unstable and fails to maintain contact between RNAP and the encapsulated nucleic acids. The stability of the transcription elongation complex is derived from i) contacts between RNAP and the RNA:DNA hybrid, ii) contacts between RNAP and single-stranded RNA in the exit channel, iii) contacts between RNAP and the downstream DNA, and iv) the base pairing of the RNA:DNA hybrid[116,121–127]. The first and last of these contacts are most likely to be altered during the termination process. Transcription through specific DNA sequences can result in stronger or weaker base pairing within the RNA:DNA hybrid, and contacts between RNAP and the nucleic acids are most easily modified by movements of the clamp domain that relieve movements of

the hybrid with respect to the core of RNAP[128–130]. Release of the nascent RNA may be possible through continued translocation in the absence of synthesis, or the RNA:DNA hybrid could be released in bulk if the clamp domain transitions from a closed to open position. The gene-dense nature of many archaeal genomes necessitates timely termination of transcription to prevent aberrant transcription of neighboring genes. It is predicted that there are two mechanisms of termination across all domains: intrinsic termination and factor-dependent termination.

Intrinsic termination

Intrinsic transcription termination is driven primarily by weak base pairing within the RNA:DNA hybrid and occurs independently of the activity of transcription factors[131,132]. Intrinsic transcription termination has been established in all three Domains[20,21,133,134], with some differences in sequence and structural requirements[131,133,135–137]. The archaeal RNAP, like eukaryotic RNAP III, is sensitive to intrinsic termination[20,134,138,139]. Eukaryotic RNAP I and RNAP II do respond to DNA sequence context in the form of pauses and arrests, but rarely release the transcript at such positions[140–142]. Archaeal intrinsic termination is characterized by a run of five to ten thymidine residues in the non-template strand, encoding a run of poly-U at the 3' end of the nascent RNA[20,21]. The weak rU:dA RNA:DNA hybrid at or near the positions of termination is seemingly insufficiently energy rich to maintain the stability of the elongation complex; RNAP III similarly spontaneously dissociates upon transcription of poly-T non-template tracts.

Identification of factor-dependent termination

Transcription factors involved in initiation and elongation have been characterized in all domains and until recently transcription termination factor(s) had only been characterized in Bacteria and Eukarya[14,143–146]. Two well-studied transcription bacterial termination factors, Rho and Mfd[13,147–151], lack clear homologs in archaeal genomes, but there are hints that analogous activities may be present in archaeal species. Interestingly, the mechanism of the only identified archaeal transcription termination factor, Eta, suggests that it is analogous to Mfd[14].

Although first characterized in *T. kodakarensis*, Eta is conserved in most archaeal lineages thus suggesting that factor-dependent transcription termination is a common transcription regulatory strategy in archaea. Eta is able to disrupt transcription elongation complexes causing release of the nascent RNA to the supernatant. Eta appears to be a slow acting transcription termination factor and likely is responsible for identifying RNAP in a paused or stalled transcription elongation complex. To terminate transcription and remove the stalled RNAP, Eta binds to the upstream DNA and in an ATP dependent manner pushes RNAP into a hypertranslocated state which causes disassociation of the complex.

The bacterial Mfd protein can remove RNAP from sites of DNA damage and initiate transcription-coupled DNA repair (TCR)[147,149,151,155]. Recent evidence that the archaeal RNAP halts synthesis and forms long-lived complexes at the site of DNA lesions *in vitro* predicts that mechanisms exist to remove RNAP from the site of damage[156]. Further the deletion of the transcription termination factor, Eta, in *Thermococcus kodakarensis* resulted in cells that were sensitive to UV irradiation, even further suggesting the analogous activity of Eta to Mfd[14].

Rho is a homohexamer helicase that represses phage transcription and mediates polar repression of downstream genes when transcription and translation become uncoupled[144,152–154]. Archaea demonstrate polar repression of downstream genes in the absence of continued translation, and it is likely that a factor or factors mediate polarity in archaea[115]. It is tempting to use the bacterial model of NusG:Rho interactions to conjure a similar picture for Spt5 KOW interactions with an archaeal transcription termination factor; Rho is capable of terminating a stalled archaeal RNAP *in vitro*[20].

Chromatin architecture affects the transcription cycle

Archaea employ two seemingly distinct mechanisms to compact, wrap, and condense their genomes to fit within the cell (Figure 1.2)[157]. Most euryarchaeal species are oligoploid and encode histone proteins that dominate chromatin architecture[158–162]; archaeal histones mimic the core eukaryotic histone-fold[163,164]. In contrast, most crenarchaeal species are diploid and are

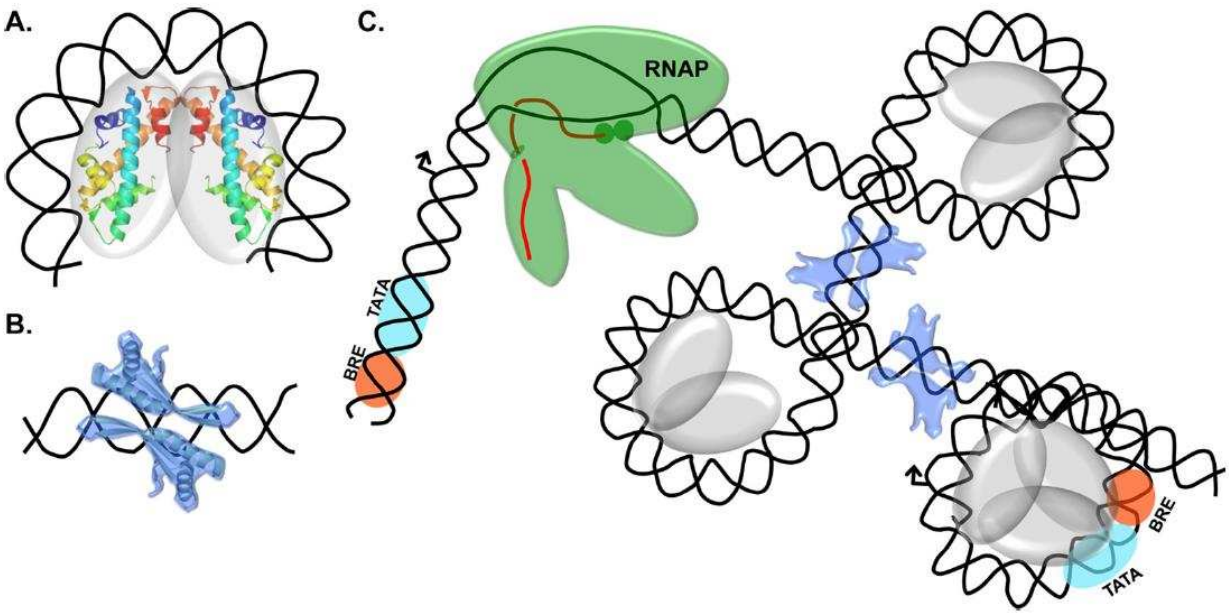

Figure 1.2. Transcription in the context of archaeal chromatin. (A) The structure of histone A from *Methothermus fervidus* (PDB ID: 1B67) is overlaid by a cartoon representation of each histone dimer with ~60 bp of DNA wrapping the complex. (B) The crystal structure of an Alba dimer from *Sulfolobus solfactaricus* (PDB ID: 1H0X) bound to DNA is overlaid by a cartoon representation. (C). Transcription elongation continues in a chromatin environment. Accessibility of the TATA box and BRE is altered by localized chromatin structure.

reliant on small, basic nucleoid proteins to organize their genomes[165,166]. Condensation demands organization of the genome and offers regulatory opportunities by controlling the accessibility of promoter sequences, the introduction of local superhelicities that may promote or inhibit promoter opening, and the potential for the introduction of chromatin-based obstacles to transcription elongation. The overall role of genome architecture with respect to archaeal transcription is an emerging area of study with several recent studies highlighting the breadth of influences genome architecture can have on transcription output at the organismal level[163,167–169].

Archaeal histone-based chromatin is known to form extended superstructures of polymerized histone-tetramers enveloped in a superhelical DNA that, in overall dimensions, mimics an extended and continuous nucleosomal configuration[163]. Archaeal histones share similar biases with eukaryotic nucleosomes for flexible DNA sequences and are, in general, absent from the core promoters of archaeal genes[163,170,171]. Archaeal histone proteins share the same core-fold as eukaryotic histones, but lack the extensions from this fold (i.e. tails) that are highly modified and essential for proper nucleosome dynamics in eukaryotes[172]. Archaeal histone-based chromatin structures present a surmountable barrier to the progression of the transcription elongation complex, although traversion does slow the elongation complex[173]. The lack of known modifications to archaeal histones, and the lack of known machinery for the repositioning or movement of archaeal histones suggests that transcription elongation complexes simply traverse histone-based chromatin complexes naturally and that chromatin organization spontaneously reforms when the histones gain access to preferred binding positions following the departure of RNAP. This mechanism of elongation through the histone structures is likely similar to the mechanism of Pol III in eukaryotes[173–175].

The activities or stimulatory effects of archaeal elongation factors on transcription through archaeal histone-based chromatin remain to be explored; the substantial pausing and delayed progress of RNAP on chromatinized-templates suggests that elongation factors will

accelerate progress of the transcription elongation complex. Any role of chromatin architecture in transcription termination is similarly unexplored. Topology of naked DNA templates does influence the positions and efficiencies of intrinsic terminators, suggesting that chromatin-templates may also influence termination patterns. Archaeal histone-based chromatin-structures are not only depleted from promoter regions, but also from predicted termination regions, suggesting a potential regulatory role for chromatin architecture on termination of transcription[170].

Histone-based regulation of transcription

Several genetic studies have addressed the role of archaeal histone-based chromatin on gene expression at the organismal level with surprisingly different results. In some halophilic species, singular histone-encoding genes are non-essential and histone proteins appear to function more akin to site-specific transcription factors, moderately influencing the expression of only a few genes[167]. These studies contrast the view of histone proteins as general organizational factors with global influence on gene expression, and minimally suggest that archaeal chromatin of some species is dependent on the activities of many nucleoid-associated proteins. When histone-encoding genes have been deleted, or attempted to be deleted from other species, more global disruption of gene expression has been noted[163,164,167–171,176–179]. Some species are reliant on at least one histone protein, and it is unclear at this point whether the noted global changes in gene expression seen in deletion strains stem from reorganization or disorganization of the archaeal genomes or the primary, secondary, and tertiary effects of localized disruptions that leads to additional differences in regulation at remote sites[163,168].

Histone Occupancy at Archaeal Promoter Sequences

Chromatin architecture at a promoter could influence or prevent transcription initiation by occluding transcription factor binding or inhibiting DNA melting[170,173,176,180]. Crenarchaeal encoded nucleoid associated proteins have been shown to influence transcription output through the acetylation/deaceytlation of Alba *in vitro*[81], although Alba has not yet been shown to

influence transcription *in vivo*. It is possible that Alba could regulate transcription given that Alba proteins can loop, condense, bridge and even saturate DNA *in vitro*, but the *in vivo* dynamics remain unknown[181–185]. In the euryarchaeal organism, *Methococcus voltae*, the deletion of the gene encoding Alba resulted in the upregulation only of a small number of genes implying that Alba-based regulation may be limited in scope[169]. Additional research may reveal a clearer picture of transcriptional regulation through binding of Alba.

The binding preferences and genomic locations of stable euryarchaeal histone protein interactions have been mapped and it has been shown that regions directly upstream from the start codon are nucleosome depleted on a global scale[170,171]. The presence of histones bound at the promoter has been correlated with a decrease in total transcription *in vitro*[180], and it was suggested that both steric and torsional effects limit binding of basal transcription factors to the DNA[180]. Although most data supports the lack of histone-based structures at the promoter, specific promoters can be regulated by histone occupancy. This appears to be a general mechanism of histone-based regulation in some halophiles, and a more specialized mechanism of regulation in other species. The transcriptional activator Ptr2 from *Methanocaldococcus jannaschii* must out-compete histones for binding to the promoter to activate transcription of select genes[186].

Evidence for transcription-coupled DNA repair in Archaea

DNA repair mechanisms are essential to maintain the genome for accurate transcription and the production of functional RNA macromolecules. In addition to the generalized DNA repair pathways, TCR has been established in Bacteria and Eukarya forming a link between transcription and DNA repair[187,188]. In Bacteria, RNAP stalls at the DNA lesion, and the footprint of RNAP necessarily occludes the lesion from DNA repair factors. In most cases, Mfd recognizes the stalled RNAP, removes it from the DNA lesion, and recruits the Uvr family of proteins to initiate excision of the damaged base and surrounding sequences; DNA polymerase I then resynthesizes DNA using the undamaged strand as a template and DNA ligase

completes repair by ligating the nick to seal the DNA[188,189]. The Eukarya TCR pathway is similarly dependent on recognition of DNA damage by elongating RNAPs, but the pathway mediating removal, repositioning or degradation of the stalled RNAP to expose the damaged base is much more complex[187]. The identity and participation of many factors is known, but the mechanistic roles played by of all the factors to direct TCR *in vivo* has yet to be established. It is known that the lack or inactivation of factors participating in eukaryotic TCR leads to debilitating human conditions, including Xeroderma Pigmetosum and Cockayne's syndrome[190–195].

Efforts to establish TCR in Archaea have produced mixed results. Studies of members of the Crenarchaea, including *Sulfolobus solfataricus*, concluded that TCR was not detectable, and instead suggested high levels of global genome repair were sufficient to maintain genomic integrity[196,197]. Evidence supportive of TCR has been established in some halophilic, euryarchaeal species, but the requirements for archaeal TCR and the molecular details of proteins involved remains unclear. The halophiles are unique among known archaea in that their genomes encode UvrABC – likely recently acquired via horizontal gene transfer from bacteria[198]. In *Haloferax volcanii*, it was demonstrated that UvrA was required for TCR, however in *Halobacterium* sp. NRC-1, UvrA was not involved in TCR suggesting a unique archaeal pathway[199,200]. Furthermore, most euryarchaeal species do not encode any Uvr protein homologues and therefore are likely reliant on unknown archaeal specific factors for mediating TCR. Most euryarchaeal genomes do encode homologs to the eukaryotic XP proteins, factors known to be critical for eukaryotic TCR[201,202]. It is tempting to speculate that these XP proteins homologs would also have a role in archaeal TCR in archaea.

1.3 DNA REPLICATION

Environmental conditions and the overall health of cells regulate genome architecture, DNA repair and replication. The initiation of DNA replication is typically the target of extensive regulation, and all life normally conforms to a long-standing and seemingly simple mechanism to regulate initiation of genomic replication. This hypothesis asserts a role for a trans-acting factor

(an initiator protein or protein-complex) whose access to and activities at a cis-acting element (DNA origin) are regulated to limit initiation of DNA replication. Although the same overall scheme is used, Bacteria use an initiation factor that is distinct from the factor shared between Archaea and Eukarya with the eukaryotic initiation complex involving more factors. A wide diversity of replicative strategies has been reported within Archaea, and this diversity may reflect the unique challenges presented by the diverse environments in which archaeal species thrive.

Replication Origins and Initiation of DNA Synthesis

DNA replication typically initiates through the assembly and loading of the replisome at distinct positions in the genome, replication origins in Bacteria and Archaea or autonomously replicating sequences (ARSs) in Eukarya. Eukaryotes have thousands of ARSs, and the mechanism of ARS recognition by the Origin Recognition Complex (ORC) varies among species. In some species, the ARS is an AT-rich sequence whereas in other species there is no consensus sequence among ARS, and the mechanism of recognition is unknown[203]. Further complicating recognition of ARS by ORC in Eukarya, chromatin plays a role in recruitment of the ORC and other replication proteins[204,205]. In contrast, Bacteria and Archaea replicate their genomes from a single or just a few genomic positions that retain conserved, sequence specific regions that recruit initiation factors and facilitate unwinding of the DNA[206]. Archaeal replication origins often have AT-rich sequences interspersed with Origin Recognition Boxes (ORB) that serve as binding sites for the origin recognition proteins. The core sequence of an ORB contains a short dyad repeat, necessary for binding of the origin recognition protein[207]. Although the exact sequence and orientation of ORBs within the origin varies, the presence of ORBs and defined replication origins is conserved across Archaea[208,209].

The first archaeal replication origin was experimentally identified in *Pyrococcus abyssi* in 2000[210]. Continued experimentation has identified and validated *in silico* predictions of archaeal replication origins in diverse domains, and the results obtained support that the *Thermococcales*

encode only a single origin of replication per chromosome[211]. The presence of multiple origins (1-4) per chromosome in some archaeal clades may have arisen through horizontal gene transfer, gene duplication events, or the retention of viral sequences in the host genome[209].

Archaeal Origin Recognition Proteins

Eukaryotic origin recognition is reliant on three complexes/proteins: Orc1 recognizes the ARS and binds directly to the DNA; Cdt1 is recruited by Orc1, and this recruitment represents a key regulatory step in replication initiation; and finally, Cdc6 loads the MCM helicases onto the DNA. Archaeal origins are recognized by a subset of homologous factors to also achieve loading of the replicative helicase (Figure 1.3). Archaeal genomes encode a protein that is homologous to both Cdc6 and Orc1, and, depending on the organism, this factor is termed Cdc6, Cdc6/Orc1, or Orc1[209,212,213]. In the *Thermococcales*, Cdc6 is the assigned annotation of the initiator protein and will be the general term used in this dissertation.

With the exception of *Methanopyrus kandleri*, bioninformatic analysis has suggested that all archaea encode at least one copy of Cdc6[212]. The *Thermococcales* encode only one Cdc6 protein but Cdc6 copy number varies considerably among archaeal clades[212]. For example, *H. volcanii* encodes 14 distinct Cdc6 proteins that function at three chromosomal origins and an integrated viral origin[214]. In some organisms, an increase in the number of Cdc6 proteins directly correlates to the number of replication origins with each Cdc6 binding only one specific origin[215]. Yet in other organisms, only one Cdc6 protein is necessary for replication initiation, while the remaining Cdc6 proteins are predicted to have other roles in the cell[216,217]. Although these roles have yet to be experimentally determined in archaea, they are hypothesized to include gene regulation, recombination, replication fork restart, or possible negative regulation of replication[212,216].

Archaeal origin binding and DNA unwinding

For replication to truly initiate, the DNA must be unwound and the minichromosome maintenance (MCM) helicase must be loaded onto the DNA. MCM is responsible for continued

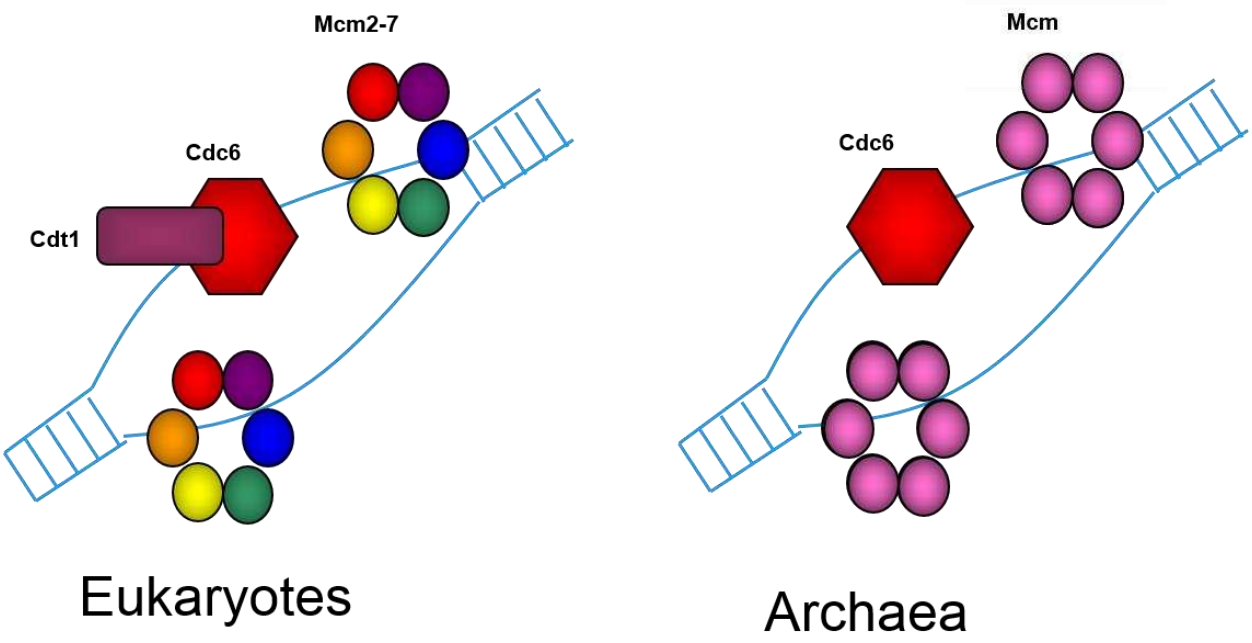

Figure 1.3. Comparison of the DNA replication initiation complexes in eukaryotes (right) and archaea (left). Eukaryotes require both Cdc6 and Cdt1 to recognize the origin of replication whereas archaea only require Cdc6. The eukaryotic Mcm hexamer is made up 6 different proteins whereas archaeal Mcm is homohexamer. In both Domains, once Mcm has unwound the DNA, the DNA replication apparatus is loaded.

unwinding of DNA throughout replication elongation. The winged-helix domain of Cdc6 binds directly to origins, and more specifically to ~4-5 bp of ORB sequences[218–220]. Binding of multiple Cdc6 proteins to adjacent ORBs and mini-ORBS is not cooperative, but is generally necessary to drive the ATP-independent unwinding/melting of origin sequences[221]. Unwinding permits loading of the replicative, ring-shaped, homohexameric archaeal MCM helicase on ssDNA. Cdc6 may assist loading the MCM as is true in Eukarya; however it is also plausible that archaeal MCM-loading is Cdc6-indepenent, and several models to accommodate MCM loading have been proposed[222–225]. MCM can self-load onto DNA at D-loops generated by recombination events; this can lead to further unwinding and recruitment and loading of additional replisome components or simply facilitate resolution of the recombination event[226,227]. Spontaneously-formed, open-ring conformations of MCM provides an alternative mechanism to load the helicase complex that does not require Cdc6 activity, but such mechanisms necessarily fail to target MCM to the presumptive and known origin sequences encoded in archaeal genomes.

Regulation of replication initiation

In both Bacteria and Eukarya, initiation of replication is tightly coupled to the cell cycle. For some archaeal clades, including the Crenarcheota, a defined cell cycle has been reported, and the cell-cycle regulated expression of Cdc6 and related replication proteins is predicted to limit initiation of replication[228,229]. In contrast, a defined cell cycle is not obvious in most archaeal clades – most prominently within the Euryacheaota – and the regulation imposed on DNA replication in these species is largely undefined. Euryarchaeal species often share additional characteristics, including retention of histone proteins, retention of the archaeal specific DNA polymerase D, and oligoploidy (e.g. the retention of many genomes per cell)[159,161–164,170,212,230]. Studies of the euryarchaeal species *P. abyssi* suggest that Cdc6 is always bound to the origin of replication, regardless of growth phase[231]. This constitutive binding of Cdc6 to origin sequences suggests regulation of replication initiation occurs through DNA opening or MCM loading.

Alternative mechanisms of initiating replication

Origin-dependent genomic replication supports rapid growth of Bacteria and Eukarya, but alternative mechanisms of replication initiation can support bacterial growth, albeit very slow growth, when either the origin or origin-recognition protein (DnaA) is missing or inactivated by mutation[232–235]. Origin-independent replication is reliant on long-lived R-loops or the induction of double strand DNA breaks. R-loops can serve as sights of replisome formation and therefore sites of replication initiation[236]. Double strand DNA breaks require recombination for repair of the break. During this repair, the replisome machinery can be loaded onto the DNA and begin origin-independent replication.

In Eukarya, the presence of so many ARS precludes the necessity for all ARSs to fire during each round of replication, but no ARS-independent mechanisms support genomic replication. The retention of Cdc6 in essentially all archaeal genomes suggests a prominent role in replication initiation, but the ability of cells to replicate in an origin-independent manner, even in the presence of Cdc6, suggests that the requirement for Cdc6 may not be absolute. Likewise, the retention of multiple defined origin sequence in many species suggests their use for normal cellular growth, but in many cases, origin-sequences can be individually deleted without compromising viability or growth rate. In an intriguing report, all replication origins can be viably deleted from *H. volcanii*, and in contrast to expectations, cells lacking all origins gained a fitness advantage[237]. Replication in these strains was predicted to occur by RDR, and in support of such, recombination factors that were dispensable in strains containing origins become essential in strains lacking origin sequences.

1.4 CONCLUDING STATEMENT

This work presented here offers insights into the regulatory strategies that regulate gene expression and cellular growth in archaeal organisms. The response of the archaeal RNAP polymerase from *T. kodakarensis* to DNA lesions will be described. Further a system for investigating TCR in *T. kodakarensis* has been developed and results from this system are

pending. This system will be applied to determine which factors, likely unique and archaeal specific, are responsible for TCR *in vivo*. Further, work will be presented describing the regulation of replication, or apparent lack thereof, in *T. kodakarensis*. This lack of regulation provides insight into some of the large evolutionary questions surrounding DNA replication initiation and cellular division in archaeal species. The work presented lays the foundation for continued elucidation of the TCR pathway in the euryarchaea in addition to better understanding of archaeal DNA replication and regulation thereof.

REFERENCES

1. 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 research* **2010**, *38*, 4040–51.
2. 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* **2011**, *43*, 263–274.
3. Armache, K.-J.; Mitterweger, S.; Meinhart, A.; Cramer, P. Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7. *The Journal of biological chemistry* **2005**, *280*, 7131–4.
4. Hirata, A.; Klein, B. J.; Murakami, K. S. The X-Ray Crystal Structure of RNA Polymerase from Archaea. *Nature* **2008**, *451*, 851–854.
5. Qureshi, S. A.; Bell, S. D.; Jackson, S. P. Factor Requirements for Transcription in the Archaeon Sulfolobus Shibatae. *EMBO J* **1997**, *16*, 2927–2936.
6. Hausner, W.; Thomm, M. Events during Initiation of Archaeal Transcription: Open Complex Formation and DNA-Protein Interactions. *Journal of bacteriology* **2001**, *183*, 3025–31.
7. Rowlands, T.; Baumann, P.; Jackson, S. P. The TATA-Binding Protein: A General Transcription Factor in Eukaryotes and Archaebacteria. *Science (New York, N.Y.)* **1994**, *264*, 1326–9.
8. Langer, D.; Hain, J.; Thuriaux, P.; Zillig, W. Transcription in Archaea: Similarity to That in Eucarya. *Proc Natl Acad Sci U S A* **1995**, *92*, 5768–5772.
9. Decker, K. B.; Hinton, D. M. Transcription Regulation at the Core: Similarities among

Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annual review of microbiology* **2013**, *67*, 113–39.

10. Carlo, S. De; Lin, S.-C.; Taatjes, D. J.; Hoenger, A. Molecular Basis of Transcription Initiation in Archaea. *Transcription 1*, 103–11.
11. Belogurov, G. A.; Mooney, R. A.; Svetlov, V.; Landick, R.; Artsimovitch, I. Functional Specialization of Transcription Elongation Factors. *The EMBO journal* **2009**, *28*, 112–22.
12. 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* **2013**, *41*, 7048–7059.
13. D'Heygère, F.; Schwartz, A.; Coste, F.; Castaing, B.; Boudvillain, M. ATP-Dependent Motor Activity of the Transcription Termination Factor Rho from Mycobacterium Tuberculosis. *Nucleic acids research* **2015**, gkv505-.
14. Walker, J. E.; Luyties, O.; Santangelo, T. J. Factor-Dependent Archaeal Transcription Termination. *Proceedings of the National Academy of Sciences* **2017**, *114*.
15. Bartlett, M. S.; Thomm, M.; Geiduschek, E. P. The Orientation of DNA in an Archaeal Transcription Initiation Complex. *Nature structural biology* **2000**, *7*, 782–5.
16. Grunberg, S.; Bartlett, M. S.; Naji, S.; Thomm, M. Transcription Factor E Is a Part of Transcription Elongation Complexes. *J Biol Chem* **2007**, *282*, 35482–35490.
17. Bell, S. D.; Brinkman, A. B.; Oost, J. van der; Jackson, S. P. The Archaeal TFIIEalpha Homologue Facilitates Transcription Initiation by Enhancing TATA-Box Recognition. *EMBO reports* **2001**, *2*, 133–8.
18. Blombach, F.; Salvadori, E.; Fouqueau, T.; Yan, J.; Reimann, J.; Sheppard, C.; Smollett, K. L.; Albers, S. V; Kay, C. W.; Thalassinos, K.; Werner, F. Archaeal TFEα/β Is a Hybrid of TFIIE and the RNA Polymerase III Subcomplex hRPC62/39. *eLife* **2015**, *4*.
19. Hanzelka, B. L.; Darcy, T. J.; Reeve, J. N. TFE, an Archaeal Transcription Factor in Methanobacterium Thermoautotrophicum Related to Eucaryal Transcription Factor

TFIIEalpha. *Journal of bacteriology* **2001**, *183*, 1813–8.

20. Santangelo, T. J.; Reeve, J. N. Archaeal RNA Polymerase Is Sensitive to Intrinsic Termination Directed by Transcribed and Remote Sequences. *J Mol Biol* **2006**, *355*, 196–210.

21. Santangelo, T. J.; Cubonova, L.; Skinner, K. M.; Reeve, J. N. Archaeal Intrinsic Transcription Termination in Vivo. *J Bacteriol* **2009**, *191*, 7102–7108.

22. Santangelo, T. J.; Cubonova, L.; James, C. L.; Reeve, J. N.; 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* **2007**, *367*, 344–357.

23. Gietl, A.; Holzmeister, P.; Blombach, F.; Schulz, S.; Voithenberg, L. V von; Lamb, D. C.; Werner, F.; Tinnefeld, P.; Grohmann, D. Eukaryotic and Archaeal TBP and TFB/TF(II)B Follow Different Promoter DNA Bending Pathways. *Nucleic Acids Res* **2014**, *42*, 6219–6231.

24. Ochs, S. M.; Thumann, S.; Richau, R.; Weirauch, M. T.; Lowe, T. M.; Thomm, M.; Hausner, W. Activation of Archaeal Transcription Mediated by Recruitment of Transcription Factor B. *The Journal of biological chemistry* **2012**, *287*, 18863–71.

25. Renfrow, M. B.; Naryshkin, N.; Lewis, L. M.; Chen, H. T.; Ebright, R. H.; Scott, R. A. Transcription Factor B Contacts Promoter DNA near the Transcription Start Site of the Archaeal Transcription Initiation Complex. *J Biol Chem* **2004**, *279*, 2825–2831.

26. Goede, B.; Naji, S.; Kampen, O. von; Ilg, K.; Thomm, M. Protein-Protein Interactions in the Archaeal Transcriptional Machinery: Binding Studies of Isolated RNA Polymerase Subunits and Transcription Factors. *J Biol Chem* **2006**, *281*, 30581–30592.

27. Walker, J. E.; Santangelo, T. J. Analyses of in Vivo Interactions between Transcription Factors and the Archaeal RNA Polymerase. *Methods (San Diego, Calif.)* **2015**, *86*, 73–9.

28. Ouhammouch, M.; Dewhurst, R. E.; Hausner, W.; Thomm, M.; Geiduschek, E. P. Activation of Archaeal Transcription by Recruitment of the TATA-Binding Protein. *Proc*

Natl Acad Sci U S A **2003**, *100*, 5097–5102.

29. Bell, S. D.; Kosa, P. L.; Sigler, P. B.; Jackson, S. P. Orientation of the Transcription Preinitiation Complex in Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 13662–7.

30. Karr, E. A. Transcription Regulation in the Third Domain. *Advances in applied microbiology* **2014**, *89*, 101–33.

31. Ng, W. V; Kennedy, S. P.; Mahairas, G. G.; Berquist, B.; Pan, M.; Shukla, H. D.; Lasky, S. R.; Baliga, N. S.; Thorsson, V.; Sbrogna, J.; Swartzell, S.; Weir, D.; Hall, J.; Dahl, T. A.; Welti, R.; Goo, Y. A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M. J.; Hough, D. W.; Maddocks, D. G.; Jablonski, P. E.; Krebs, M. P.; Angevine, C. M.; Dale, H.; Isenbarger, T. A.; Peck, R. F.; Pohlschroder, M.; Spudich, J. L.; Jung, K. W.; Alam, M.; Freitas, T.; Hou, S.; Daniels, C. J.; Dennis, P. P.; Omer, A. D.; Ebhardt, H.; Lowe, T. M.; Liang, P.; Riley, M.; Hood, L.; DasSarma, S. Genome Sequence of Halobacterium Species NRC-1. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 12176–81.

32. Baliga, N. S.; Bonneau, R.; Facciotti, M. T.; Pan, M.; Glusman, G.; Deutsch, E. W.; Shannon, P.; Chiu, Y.; Weng, R. S.; Gan, R. R.; Hung, P.; Date, S. V; Marcotte, E.; Hood, L.; Ng, W. V. Genome Sequence of Haloarcula Marismortui: A Halophilic Archaeon from the Dead Sea. *Genome research* **2004**, *14*, 2221–34.

33. Goo, Y. A.; Yi, E. C.; Baliga, N. S.; Tao, W. A.; Pan, M.; Aebersold, R.; Goodlett, D. R.; Hood, L.; Ng, W. V Proteomic Analysis of an Extreme Halophilic Archaeon, Halobacterium Sp. NRC-1. *Molecular & cellular proteomics : MCP* **2003**, *2*, 506–24.

34. Micorescu, M.; Grunberg, S.; Franke, A.; Cramer, P.; Thomm, M.; Bartlett, M. Archaeal Transcription: Function of an Alternative Transcription Factor B from Pyrococcus Furiosus. *Journal of Bacteriology* **2007**, *190*, 157–167.

35. Reichlen, M. J.; Murakami, K. S.; Ferry, J. G. Functional Analysis of the Three TATA

Binding Protein Homologs in Methanosarcina Acetivorans. - PubMed - NCBI http://www.ncbi.nlm.nih.gov/pubmed/20081030 (accessed Jan 5, 2016).

36. Bartlett, M. S.; Thomm, M.; Geiduschek, E. P. Topography of the Euryarchaeal Transcription Initiation Complex. *J Biol Chem* **2004**, *279*, 5894–5903.

37. Coker, J. A.; DasSarma, S. Genetic and Transcriptomic Analysis of Transcription Factor Genes in the Model Halophilic Archaeon: Coordinate Action of TbpD and TfbA. *BMC genetics* **2007**, *8*, 61.

38. Hidese, R.; Nishikawa, R.; Gao, L.; Katano, M.; Imai, T.; Kato, S.; Kanai, T.; Atomi, H.; Imanaka, T.; Fujiwara, S. Different Roles of Two Transcription Factor B Proteins in the Hyperthermophilic Archaeon Thermococcus Kodakarensis. *Extremophiles : life under extreme conditions* **2014**, *18*, 573–88.

39. Bonneau, R.; Facciotti, M. T.; Reiss, D. J.; Schmid, A. K.; Pan, M.; Kaur, A.; Thorsson, V.; Shannon, P.; Johnson, M. H.; Bare, J. C.; Longabaugh, W.; Vuthoori, M.; Whitehead, K.; Madar, A.; Suzuki, L.; Mori, T.; Chang, D.-E.; Diruggiero, J.; Johnson, C. H.; Hood, L.; Baliga, N. S. A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell* **2007**, *131*, 1354–65.

40. Baliga, N. S.; Goo, Y. A.; Ng, W. V; Hood, L.; Daniels, C. J.; DasSarma, S. Is Gene Expression in Halobacterium NRC-1 Regulated by Multiple TBP and TFB Transcription Factors? *Molecular microbiology* **2000**, *36*, 1184–5.

41. Turkarslan, S.; Reiss, D. J.; Gibbins, G.; Su, W. L.; Pan, M.; Bare, J. C.; Plaisier, C. L.; Baliga, N. S. Niche Adaptation by Expansion and Reprogramming of General Transcription Factors. *Molecular systems biology* **2011**, *7*, 554.

42. Paytubi, S.; White, M. F. The Crenarchaeal DNA Damage-Inducible Transcription Factor B Paralogue TFB3 Is a General Activator of Transcription. *Molecular microbiology* **2009**, *72*, 1487–99.

43. Österberg, S.; Peso-Santos, T. del; Shingler, V. Regulation of Alternative Sigma Factor

Use. *Annual review of microbiology* **2011**, *65*, 37–55.

44. Gehring, A. M.; Santangelo, T. J. Manipulating Archaeal Systems to Permit Analyses of Transcription Elongation-Termination Decisions in Vitro. *Methods in molecular biology (Clifton, N.J.)* **2015**, *1276*, 263–79.

45. 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* **2008**, *70*, 623–633.

46. DeDecker, B. S.; O'Brien, R.; Fleming, P. J.; Geiger, J. H.; Jackson, S. P.; Sigler, P. B. The Crystal Structure of a Hyperthermophilic Archaeal TATA-Box Binding Protein. *Journal of molecular biology* **1996**, *264*, 1072–84.

47. Littlefield, O.; Korkhin, Y.; Sigler, P. B. The Structural Basis for the Oriented Assembly of a TBP/TFB/promoter Complex. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 13668–73.

48. Kassavetis, G. A.; Joazeiro, C. A.; Pisano, M.; Geiduschek, E. P.; Colbert, T.; Hahn, S.; Blanco, J. A. The Role of the TATA-Binding Protein in the Assembly and Function of the Multisubunit Yeast RNA Polymerase III Transcription Factor, TFIIIB. *Cell* **1992**, *71*, 1055–64.

49. Knutson, B. A.; Hahn, S. Yeast Rrn7 and Human TAF1B Are TFIIB-Related RNA Polymerase I General Transcription Factors. *Science (New York, N.Y.)* **2011**, *333*, 1637–40.

50. Adachi, N.; Senda, M.; Natsume, R.; Senda, T.; Horikoshi, M. Crystal Structure of Methanococcus Jannaschii TATA Box-Binding Protein. *Genes to cells : devoted to molecular & cellular mechanisms* **2008**, *13*, 1127–40.

51. Qureshi, S. A.; Jackson, S. P. Sequence-Specific DNA Binding by the S. Shibatae TFIIB Homolog, TFB, and Its Effect on Promoter Strength. *Molecular Cell* **1998**, *1*, 389–400.

52. Bleiholder, A.; Frommherz, R.; Teufel, K.; Pfeifer, F. Expression of Multiple Tfb Genes in Different Halobacterium Salinarum Strains and Interaction of TFB with Transcriptional Activator GvpE. *Archives of microbiology* **2012**, *194*, 269–79.

53. 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. *Nature Communications* **2015**, *6*, 6161.

54. Werner, F.; Weinzierl, R. O. Direct Modulation of RNA Polymerase Core Functions by Basal Transcription Factors. *Mol Cell Biol* **2005**, *25*, 8344–8355.

55. Ohkuma, Y.; Sumimoto, H.; Horikoshi, M.; Roeder, R. G. Factors Involved in Specific Transcription by Mammalian RNA Polymerase II: Purification and Characterization of General Transcription Factor TFIIE. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 9163–7.

56. Gregory Peterson, M.; Inostroza, J.; Maxon, M. E.; Flores, O.; Admon, A.; Reinberg, D.; Tjian, R. Structure and Functional Properties of Human General Transcription Factor IIE. *Nature* **1991**, *354*, 369–373.

57. Vannini, A.; Cramer, P. Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Molecular Cell* **2012**, *45*, 439–446.

58. Lefèvre, S.; Dumay-Odelot, H.; El-Ayoubi, L.; Budd, A.; Legrand, P.; Pinaud, N.; Teichmann, M.; Fribourg, S. Structure-Function Analysis of hRPC62 Provides Insights into RNA Polymerase III Transcription Initiation. *Nature Structural & Molecular Biology* **2011**, *18*, 352–358.

59. Carter, R.; Drouin, G. The Increase in the Number of Subunits in Eukaryotic RNA Polymerase III Relative to RNA Polymerase II Is due to the Permanent Recruitment of General Transcription Factors. *Molecular biology and evolution* **2010**, *27*, 1035–43.

60. Naji, S.; Grunberg, S.; Thomm, M. The RPB7 Orthologue E' Is Required for Transcriptional Activity of a Reconstituted Archaeal Core Enzyme at Low Temperatures

and Stimulates Open Complex Formation. *J Biol Chem* **2007**, *282*, 11047–11057.

61. Werner, F.; Weinzierl, R. O. A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription. *Mol Cell* **2002**, *10*, 635–646.

62. Meinhart, A.; Blobel, J.; Cramer, P. An Extended Winged Helix Domain in General Transcription Factor E/IIE Alpha. *The Journal of biological chemistry* **2003**, *278*, 48267–74.

63. Okuda, M.; Tanaka, A.; Arai, Y.; Satoh, M.; Okamura, H.; Nagadoi, A.; Hanaoka, F.; Ohkuma, Y.; Nishimura, Y. A Novel Zinc Finger Structure in the Large Subunit of Human General Transcription Factor TFIIE. *The Journal of biological chemistry* **2004**, *279*, 51395–403.

64. Hirtreiter, A.; Grohmann, D.; Werner, F. Molecular Mechanisms of RNA Polymerase--the F/E (RPB4/7) Complex Is Required for High Processivity in Vitro. *Nucleic Acids Res* **2010**, *38*, 585–596.

65. Ouhammouch, M.; Werner, F.; Weinzierl, R. O. J.; Geiduschek, E. P. A Fully Recombinant System for Activator-Dependent Archaeal Transcription. *Journal of Biological Chemistry* **2004**, *279*, 51719–51721.

66. Thomm, M.; Wich, G. An Archaebacterial Promoter Element for Stable RNA Genes with Homology to the TATA Box of Higher Eukaryotes. *Nucleic acids research* **1988**, *16*, 151–63.

67. Reiter, W. D.; Hüdepohl, U.; Zillig, W. Mutational Analysis of an Archaebacterial Promoter: Essential Role of a TATA Box for Transcription Efficiency and Start-Site Selection in Vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 9509–13.

68. Reiter, W. D.; Palm, P.; Zillig, W. Analysis of Transcription in the Archaebacterium Sulfolobus Indicates That Archaebacterial Promoters Are Homologous to Eukaryotic Pol II Promoters. *Nucleic acids research* **1988**, *16*, 1–19.

69. Hain, J.; Reiter, W. D.; Hüdepohl, U.; Zillig, W. Elements of an Archaeal Promoter Defined by Mutational Analysis. *Nucleic acids research* **1992**, *20*, 5423–8.

70. Peng, N.; Xia, Q.; Chen, Z.; Liang, Y. X.; She, Q. An Upstream Activation Element Exerting Differential Transcriptional Activation on an Archaeal Promoter. *Molecular microbiology* **2009**, *74*, 928–39.

71. Brenneis, M.; Hering, O.; Lange, C.; Soppa, J. Experimental Characterization of Cis-Acting Elements Important for Translation and Transcription in Halophilic Archaea. *PLoS genetics* **2007**, *3*, e229.

72. Ao, X.; Li, Y.; Wang, F.; Feng, M.; Lin, Y.; Zhao, S.; Liang, Y.; Peng, N. The Sulfolobus Initiator Element Is an Important Contributor to Promoter Strength. *Journal of bacteriology* **2013**, *195*, 5216–22.

73. Chakraborty, A.; Wang, D.; Ebright, Y. W.; Korlann, Y.; Kortkhonjia, E.; Kim, T.; Chowdhury, S.; Wigneshweraraj, S.; Irschik, H.; Jansen, R.; Nixon, B. T.; Knight, J.; Weiss, S.; Ebright, R. H. Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science* **2012**, *337*, 591–595.

74. Blombach, F.; Daviter, T.; Fielden, D.; Grohmann, D.; Smollett, K.; Werner, F. Archaeology of RNA Polymerase: Factor Swapping during the Transcription Cycle. *Biochem Soc Trans* **2013**, *41*, 362–367.

75. Jun, S.-H. H.; Hirata, A.; Kanai, T.; Santangelo, T. J.; Imanaka, T.; Murakami, K. S. The X-Ray Crystal Structure of the Euryarchaeal RNA Polymerase in an Open-Clamp Configuration. *Nat Commun* **2014**, *5*, 5132.

76. Hartzog, G. a; Kaplan, C. D. Competing for the Clamp: Promoting RNA Polymerase Processivity and Managing the Transition from Initiation to Elongation. *Molecular cell* **2011**, *43*, 161–3.

77. Grohmann, D.; Klose, D.; Klare, J. P.; Kay, C. W. M.; Steinhoff, H.-J.; Werner, F. RNA-Binding to Archaeal RNA Polymerase Subunits F/E: A DEER and FRET Study. *Journal of*

the American Chemical Society **2010**, *132*, 5954–5955.

78. Wardleworth, B. N.; Russell, R. J. M.; Bell, S. D.; Taylor, G. L.; White, M. F. Structure of Alba: An Archaeal Chromatin Protein Modulated by Acetylation. *The EMBO journal* **2002**, *21*, 4654–62.

79. Marsh, V. L.; Peak-Chew, S. Y.; Bell, S. D. Sir2 and the Acetyltransferase, Pat, Regulate the Archaeal Chromatin Protein, Alba. *The Journal of biological chemistry* **2005**, *280*, 21122–8.

80. Goyal, M.; Banerjee, C.; Nag, S.; Bandyopadhyay, U. The Alba Protein Family: Structure and Function. *Biochimica et biophysica acta* **2016**, *1864*, 570–583.

81. Bell, S. D.; Botting, C. H.; Wardleworth, B. N.; Jackson, S. P.; White, M. F. The Interaction of Alba, a Conserved Archaeal Chromatin Protein, with Sir2 and Its Regulation by Acetylation. *Science (New York, N.Y.)* **2002**, *296*, 148–51.

82. Tang, T. H.; Rozhdestvensky, T. S.; d'Orval, B. C.; Bortolin, M.-L.; Huber, H.; Charpentier, B.; Branlant, C.; Bachellerie, J.-P.; Brosius, J.; Hüttenhofer, A. RNomics in Archaea Reveals a Further Link between Splicing of Archaeal Introns and rRNA Processing. *Nucleic acids research* **2002**, *30*, 921–30.

83. Tocchini-Valentini, G. D.; Fruscoloni, P.; Tocchini-Valentini, G. P. Structure, Function, and Evolution of the tRNA Endonucleases of Archaea: An Example of Subfunctionalization. *Proceedings of the National Academy of Sciences* **2005**, *102*, 8933–8938.

84. Kjems, J.; Jensen, J.; Olesen, T.; Garrett, R. A. Comparison of Transfer RNA and Ribosomal RNA Intron Splicing in the Extreme Thermophile and Archaebacterium Desulfurococcus Mobilis. *Canadian journal of microbiology* **1989**, *35*, 210–4.

85. Martinez-Rucobo, F. W.; Sainsbury, S.; Cheung, A. C. M.; Cramer, P. Architecture of the RNA Polymerase-Spt4/5 Complex and Basis of Universal Transcription Processivity. *The EMBO journal* **2011**, *30*, 1302–10.

86. Zhou, H.; Liu, Q.; Gao, Y.; Teng, M.; Niu, L. Crystal Structure of NusG N-Terminal (NGN) Domain from Methanocaldococcus Jannaschii and Its Interaction with rpoE''. *Proteins* **2009**, *76*, 787–93.

87. Hausner, W.; Lange, U.; Musfeldt, M. Transcription Factor S, a Cleavage Induction Factor of the Archaeal RNA Polymerase. *The Journal of biological chemistry* **2000**, *275*, 12393–9.

88. Langer, D.; Zillig, W. Putative tfIIs Gene of Sulfolobus Acidocaldarius Encoding an Archaeal Transcription Elongation Factor Is Situated Directly Downstream of the Gene for a Small Subunit of DNA-Dependent RNA Polymerase. *Nucleic acids research* **1993**, *21*, 2251.

89. Klein, B. J.; Bose, D.; Baker, K. J.; Yusoff, Z. M.; Zhang, X.; Murakami, K. S. RNA Polymerase and Transcription Elongation Factor Spt4/5 Complex Structure. *Proc Natl Acad Sci U S A* **2011**, *108*, 546–550.

90. 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 MjSpt4p:Spt5 Complex at the Exit Tunnel of RNAPII. *Journal of structural biology* **2015**.

91. Marr, M. T.; Roberts, J. W. Function of Transcription Cleavage Factors GreA and GreB at a Regulatory Pause Site. *Molecular cell* **2000**, *6*, 1275–85.

92. Erie, D. A.; Hajiseyedjavadi, O.; Young, M. C.; Hippel, P. H. von Multiple RNA Polymerase Conformations and GreA: Control of the Fidelity of Transcription. *Science (New York, N.Y.)* **1993**, *262*, 867–73.

93. Grünberg, S.; Reich, C.; Zeller, M. E.; Bartlett, M. S.; Thomm, M. Rearrangement of the RNA Polymerase Subunit H and the Lower Jaw in Archaeal Elongation Complexes. *Nucleic Acids Research* **2010**, *38*, 1950–1963.

94. Lange, U.; Hausner, W. Transcriptional Fidelity and Proofreading in Archaea and Implications for the Mechanism of TFS-Induced RNA Cleavage. *Molecular microbiology*

2004, *52*, 1133–43.

95. Guo, M.; Xu, F.; Yamada, J.; Egelhofer, T.; Gao, Y.; Hartzog, G. A.; Teng, M.; Niu, L. Core Structure of the Yeast spt4-spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. *Structure (London, England : 1993)* **2008**, *16*, 1649–58.
96. Meyer, P. A.; Li, S.; Zhang, M.; Yamada, K.; Takagi, Y.; Hartzog, G. A.; Fu, J. Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Molecular and cellular biology* **2015**, *35*, 3354–69.
97. Yakhnin, A. V; Babitzke, P. NusG/Spt5: Are There Common Functions of This Ubiquitous Transcription Elongation Factor? *Current opinion in microbiology* **2014**, *18*, 68–71.
98. Sevostyanova, A.; Belogurov, G. a; Mooney, R. a; Landick, R.; Artsimovitch, I. The β Subunit Gate Loop Is Required for RNA Polymerase Modification by RfaH and NusG. *Molecular cell* **2011**, *43*, 253–62.
99. Mooney, R. A.; Schweimer, K.; Rösch, P.; Gottesman, M.; Landick, R. Two Structurally Independent Domains of E. Coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. *Journal of Molecular Biology* **2009**, *391*, 341–358.
100. Drögemüller, J.; Strauß, M.; Schweimer, K.; Jurk, M.; Rösch, P.; Knauer, S. H. Determination of RNA Polymerase Binding Surfaces of Transcription Factors by NMR Spectroscopy. *Scientific Reports* **2015**, *5*, 16428.
101. Burova, E.; Hung, S. C.; Sagitov, V.; Stitt, B. L.; Gottesman, M. E. Escherichia Coli NusG Protein Stimulates Transcription Elongation Rates in Vivo and in Vitro. *Journal of bacteriology* **1995**, *177*, 1388–92.
102. Herbert, K. M.; Zhou, J.; Mooney, R. a; Porta, A. La; Landick, R.; Block, S. M. E. Coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase. *Journal of molecular biology* **2010**, *399*, 17–30.

103. Belogurov, G. A.; Artsimovitch, I. Regulation of Transcript Elongation. *Annual Review of Microbiology* **2015**, *69*, 49–69.

104. Yakhnin, A. V; Murakami, K. S.; Babitzke, P. NusG Is a Sequence-Specific RNA Polymerase Pause Factor That Binds to the Non-Template DNA within the Paused Transcription Bubble. *The Journal of biological chemistry* **2016**, *291*, 5299–308.

105. Crickard, J. B.; Fu, J.; Reese, J. C. Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. *The Journal of biological chemistry* **2016**, *291*, 9853–70.

106. Yakhnin, A. V; Yakhnin, H.; Babitzke, P. Function of the Bacillus Subtilis Transcription Elongation Factor NusG in Hairpin-Dependent RNA Polymerase Pausing in the Trp Leader. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 16131–6.

107. Anderson, S. J.; Sikes, M. L.; Zhang, Y.; French, S. L.; Salgia, S.; Beyer, A. L.; Nomura, M.; Schneider, D. A. The Transcription Elongation Factor Spt5 Influences Transcription by RNA Polymerase I Positively and Negatively. *The Journal of biological chemistry* **2011**, *286*, 18816–24.

108. Tomar, S. K.; Artsimovitch, I. NusG-Spt5 Proteins-Universal Tools for Transcription Modification and Communication. *Chemical reviews* **2013**, *113*, 8604–19.

109. Banerjee, S.; Chalissery, J.; Bandey, I.; Sen, R. Rho-Dependent Transcription Termination: More Questions than Answers. *J Microbiol* **2006**, *44*, 11–22.

110. Burmann, B. M.; Luo, X.; Rösch, P.; Wahl, M. C.; Gottesman, M. E. Fine Tuning of the E. Coli NusB:NusE Complex Affinity to BoxA RNA Is Required for Processive Antitermination. *Nucleic acids research* **2010**, *38*, 314–26.

111. Proshkin, S.; Rahmouni, A. R.; Mironov, A.; Nudler, E. Cooperation between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science (New York, N.Y.)* **2010**, *328*, 504–8.

112. Peters, J. M.; Mooney, R. A.; Grass, J. A.; Jessen, E. D.; Tran, F.; Landick, R. Rho and NusG Suppress Pervasive Antisense Transcription in Escherichia Coli. *Genes Dev* **2012**, *26*, 2621–2633.

113. Burns, C. M.; Richardson, J. P. NusG Is Required to Overcome a Kinetic Limitation to Rho Function at an Intragenic Terminator. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 4738–42.

114. French, S. L.; Santangelo, T. J.; Beyer, A. L.; Reeve, J. N. Transcription and Translation Are Coupled in Archaea. *Mol Biol Evol* **2007**, *24*, 893–895.

115. Santangelo, T. J.; Cubonova, L.; Matsumi, R.; Atomi, H.; Imanaka, T.; Reeve, J. N. Polarity in Archaeal Operon Transcription in Thermococcus Kodakaraensis. *J Bacteriol* **2008**, *190*, 2244–2248.

116. Gnatt, A. L.; Cramer, P.; Fu, J.; Bushnell, D. A.; Kornberg, R. D. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. *Science (New York, N.Y.)* **2001**, *292*, 1876–82.

117. Werner, F.; Grohmann, D. Evolution of Multisubunit RNA Polymerases in the Three Domains of Life. *Nat Rev Microbiol* **2011**, *9*, 85–98.

118. 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. *Proceedings of the National Academy of Sciences of the United States of America* **2016**.

119. Grohmann, D.; Klose, D.; Fielden, D.; Werner, F. FRET (Fluorescence Resonance Energy Transfer) Sheds Light on Transcription. *Biochemical Society Transactions* **2011**, *39*, 122–127.

120. Tanaka, A.; Akimoto, Y.; Kobayashi, S.; Hisatake, K.; Hanaoka, F.; Ohkuma, Y. Association of the Winged Helix Motif of the TFIIEα Subunit of TFIIE with Either the TFIIEβ Subunit or TFIIB Distinguishes Its Functions in Transcription. *Genes to Cells* **2015**, *20*, 203–216.

121. Vassylyev, D. G.; Vassylyeva, M. N.; Perederina, A.; Tahirov, T. H.; Artsimovitch, I. Structural Basis for Transcription Elongation by Bacterial RNA Polymerase. *Nature* **2007**, *448*, 157–162.

122. Kettenberger, H.; Armache, K.-J.; Cramer, P. Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Molecular Cell* **2004**, *16*, 955–965.

123. Kireeva, M. L.; Komissarova, N.; Waugh, D. S.; Kashlev, M. The 8-Nucleotide-Long RNA:DNA Hybrid Is a Primary Stability Determinant of the RNA Polymerase II Elongation Complex. *The Journal of biological chemistry* **2000**, *275*, 6530–6.

124. Andrecka, J.; Treutlein, B.; Arcusa, M. A. I.; Muschielok, A.; Lewis, R.; Cheung, A. C. M.; Cramer, P.; Michaelis, J. Nano Positioning System Reveals the Course of Upstream and Nontemplate DNA within the RNA Polymerase II Elongation Complex. *Nucleic acids research* **2009**, *37*, 5803–9.

125. Reeder, T. C.; Hawley, D. K. Promoter Proximal Sequences Modulate RNA Polymerase II Elongation by a Novel Mechanism. *Cell* **1996**, *87*, 767–77.

126. Nudler, E.; Avetissova, E.; Markovtsov, V.; Goldfarb, A. Transcription Processivity: Protein-DNA Interactions Holding Together the Elongation Complex. *Science (New York, N.Y.)* **1996**, *273*, 211–7.

127. Sidorenkov, I.; Komissarova, N.; Kashlev, M. Crucial Role of the RNA:DNA Hybrid in the Processivity of Transcription. *Molecular cell* **1998**, *2*, 55–64.

128. Hein, P. P.; Kolb, K. E.; Windgassen, T.; Bellecourt, M. J.; Darst, S. A.; Mooney, R. A.; Landick, R. RNA Polymerase Pausing and Nascent-RNA Structure Formation Are Linked through Clamp-Domain Movement. *Nat Struct Mol Biol* **2014**, *21*, 794–802.

129. Komissarova, N.; Becker, J.; Solter, S.; Kireeva, M.; Kashlev, M. Shortening of RNA:DNA Hybrid in the Elongation Complex of RNA Polymerase Is a Prerequisite for Transcription Termination. *Molecular cell* **2002**, *10*, 1151–62.

130. Bochkareva, A.; Yuzenkova, Y.; Tadigotla, V. R.; Zenkin, N. Factor-Independent Transcription Pausing Caused by Recognition of the RNA-DNA Hybrid Sequence. *The EMBO Journal* **2012**, *31*, 630–639.

131. Gusarov, I.; Nudler, E. The Mechanism of Intrinsic Transcription Termination. *Molecular cell* **1999**, *3*, 495–504.

132. Martin, F. H.; Tinoco, I. DNA-RNA Hybrid Duplexes Containing oligo(dA:rU) Sequences Are Exceptionally Unstable and May Facilitate Termination of Transcription. *Nucleic acids research* **1980**, *8*, 2295–9.

133. Czyz, A.; Mooney, R. A.; Iaconi, A.; Landick, R. Mycobacterial RNA Polymerase Requires a U-Tract at Intrinsic Terminators and Is Aided by NusG at Suboptimal Terminators. *mBio* **2014**, *5*, e00931.

134. Arimbasseri, A. G.; Rijal, K.; Maraia, R. J. Transcription Termination by the Eukaryotic RNA Polymerase III. *Biochimica et biophysica acta 1829*, 318–30.

135. Reiter, W. D.; Palm, P.; Zillig, W. Transcription Termination in the Archaebacterium Sulfolobus: Signal Structures and Linkage to Transcription Initiation. *Nucleic acids research* **1988**, *16*, 2445–59.

136. Santangelo, T. J.; Roberts, J. W. Forward Translocation Is the Natural Pathway of RNA Release at an Intrinsic Terminator. *Mol Cell* **2004**, *14*, 117–126.

137. Reynolds, R.; Bermúdez-Cruz, R. M.; Chamberlin, M. J. Parameters Affecting Transcription Termination by Escherichia Coli RNA Polymerase. I. Analysis of 13 Rho-Independent Terminators. *Journal of molecular biology* **1992**, *224*, 31–51.

138. Spitalny, P.; Thomm, M. A Polymerase III-like Reinitiation Mechanism Is Operating in Regulation of Histone Expression in Archaea. *Mol Microbiol* **2008**, *67*, 958–970.

139. Dieci, G.; Sentenac, A. Facilitated Recycling Pathway for RNA Polymerase III. *Cell* **1996**, *84*, 245–52.

140. Dedrick, R. L.; Kane, C. M.; Chamberlin, M. J. Purified RNA Polymerase II Recognizes

Specific Termination Sites during Transcription in Vitro. *The Journal of biological chemistry* **1987**, *262*, 9098–108.

141. Palangat, M.; Hittinger, C. T.; Landick, R. Downstream DNA Selectively Affects a Paused Conformation of Human RNA Polymerase II. *Journal of Molecular Biology* **2004**, *341*, 429–442.

142. Kerppola, T. K.; Kane, C. M. Intrinsic Sites of Transcription Termination and Pausing in the c-Myc Gene. *Molecular and cellular biology* **1988**, *8*, 4389–94.

143. 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* **2002**, *109*, 757–767.

144. Roberts, J. W. Termination Factor for RNA Synthesis. *Nature* **1969**, *224*, 1168–74.

145. 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. *Molecular Cell* **2015**, *60*, 256–267.

146. Merkl, P.; Perez-Fernandez, J.; Pilsl, M.; Reiter, A.; Williams, L.; Gerber, J.; Bohm, M.; Deutzmann, R.; Griesenbeck, J.; Milkereit, P.; Tschochner, H. Binding of the Termination Factor Nsi1 to Its Cognate DNA Site Is Sufficient To Terminate RNA Polymerase I Transcription In Vitro and To Induce Termination In Vivo. *Molecular and Cellular Biology* **2014**, *34*, 3817–3827.

147. Howan, K.; Smith, A. J.; Westblade, L. F.; Joly, N.; Grange, W.; Zorman, S.; Darst, S. A.; Savery, N. J.; Strick, T. R. Initiation of Transcription-Coupled Repair Characterized at Single-Molecule Resolution. *Nature* **2012**, *490*, 431–4.

148. Leela, J. K.; Syeda, A. H.; Anupama, K.; Gowrishankar, J. Rho-Dependent Transcription Termination Is Essential to Prevent Excessive Genome-Wide R-Loops in Escherichia Coli. *Proceedings of the National Academy of Sciences of the United States of America*

2013, *110*, 258–63.

149. Smith, A. J.; Pernstich, C.; Savery, N. J. Multipartite Control of the DNA Translocase, Mfd. *Nucleic acids research* **2012**, *40*, 10408–16.

150. Park, J.-S.; Roberts, J. W. Role of DNA Bubble Rewinding in Enzymatic Transcription Termination. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 4870–5.

151. Haines, N. M.; Kim, Y.-I. T.; Smith, A. J.; Savery, N. J. Stalled Transcription Complexes Promote DNA Repair at a Distance. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 4037–42.

152. Adhya, S.; Gottesman, M.; Crombrugghe, B. De Release of Polarity in Escherichia Coli by Gene N of Phage Lambda: Termination and Antitermination of Transcription. *Proceedings of the National Academy of Sciences of the United States of America* **1974**, *71*, 2534–8.

153. Menouni, R.; Champ, S.; Espinosa, L.; Boudvillain, M.; Ansaldi, M. Transcription Termination Controls Prophage Maintenance in Escherichia Coli Genomes. *Proceedings of the National Academy of Sciences* **2013**, *110*, 14414–14419.

154. 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 (New York, N.Y.)* **2008**, *320*, 935–8.

155. Witkin, E. M. Radiation-Induced Mutations and Their Repair. *Science (New York, N.Y.)* **1966**, *152*, 1345–53.

156. Gehring, A. M.; Santangelo, T. J. Archaeal RNA Polymerase Arrests Transcription at DNA Lesions. *Transcription* **2017**, e1324941.

157. Peeters, E.; Driessen, R. P. C.; Werner, F.; Dame, R. T. The Interplay between Nucleoid Organization and Transcription in Archaeal Genomes. *Nature reviews. Microbiology* **2015**, *13*, 333–41.

158. Zerulla, K.; Soppa, J. Polyploidy in Haloarchaea: Advantages for Growth and Survival. *Frontiers in microbiology* **2014**, *5*, 274.

159. Spaans, S. K.; Oost, J. van der; Kengen, S. W. M. The Chromosome Copy Number of the Hyperthermophilic Archaeon Thermococcus Kodakarensis KOD1. *Extremophiles : life under extreme conditions* **2015**, *19*, 741–50.

160. Jaakkola, S. T.; Zerulla, K.; Guo, Q.; Liu, Y.; Ma, H.; Yang, C.; Bamford, D. H.; Chen, X.; Soppa, J.; Oksanen, H. M. Halophilic Archaea Cultivated from Surface Sterilized Middle-Late Eocene Rock Salt Are Polyploid. *PloS one* **2014**, *9*, e110533.

161. Hildenbrand, C.; Stock, T.; Lange, C.; Rother, M.; Soppa, J. Genome Copy Numbers and Gene Conversion in Methanogenic Archaea. *Journal of bacteriology* **2011**, *193*, 734–43.

162. Breuert, S.; Allers, T.; Spohn, G.; Soppa, J. Regulated Polyploidy in Halophilic Archaea. *PloS one* **2006**, *1*, e92.

163. Mattiroli, 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* **2017**, *357*, 609–612.

164. Pereira, S. L.; Grayling, R. A.; Lurz, R.; Reeve, J. N. Archaeal Nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 12633–7.

165. Driessen, R. P. C.; Dame, R. T. Structure and Dynamics of the Crenarchaeal Nucleoid. *Biochemical Society transactions* **2013**, *41*, 321–5.

166. Driessen, R. P. C.; Dame, R. T. Nucleoid-Associated Proteins in Crenarchaea. *Biochemical Society transactions* **2011**, *39*, 116–21.

167. Dulmage, K. A.; Todor, H.; Schmid, A. K. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *mBio* **2015**, *6*, e00649-15.

168. Cubonovaa, 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* **2012**, *194*, 6864–6874.

169. Heinicke, I.; Müller, J.; Pittelkow, M.; Klein, A. Mutational Analysis of Genes Encoding Chromatin Proteins in the Archaeon Methanococcus Voltae Indicates Their Involvement in the Regulation of Gene Expression. *Molecular genetics and genomics : MGG* **2004**, *272*, 76–87.
170. Ammar, R.; Torti, D.; Tsui, K.; Gebbia, M.; Durbic, T.; Bader, G. D.; Giaever, G.; Nislow, C. Chromatin Is an Ancient Innovation Conserved between Archaea and Eukarya. *eLife* **2012**, *1*, e00078.
171. 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* **2013**, *14*, 391.
172. Forbes, A. J.; Patrie, S. M.; Taylor, G. K.; Kim, Y.-B.; Jiang, L.; Kelleher, N. L. Targeted Analysis and Discovery of Posttranslational Modifications in Proteins from Methanogenic Archaea by Top-down MS. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 2678–83.
173. Xie, Y.; Reeve, J. N. Transcription by an Archaeal RNA Polymerase Is Slowed but Not Blocked by an Archaeal Nucleosome. *J Bacteriol* **2004**, *186*, 3492–3498.
174. Studitsky, V. M. Mechanism of Transcription Through the Nucleosome by Eukaryotic RNA Polymerase. *Science* **1997**, *278*, 1960–1963.
175. Clark, D. J.; Felsenfeld, G. A Nucleosome Core Is Transferred out of the Path of a Transcribing Polymerase. *Cell* **1992**, *71*, 11–22.
176. Maruyama, H.; Harwood, J. C.; Moore, K. M.; Paszkiewicz, K.; Durley, S. C.; Fukushima, H.; Atomi, H.; Takeyasu, K.; Kent, N. A. An Alternative Beads-on-a-String Chromatin Architecture in Thermococcus Kodakarensis. *EMBO reports* **2013**, *14*, 711–7.
177. Weidenbach, K.; Glöer, J.; Ehlers, C.; Sandman, K.; Reeve, J. N.; Schmitz, R. A. Deletion

of the Archaeal Histone in Methanosarcina Mazei Gö1 Results in Reduced Growth and Genomic Transcription. *Molecular microbiology* **2008**, *67*, 662–71.

178. Dinger, M. E.; Baillie, G. J.; Musgrave, D. R. Growth Phase-Dependent Expression and Degradation of Histones in the Thermophilic Archaeon Thermococcus Zilligii. *Molecular microbiology* **2000**, *36*, 876–85.

179. Sandman, K.; Grayling, R. A.; Dobrinski, B.; Lurz, R.; Reeve, J. N. Growth-Phase-Dependent Synthesis of Histones in the Archaeon Methanothermus Fervidus. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 12624–8.

180. Soares, D.; Dahlke, I.; Li, W.-T.; Sandman, K.; Hethke, C.; Thomm, M.; Reeve, J. N. Archaeal Histone Stability, DNA Binding, and Transcription Inhibition above 90°C. *Extremophiles* **1998**, *2*, 75–81.

181. Jelinska, C.; Petrovic-Stojanovska, B.; Ingledew, W. J.; White, M. F. Dimer-Dimer Stacking Interactions Are Important for Nucleic Acid Binding by the Archaeal Chromatin Protein Alba. *The Biochemical journal* **2010**, *427*, 49–55.

182. Lurz, R.; Grote, M.; Dijk, J.; Reinhardt, R.; Dobrinski, B. Electron Microscopic Study of DNA Complexes with Proteins from the Archaebacterium Sulfolobus Acidocaldarius. *The EMBO journal* **1986**, *5*, 3715–21.

183. Jelinska, C.; Conroy, M. J.; Craven, C. J.; Hounslow, A. M.; Bullough, P. A.; Waltho, J. P.; Taylor, G. L.; White, M. F. Obligate Heterodimerization of the Archaeal Alba2 Protein with Alba1 Provides a Mechanism for Control of DNA Packaging. *Structure (London, England : 1993)* **2005**, *13*, 963–71.

184. Laurens, N.; Driessen, R. P. C.; Heller, I.; Vorselen, D.; Noom, M. C.; Hol, F. J. H.; White, M. F.; Dame, R. T.; Wuite, G. J. L. Alba Shapes the Archaeal Genome Using a Delicate Balance of Bridging and Stiffening the DNA. *Nature communications* **2012**, *3*, 1328.

185. Tanaka, T.; Padavattan, S.; Kumarevel, T. Crystal Structure of Archaeal Chromatin

Protein Alba2-Double-Stranded DNA Complex from Aeropyrum Pernix K1. *The Journal of biological chemistry* **2012**, *287*, 10394–402.

186. Wilkinson, S. P.; Ouhammouch, M.; Geiduschek, E. P. Transcriptional Activation in the Context of Repression Mediated by Archaeal Histones. *Proc Natl Acad Sci U S A* **2010**, *107*, 6777–6781.
187. Hanawalt, P. C.; Spivak, G. Transcription-Coupled DNA Repair: Two Decades of Progress and Surprises. *Nat Rev Mol Cell Biol* **2008**, *9*, 958–970.
188. Ganesan, A.; Spivak, G.; Hanawalt, P. C. Transcription-Coupled DNA Repair in Prokaryotes. *Progress in molecular biology and translational science* **2012**, *110*, 25–40.
189. Svejstrup, J. Q. Mechanisms of Transcription-Coupled DNA Repair. *Nat Rev Mol Cell Biol* **2002**, *3*, 21–29.
190. Saijo, M. The Role of Cockayne Syndrome Group A (CSA) Protein in Transcription-Coupled Nucleotide Excision Repair. *Mech Ageing Dev* **2013**, *134*, 196–201.
191. Cho, I.; Tsai, P. F.; Lake, R. J.; Basheer, A.; Fan, H. Y. ATP-Dependent Chromatin Remodeling by Cockayne Syndrome Protein B and NAP1-like Histone Chaperones Is Required for Efficient Transcription-Coupled DNA Repair. *PLoS Genet* **2013**, *9*, e1003407.
192. Wilson, B. T.; Stark, Z.; Sutton, R. E.; Danda, S.; Ekbote, A. V; Elsayed, S. M.; Gibson, L.; Goodship, J. A.; Jackson, A. P.; Keng, W. T.; King, M. D.; McCann, E.; Motojima, T.; Murray, J. E.; Omata, T.; Pilz, D.; Pope, K.; Sugita, K.; White, S. M.; Wilson, I. J. The Cockayne Syndrome Natural History (CoSyNH) Study: Clinical Findings in 102 Individuals and Recommendations for Care. *Genetics in medicine : official journal of the American College of Medical Genetics* **2015**.
193. Kleijer, W. J.; Laugel, V.; Berneburg, M.; Nardo, T.; Fawcett, H.; Gratchev, A.; Jaspers, N. G. J.; Sarasin, A.; Stefanini, M.; Lehmann, A. R. Incidence of DNA Repair Deficiency Disorders in Western Europe: Xeroderma Pigmentosum, Cockayne Syndrome and

Trichothiodystrophy. *DNA repair* **2008**, *7*, 744–50.

194. Rapin, I. Disorders of Nucleotide Excision Repair. *Handbook of clinical neurology* **2013**, *113*, 1637–50.

195. Fan, L.; Fuss, J. O.; Cheng, Q. J.; Arvai, A. S.; Hammel, M.; Roberts, V. A.; Cooper, P. K.; Tainer, J. A. XPD Helicase Structures and Activities: Insights into the Cancer and Aging Phenotypes from XPD Mutations. *Cell* **2008**, *133*, 789–800.

196. Dorazi, R.; Gotz, D.; Munro, S.; Bernander, R.; White, M. F. Equal Rates of Repair of DNA Photoproducts in Transcribed and Non-Transcribed Strands in Sulfolobus Solfataricus. *Mol Microbiol* **2007**, *63*, 521–529.

197. Romano, V.; Napoli, A.; Salerno, V.; Valenti, A.; Rossi, M.; Ciaramella, M. Lack of Strand-Specific Repair of UV-Induced DNA Lesions in Three Genes of the Archaeon Sulfolobus Solfataricus. *J Mol Biol* **2007**, *365*, 921–929.

198. Crowley, D. J.; Boubriak, I.; Berquist, B. R.; Clark, M.; Richard, E.; Sullivan, L.; DasSarma, S.; McCready, S. The uvrA, uvrB and uvrC Genes Are Required for Repair of Ultraviolet Light Induced DNA Photoproducts in Halobacterium Sp. NRC-1. *Saline systems* **2006**, *2*, 11.

199. Stantial, N.; Dumpe, J.; Pietrosimone, K.; Baltazar, F.; Crowley, D. J. Transcription-Coupled Repair of UV Damage in the Halophilic Archaea. *DNA repair* **2016**, *41*, 63–8.

200. Rouillon, C.; White, M. F. The Evolution and Mechanisms of Nucleotide Excision Repair Proteins. *Res Microbiol* **2011**, *162*, 19–26.

201. White, M. F. Structure, Function and Evolution of the XPD Family of Iron–sulfur-Containing 5′→3′ DNA Helicases. *Biochemical Society Transactions* **2009**, *37*.

202. Spivak, G. Nucleotide Excision Repair in Humans. *DNA repair* **2015**, *36*, 13–8.

203. Masai, H.; Matsumoto, S.; You, Z.; Yoshizawa-Sugata, N.; Oda, M. Eukaryotic Chromosome DNA Replication: Where, When, and How? *Annual Review of Biochemistry* **2010**, *79*, 89–130.

204. Raghuraman, M. K.; Winzeler, E. A.; Collingwood, D.; Hunt, S.; Wodicka, L.; Conway, A.; Lockhart, D. J.; Davis, R. W.; Brewer, B. J.; Fangman, W. L. Replication Dynamics of the Yeast Genome. *Science (New York, N.Y.)* **2001**, *294*, 115–21.

205. Pohl, T. J.; Brewer, B. J.; Raghuraman, M. K. Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in Saccharomyces Cerevisiae. *PLoS genetics* **2012**, *8*, e1002677.

206. Leonard, A. C.; Mechali, M. DNA Replication Origins. *Cold Spring Harb Perspect Biol* **2013**, *5*, a010116.

207. Robinson, N. P.; Dionne, I.; Lundgren, M.; Marsh, V. L.; Bernander, R.; Bell, S. D. Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus Solfataricus. *Cell* **2004**, *116*, 25–38.

208. Kelman, Z. The Replication Origin of Archaea Is Finally Revealed. *Trends in biochemical sciences* **2000**, *25*, 521–3.

209. Ausiannikava, D.; Allers, T. Diversity of DNA Replication in the Archaea. *Genes* **2017**, *8*, 56.

210. Myllykallio, H.; Lopez, P.; Lopez-Garcia, P.; Heilig, R.; Saurin, W.; Zivanovic, Y.; Philippe, H.; Forterre, P. Bacterial Mode of Replication with Eukaryotic-like Machinery in a Hyperthermophilic Archaeon. *Science* **2000**, *288*, 2212–2215.

211. Ojha, K. K.; Swati, D. Mapping of Origin of Replication in Themococcales. *Bioinformation* **2010**, *5*, 213–218.

212. Makarova, K. S.; Koonin, E. V. Archaeology of Eukaryotic DNA Replication. *Cold Spring Harbor Perspectives in Biology* **2013**, *5*, a012963–a012963.

213. Akita, M.; Adachi, A.; Takemura, K.; Yamagami, T.; Matsunaga, F.; Ishino, Y. Cdc6/Orc1 from Pyrococcus Furiosus May Act as the Origin Recognition Protein and Mcm Helicase Recruiter. *Genes Cells* **2010**, *15*, 537–552.

214. Norais, C.; Hawkins, M.; Hartman, A. L.; Eisen, J. A.; Myllykallio, H.; Allers, T. Genetic

and Physical Mapping of DNA Replication Origins in Haloferax Volcanii. *PLoS genetics* **2007**, *3*, e77.

215. Samson, R. Y.; Xu, Y.; Gadelha, C.; Stone, T. A.; Faqiri, J. N.; Li, D.; Qin, N.; Pu, F.; Liang, Y. X.; She, Q.; Bell, S. D. Specificity and Function of Archaeal DNA Replication Initiator Proteins. *Cell Rep* **2013**, *3*, 485–496.
216. Raymann, K.; Forterre, P.; Brochier-Armanet, C.; Gribaldo, S. Global Phylogenomic Analysis Disentangles the Complex Evolutionary History of DNA Replication in Archaea. *Genome biology and evolution* **2014**, *6*, 192–212.
217. Wu, Z.; Liu, J.; Yang, H.; Liu, H.; Xiang, H. Multiple Replication Origins with Diverse Control Mechanisms in Haloarcula Hispanica. *Nucleic Acids Res* **2014**, *42*, 2282–2294.
218. Dueber, E. C.; Costa, A.; Corn, J. E.; Bell, S. D.; Berger, J. M. Molecular Determinants of Origin Discrimination by Orc1 Initiators in Archaea. *Nucleic Acids Res* **2011**, *39*, 3621–3631.
219. Dueber, E. L. C.; Corn, J. E.; Bell, S. D.; Berger, J. M. Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex. *Science* **2007**, *317*, 1210–1213.
220. Gaudier, M.; Schuwirth, B. S.; Westcott, S. L.; Wigley, D. B. Structural Basis of DNA Replication Origin Recognition by an ORC Protein. *Science* **2007**, *317*.
221. Grainge, I.; Gaudier, M.; Schuwirth, B. S.; Westcott, S. L.; Sandall, J.; Atanassova, N.; Wigley, D. B. Biochemical Analysis of a DNA Replication Origin in the Archaeon Aeropyrum Pernix. *Journal of Molecular Biology* **2006**, *363*, 355–369.
222. 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* **2007**, *282*, 4908–4915.
223. Samson, R. Y.; Abeyrathne, P. D.; Bell, S. D. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Molecular Cell* **2016**, *61*, 287–296.

224. Matsunaga, F.; Forterre, P.; Ishino, Y.; Myllykallio, H. In Vivo Interactions of Archaeal Cdc6/Orc1 and Minichromosome Maintenance Proteins with the Replication Origin. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 11152–7.

225. Matsunaga, F.; Takemura, K.; Akita, M.; Adachi, A.; Yamagami, T.; Ishino, Y. Localized Melting of Duplex DNA by Cdc6/Orc1 at the DNA Replication Origin in the Hyperthermophilic Archaeon Pyrococcus Furiosus. *Extremophiles* **2010**, *14*, 21–31.

226. Sakakibara, N.; Kelman, L. M.; Kelman, Z. How Is the Archaeal MCM Helicase Assembled at the Origin? Possible Mechanisms. *Biochemical Society transactions* **2009**, *37*, 7–11.

227. Sakakibara, N.; Kelman, L. M.; Kelman, Z. Unwinding the Structure and Function of the Archaeal MCM Helicase. *Molecular microbiology* **2009**, *72*, 286–96.

228. Lindås, A.-C.; Bernander, R. The Cell Cycle of Archaea. *Nature Reviews Microbiology* **2013**, *11*, 627–638.

229. Bernander, R. Chromosome Replication, Nucleoid Segregation and Cell Division in Archaea. *Trends in microbiology* **2000**, *8*, 278–83.

230. Cubonova, L.; Richardson, T.; Burkhart, B. W.; Kelman, Z.; Connolly, B. A.; Reeve, J. N.; Santangelo, T. J. Archaeal DNA Polymerase D but Not DNA Polymerase B Is Required for Genome Replication in Thermococcus Kodakarensis. *J Bacteriol* **2013**, *195*, 2322–2328.

231. Matsunaga, F.; Glatigny, A.; Mucchielli-Giorgi, M.-H.; Agier, N.; Delacroix, H.; Marisa, L.; Durosay, P.; Ishino, Y.; Aggerbeck, L.; Forterre, P. Genomewide and Biochemical Analyses of DNA-Binding Activity of Cdc6/Orc1 and Mcm Proteins in Pyrococcus Sp. *Nucleic Acids Research* **2007**, *35*, 3214–3222.

232. Kogoma, T. Stable DNA Replication: Interplay between DNA Replication, Homologous Recombination, and Transcription. *Microbiology and molecular biology reviews : MMBR*

1997, *61*, 212–38.

233. Maduike, N. Z.; Tehranchi, A. K.; Wang, J. D.; Kreuzer, K. N. Replication of the Escherichia Coli Chromosome in RNase HI-Deficient Cells: Multiple Initiation Regions and Fork Dynamics. *Molecular microbiology* **2014**, *91*, 39–56.

234. Masai, H.; Asai, T.; Kubota, Y.; Arai, K.; Kogoma, T. Escherichia Coli PriA Protein Is Essential for Inducible and Constitutive Stable DNA Replication. *The EMBO journal* **1994**, *13*, 5338–45.

235. Ogawa, T.; Pickett, G. G.; Kogoma, T.; Kornberg, A. RNase H Confers Specificity in the dnaA-Dependent Initiation of Replication at the Unique Origin of the Escherichia Coli Chromosome in Vivo and in Vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 1040–4.

236. Stuckey, R.; García-Rodríguez, N.; Aguilera, A.; Wellinger, R. E. Role for RNA:DNA Hybrids in Origin-Independent Replication Priming in a Eukaryotic System. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 5779–84.

237. Hawkins, M.; Malla, S.; Blythe, M. J.; Nieduszynski, C. A.; Allers, T. Accelerated Growth in the Absence of DNA Replication Origins. *Nature* **2013**, *503*, 544–547.

CHAPTER 2

ARCHAEAL RNA POLYMERASE ARRESTS TRANSCRIPTION AT DNA LESIONS[2]

2.1 INTRODUCTION

All life is dependent on multi-subunit RNA polymerases (RNAP) that share a double β-barrel core structure[1–9]. Despite a highly conserved active center and hybrid binding channel, the overall fidelity, elongation rate, processivity and response to sites of DNA damage differs among RNAPs from each Domain (Table 2.1). Bacterial RNAP, eukaryotic Pol II, and phage RNAPs can respond to a variety of distinct DNA lesions, although the response elicited (e.g. lesion bypass, pausing, backtracking, arrest) varies dependent on the source of RNAPs[10–24].

Archaeal-encoded RNAPs share significant structural homology with eukaryotic RNAP II (Pol II) and are also reliant on conserved archaeal-eukaryotic factors for promoter-directed initiation[8,9,25,26]. Many model archaeal species thrive in harsh environments that likely result in continuous damage to the genomes, and this damage likely impedes both replication and transcription[27–33]. Here we use RNAP purified from *Thermococcus kodakarensis*, a marine, hyperthermophilic, anaerobic euryarchaeaon to monitor the response of the archaeal RNAP to site- and strand-specific DNA lesions *in vitro*. The archaeal RNAP arrests transcription at a variety of template strand lesions whereas non-template strand lesions do not hinder transcription. Depending on the type of lesion, archaeal RNAP often stalls adjacent to the lesion, demonstrating that incorporation across from the damaged base is often possible but

---

[2] This chapter was published in June 2017 under the same title.

TJS and I conceived the content. I wrote the original manuscript with input from TJS. TJS and I designed all the experiments. I performed all experiments and analyzed all data.

Gehring, A.M.; Santangelo, T.J. Archaeal RNA Polymerase Arrests Transcript at DNA Lesions. *Transcription* **2017**. PMID: 28598254.

Table 2.1. Transcription arrest at DNA lesions *in vitro*

	Abasic	8oG	8oA	TG	dU	Ref.
Phage	-/+	-/+	+	+	-	10-15
Bacteria	-	-	n.d.	n.d.	-	16-19
Archaea	+	+	+	+	-/+	This study
Eukarya (Pol II)	-/+	-/+	+	-/+	-	11,15,20-24

that continued elongation is severely hindered as the misaligned base pair moves through the hybrid binding channel. Our results are supportive of the recent demonstration of transcription-coupled DNA repair (TCR) in some Archaea, a pathway dependent the strand-specific recognition of DNA lesions by RNAP[34].

2.2 RESULTS

Archaeal RNAP is sensitive to lesions in the template strand of DNA

All multi-subunit RNAPs normally initiate transcription *de novo* and promoter-initiated archaeal transcription minimally requires the addition of the basal transcription factors TBP (TATA-binding protein) and TFB (transcription factor B)[25,35]. An alternative mechanism of transcription is possible *in vitro* that removes the requirements of additional factors and promoter sequences[36]. This system is reliant only on addition of RNAP and a nucleic acid scaffold composed of i) a short RNA that is partially complementary to the template strand of DNA and ii) complementary non-template and template strand DNAs (Figure 2.1). This scaffold supports RNAP binding, encapsulation of the preformed RNA/DNA hybrid into the hybrid-binding site within the main channel of RNAP, and alignment the RNA transcript 3'-OH for continued catalysis upon addition of NTPs to the reaction. Scaffold-assembled transcription systems offer several advantages over promoter-directed transcription for study of transcription elongation. First, the reaction is dependent only on RNAP and the assembled nucleic acid scaffold, thus removing any influences from transcription factors that may stay associated with RNAP during the early stages of elongation. Second, elongation with a subset of NTPs is simplified, permitting RNAP elongation to any predefined template position by withholding one or more NTP substrates. Finally, and most importantly for these assays, the scaffold can be assembled from purified single-strands of nucleic acids. Chemical synthesis of each deoxyoligonucleotide permits incorporation of site- and strand-specific modified bases. The DNA templates used in this study thus allowed the response of the archaeal RNAP to a DNA lesion in either the template or the non-template strand of DNA to be investigated in isolation (Figure 2.1).

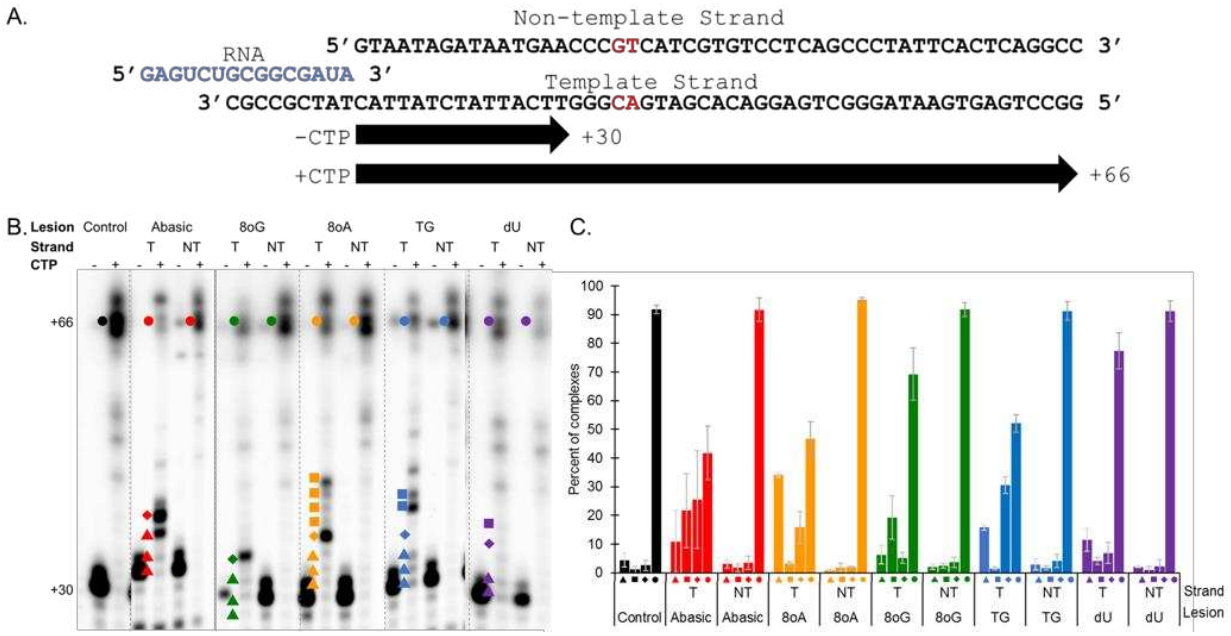

Figure 2.1. RNAP polymerase selectively recognizes and arrests at template strand DNA lesions. (A) The scaffolds were constructed from individual nucleic acid deoxyoligonucleotides (15 nt RNA, 60 nt template strand with 9bp complementary to RNA, 51 nt non-template strand). Modified bases were incorporated into either the template or non-template strand at either +34 or +35 (red nucleotides). The position of the elongation complex is designated by the total length of the RNA. The initial +15 complex will resume synthesis upon NTP addition to +30 (-CTP) or +66. (B) Transcription on scaffold templates with the noted lesion incorporated in either the template (T) or non-template (NT) strand. Reactions on control templates demonstrate the ability to limit elongation to +30 (-CTP) or allow full-length transcript synthesis (+66). Lesions in the NT strand do not inhibit transcription elongation and permit production of full-length transcripts (●). Lesions in the T strand result in significant and lesion-specific pausing before (▲), at (♦), and post (■) DNA lesion. (C) Quantification of the total lengths of RNA transcripts after 3 minutes of elongation on scaffolds containing site- and strand-specific DNA lesions.

Transcription elongation from a scaffold wherein the non-template (NT) and template (T) strands of DNA contained only the four natural bases (Figure 2.1B; Control) resulted in the expected elongation to +30 when reactions were supplemented with only ATP, GTP, and UTP; note that some elongation to +31 is seen for each template and this additional incorporation is likely the result of both small amounts of contaminating CTP and NTP misincorporation during the 3 minute incubation at 85˚C. In contrast, reactions supplemented with all 4 NTPs resulted in >90% of the initial +15 nucleotide (nt) RNA elongated to a full-length +66 nt transcript (Figure 2.1B, lanes 1 and 2).

Common DNA lesions, including deoxyuridine (dU), an abasic site, 8-oxoguanine (8oG), 8-oxyadenine (8oA), and thymidine glycol (TG) were individually site- and strand-specifically incorporated into the deoxynucleotide sequences that supported scaffold-dependent transcription. The modified or missing base was incorporated at position +34 (dU, abasic, 8oG, TG) or +35 (8oA) in either the template or non-template strand. Elongation in the absence of CTP from all templates resulted in efficient elongation to +30/+31 indicating that each scaffold supported RNAP binding and template-directed transcription elongation (Figure 2.1B, odd lanes). Elongation to +30/+31 does not permit RNAP to engage the modified or missing base within its active center. Furthermore, elongation in the presence of all 4 NTPs from scaffolds where the modified or missing base was incorporated into the non-template strand resulted in >90% of +15 initial transcripts being elongated to full-length +66 transcripts without evidence of pausing, arrest, or delay along the length of DNA template compared to the control template.

Template-strand specific incorporation of an abasic position, 8oA, 8oG, or TG resulted in the majority of archaeal RNAPs arresting at or adjacent to the position of the modified or missing base (Figure 2.1). The exact response of the archaeal RNAP to each specific lesion was unique, and with the exception of dU, even after 3 minutes of elongation with 200μM NTPs at the optimal temperature of 85˚C, nearly half of all RNAPs failed to generate full-length transcripts (Figure 2.1C).

An abasic site at position +34 of the template strand resulted, after 3 minutes of elongation, primarily in transcripts that were +33 nts and +34 nts. These lengths indicate that the archaeal RNAP continued synthesis until the abasic position resided in the active center of RNAP (+33) and that RNAP could often incorporate a base across from the abasic template position (+34). Elongation beyond the abasic site was not seen in most cases, arguing that passage of the abasic site through the hybrid-binding pocket of the archaeal RNAP was not tolerated. A similar pattern of elongation and arrest is seen on scaffolds containing 8oG at position +34, although elongation across from 8oG is better supported than incorporation across from an abasic site. On both templates a modest but reproducible percentage of RNAPs fail to elongate a few nts before the lesion. It is possible that transcripts at +30-32 nts result from backtracking and endonucleolytic cleavage of the RNA that was initially elongated to position +33 or +34, or that RNAP failed to elongate efficiently on templates with downstream lesions.

Elongation on scaffolds with 8oA incorporated into the template strand at +35 resulted in a mixture of arrested elongation complexes with RNA lengths varying from +30-38 nts. The dominant position is +34, suggesting the most elongation complexes were not capable of incorporation across from 8oA and that limited synthesis occurred after such incorporation. Specific +36, +37, and +38 nt transcripts were noted, suggesting that continued elongation was severely hindered by an altered RNA/DNA hybrid configuration containing the modified template DNA base. Evidence of hindered elongation is present for only 3 nts after the lesion, suggesting that once the modified template position moved far enough through the RNA/DNA hybrid that any distortion of the hybrid was tolerated to permit more rapid catalysis. Scaffolds containing a thymidine glycol (TG) derivative at +34 resulted in minimal interference with incorporation across from the DNA lesion but rather resulted in apparent arrest of the elongation complex several base pairs downstream of the lesion. Multiple positions of arrest downstream of TG suggest that like 8oA, movement of TG through the RNA/DNA hybrid was not well tolerated within the elongation complex. 8oA- and TG-containing scaffolds also result in production of

shorter transcripts that suggest that RNAP either failed to properly elongate due to the presence of a downstream lesion or that these transcripts result from synthesis, backtracking and cleavage of RNAs that were not elongated to or beyond the modified base. Elongation on scaffolds containing dU incorporated at +34 of the template strand resulted in minimal defects to elongation after 3 minutes. Approximately 10% of complexes appear to arrest before or at the site of the lesion, but the effects of dU are more modest than the elongation defects resultant from the other lesions tested here.

Prolonged pausing and arrest at template-strand DNA lesions

All multi-subunit RNAPs appear capable of forward and retrograde translocation during elongation. Given the apparent varied response of the archaeal RNAP to each lesion it was of interest to determine if the lengths of RNAs resultant after 3 minutes of elongation were the result of RNAP arrest, pausing, hindered elongation, backtracking and cleavage, etc. or some combination thereof. Previous *in vitro* studies[37] are in agreement with elongation kinetics measured here that show the archaeal RNAP typically elongates at ~20-25 nts/sec, permitting elongation of +15 RNAs to full-length +66 transcripts in approximately 3 seconds. To better understand how the archaeal RNAP responds to each lesion, we performed *in vitro* transcription on the same templates and removed aliquots over time to monitor the elongation kinetics (Figure 2.2A). For the control scaffold, as well as for all scaffolds containing lesions in the non-template strand, elongation was unhindered and resulted in nearly all full-length transcripts in just 2-5 seconds. In contrast, scaffolds containing template-strand lesions resulted in significant delays and overall reductions in production of full-length transcripts. Nearly all elongation complexes on these templates were affected by the modified template position (Figure 2.3B), and ~70-80% of all complexes demonstrated a prolonged pause at the exact position of the DNA lesion on each scaffold (Figure 2.3A). Each scaffold containing a template-strand DNA lesion did permit some elongation complexes to generate full-length transcripts, with TG, 8oA, and an abasic position limiting elongation of greater than half of all complexes (Figure 2.2B). dU

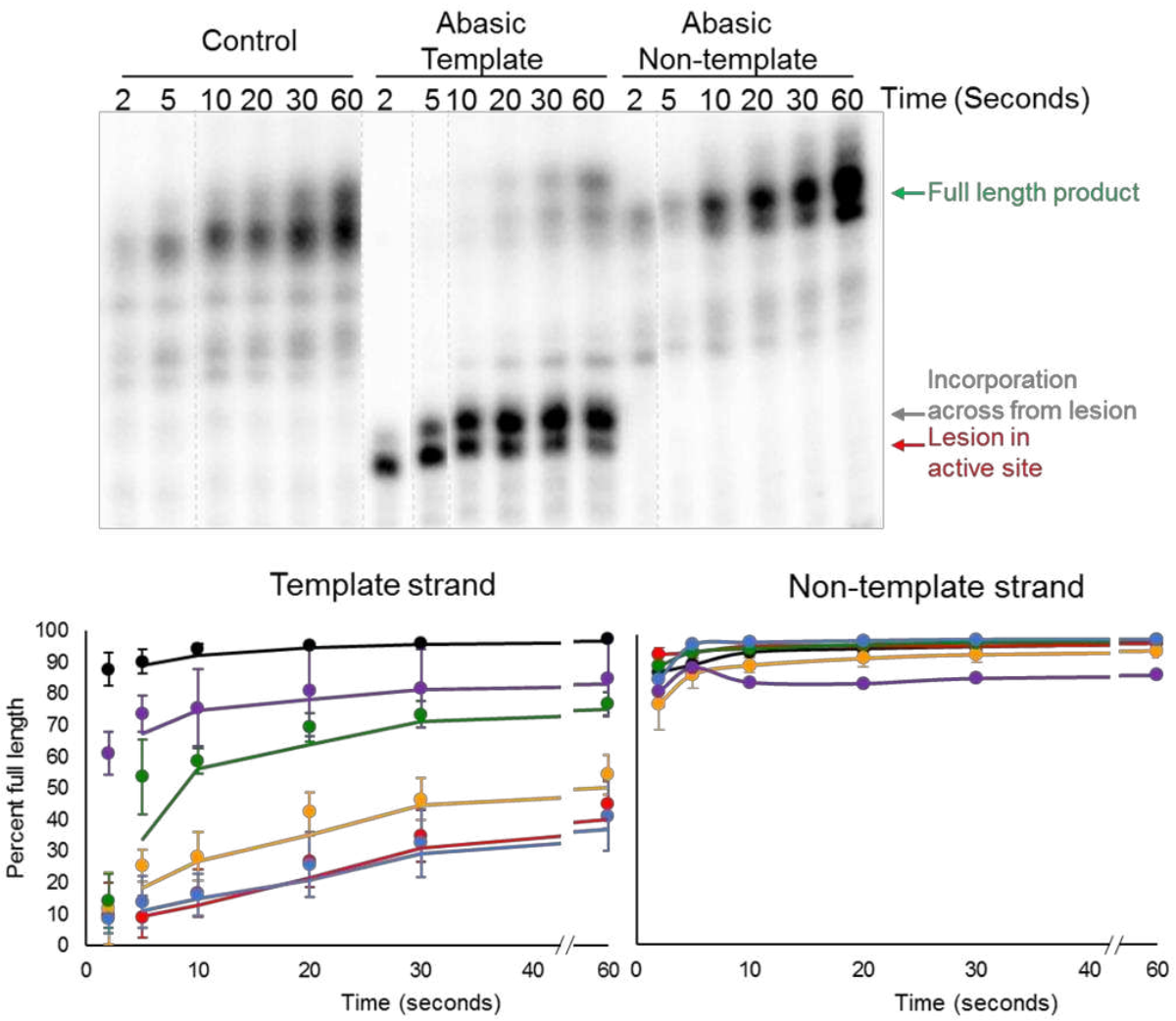

Figure 2.2. Template strand lesions arrest transcription elongation. (A) Monitoring the rate of *in vitro* transcription on scaffold templates containing a template strand-specific lesion reveals that elongation is generally unhindered until the lesion enters the RNAP active site and arrests elongation. Identically modified bases in the non-template strand do not block elongation. (B) The time required to generate full length transcripts is increased when the DNA lesion, abasic (red), 8oA (gold), 8oG (green), TG (blue), or dU (purple), is present in the template strand.

and 8oG permitted most complexes to eventually yield full-length transcripts, however, in all cases the time required for full-length transcript production was dominated by pausing at or near the site of the lesion (Figures 2.2B and 2.3). For each scaffold with template-strand DNA lesions, we observed some complexes that remained indefinitely stalled at the DNA lesion.

We calculated the half-life of complexes stalled at the lesion (Figure 2.3A) and the collective half-life of all complexes stalled at or adjacent to the lesion (Figure 2.3B). The half-life of RNAP stalled at the DNA lesion ranged from approximately 4-7 seconds for all lesions except dU (Figure 2.3). For 8oG the pause at the lesion was responsible for the bulk of the time delay in generating full-length transcripts, but for TG, 8oA, dU and an abasic position, the half-life of the pauses at adjacent positions dominated the delay in full-length transcript synthesis (Figure 2.3A and 2.3B). Except for a slow incorporation across from the abasic site resulting in a slow shift in the percentage of complexes stalled before versus at the template strand lesion, the distribution of complexes stalled at or adjacent to template strand DNA lesions was largely invariant over time (Figure 2.3E). This invariance suggests that most complexes are immediately arrested rather than paused at or near the template-strand DNA lesions. We failed to observe any obvious retrograde movement or cleavage of transcripts over time on any template.

2.3 DISCUSSION

Failure of the archaeal RNAP to respond to non-template strand DNA lesions is consistent with the activities of multi-subunit RNAPs from Bacteria and Eukarya and with the predicted path of the nucleic acids through RNAP[38–41]. The non-template strand is not predicted to make any contacts with the active center of RNAP and thus lesions within the non-template strand are not predicted to influence catalysis.

Each template-strand lesion resulted in the majority of elongation complexes pausing or arresting at or near the lesion, and with the exception of scaffolds containing dU, the bulk of elongation complexes failed to elongate significantly beyond the position of the lesion. We

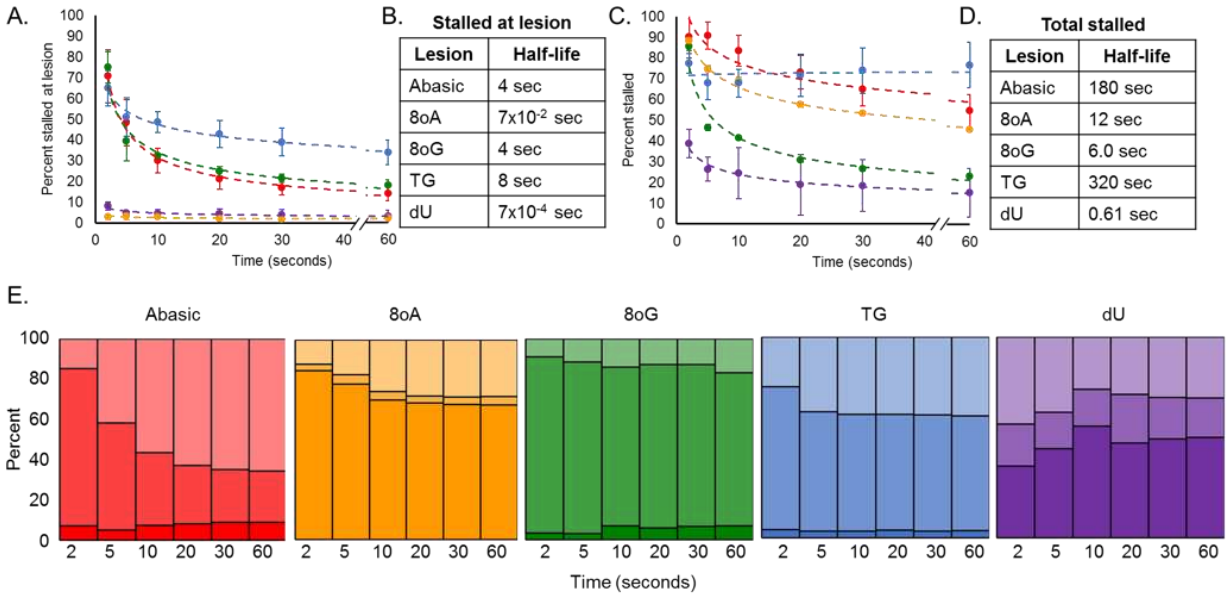

Figure 2.3. Transcription arrest occurs at and after the template strand incorporated lesion. (A) RNAP arrests at DNA lesions, abasic (red), 8oA (gold), 8oG (green), TG (blue), or dU (purple), in the template strand demonstrating that the DNA lesions are a significant block to RNAP elongation. (B) Quantification of the half-life of RNAP arrested at a DNA lesion. (C) Percent of elongation complexes arrested before, at, or after the DNA lesions. (D) Quantification of half-life of elongation complexes arrested prior to, at, or downstream of the DNA lesions. (E) Percent of complexes stalled before the DNA lesion (dark shading), at the lesions (medium shading), and after the DNA lesion (light shading).

demonstrate that the archaeal RNAP is capable of elongation up to each modified base, although elongation up to dU, 8oA, 8oG, TG, and abasic sites is hindered even before the modified base enters the bipartite RNAP active center (Figure 2.2A). Incorporation across from each base is possible, with obvious difficulties observed for dU, TG, and 8oG, whereas incorporation across from an abasic site or 8oA was extremely slow. Elongation after incorporation at the modified template position was affected for all templates, with movement of the modified base through the RNA/DNA hybrid limiting elongation rates at several positions downstream.

Elongation on the control scaffold, as well as for all scaffolds containing lesions in the non-template strand, was unhindered and resulted in the expected completion of nearly all full-length transcripts in just ~2-5 seconds (Figure 2.2B). In contrast, scaffolds containing template-strand lesions resulted in significant delays and overall reductions in production of full-length transcripts. Nearly all elongation complexes on these templates were affected by the modified template position (Figure 2.3A), and ~70-80% of all complexes demonstrated a transcriptional arrest at or adjacent to the position of the DNA lesion. For each scaffold with template-strand DNA lesions, we observed some complexes that remained indefinitely stalled at the DNA lesion. We define complexes that fail to generate full length transcripts and dwell at or near the position of the lesion for > 4 seconds on average, representing a ~100-fold greater residence time over the average dwell time of just ~40 milliseconds, as arrested. ~40% of complexes suffer only a long pause and can slowly bypass the lesions on templates with TG, 8oA, or an abasic position (Figure 2.2B).

The archaeal RNAP appears to recognize and halt elongation in response to a greater number of template-strand lesions than RNAPs from bacteriophages, Bacteria, and Eukarya (Table 2.1)[10–24]. The increased sensitivity of the archaeal RNAP to a range of DNA lesions may reflect a relatively rigid active center configuration that is not tolerant of deviations from idealized geometry for synthesis. Although a remarkable degree of structural conservation among the

core structure of the multi-subunit RNAPs exists[8,9,26,42], subtle differences in active site geometries could readily account for the differential sensitivity of the RNAPs from different Domains to template-strand lesions. The ability of the archaeal RNAP to recognize and arrest transcription at a variety of common DNA lesions is supportive of a role for RNAP in initiating the recently described more rapid repair of template versus non-template DNA lesions *in vivo*[34]. We demonstrate that transcription elongation by the archaeal RNAP is arrested at a variety of template-strand DNA lesions *in vitro*. Identical non-template strand lesions do not result in any notable reduction in elongation kinetics. The strand-bias and efficiency of recognition of a variety of template-strand DNA lesions is supportive of the archaeal RNAP serving as a sensor of DNA damage *in vivo* that may serve to signal TCR mediated DNA repair pathways in archaeal species. Repair of the template-strand lesions that arrest RNAP likely requires either the removal of RNAP from the DNA template to expose the lesion, or the repositioning of RNAP on the template in a significantly backtracked position to expose the DNA lesion. Such activities would be consistent with TCR pathways in Bacteria and Eukarya[43–45], respectively, although no factor capable of directing termination or dissociation of the archaeal elongation complex has yet been described. Furthermore, any archaeal encoded factors responsible for initiating template-strand DNA repair following exposure of the DNA lesion are not yet known. How the arrested RNAP is recognized, removed or repositioned, and the factors responsible for initiating DNA repair will require additional studies to elucidate the mechanisms and mechanics of archaeal transcription-coupled DNA repair.

2.4 MATERIALS AND METHODS

DNA templates

DNA templates were assembled from complementary, individual HPLC-purified oligodeoxynucleotides from TriLink Biotechnologies or Eurofins MWG Operon (Figure 2.1). (abasic, dSpacer; 8oA, 8-Oxo-2-deoxyadenosine; 8oG, 8-Oxo-2-deoxyguanosine; TG, thymidine glycol; dU, deoxyuridine) Lesion-containing oligodeoxynucleotides were paired with

complementary unmodified oligodeoxynucleotides in all cases. Selective digestions of such pairings with enzymes with specificity for each lesion (T4 PDG, Fpg, and EndoV) revealed that >90% of such double-stranded DNAs contained recognizable mispairings. Similar results were obtained from assembled transcription reactions, confirming incorporation of both strands of DNA into the scaffold-assembled transcription elongation complexes.

Purification of RNA polymerase

RNAP was purified as previously described[35,46]. Briefly, RNAP was purified from *T. kodakarensis* strain TS413 by a combination of Ni^{2+} affinity and MonoQ chromatography.

In vitro transcription assays

In vitro transcription reactions were assembled using a scaffold DNA template (Figure 2.1)[36]. Equimolar amounts (20 nM) of the template strand of DNA and [^{32}P]-labeled RNA were combined, heated to 65°C and slowly cooled to room temperature. RNAP (40 nM) was added and the reaction adjusted to 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 5 mM DTT. After incubation at room temperature for 10 minutes, the non-template strand was then added in 3x molar excess (60 nM) and incubation was extended for an additional 10 minutes at room temperature. Complexes were then heated to 85°C for 1 minute and rNTPs (each NTP at a final concentration of 200 µM) were added. A DNA oligonucleotide (5′ TATCGCCGCAGACTC 3′) complementary to the RNA was added at a final concentration of 10nM to prevent additional rounds of transcription initiation[47]. Transcription elongation was permitted for the experiment specific amount of time and was then stopped by addition of 100µl 1.2x stop buffer (0.6 M Tris-HCl pH 8.0, 12 mM EDTA). Reactions were subjected to equal volume phenol:chloroform:isoamyl alcohol (25:24:1) extractions and the [^{32}P]-labeled RNA transcripts were purified from the aqueous phase by alcohol-precipitation. Purified transcripts were resolved through 15% denaturing 1X-TBE polyacrylamide gels containing 8M urea (National Diagnostic Urea Gel)[35]. Radiolabeled RNA was detected on phosphorimaging plates and quantified using a Typhoon FLA 9500 (GE Healthcare). Images were analyzed using GE

Imagequant Software 5.2. The percentage of full length transcripts and transcripts within paused/arrested complexes on each template was calculated as the percentage of all transcripts +30nt or greater. Error bars represent the standard error of the mean from at least three independent assays.

REFERENCES

1. Gnatt, A. L. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. *Science* **2001**, *292*, 1876–1882.

2. Cramer, P. Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science* **2001**, *292*, 1863–1876.

3. Jasiak, A. J.; Armache, K.-J.; Martens, B.; Jansen, R.-P.; Cramer, P. Structural Biology of RNA Polymerase III: Subcomplex C17/25 X-Ray Structure and 11 Subunit Enzyme Model. *Molecular Cell* **2006**, *23*, 71–81.

4. Armache, K.-J.; Kettenberger, H.; Cramer, P. Architecture of Initiation-Competent 12-Subunit RNA Polymerase II. *Proceedings of the National Academy of Sciences* **2003**, *100*, 6964–6968.

5. Cramer, P.; Armache, K.-J.; Baumli, S.; Benkert, S.; Brueckner, F.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jasiak, A. J.; Jawhari, A.; Jennebach, S.; Kamenski, T.; Kettenberger, H.; Kuhn, C.-D.; Lehmann, E.; Leike, K.; Sydow, J. F.; Vannini, A. Structure of Eukaryotic RNA Polymerases. *Annual Review of Biophysics* **2008**, *37*, 337–352.

6. Vassylyev, D. G.; Sekine, S.; Laptenko, O.; Lee, J.; Vassylyeva, M. N.; Borukhov, S.; Yokoyama, S. Crystal Structure of a Bacterial RNA Polymerase Holoenzyme at 2.6 Å Resolution. *Nature* **2002**, *417*, 712–719.

7. Zhang, G.; Campbell, E. A.; Minakhin, L.; Richter, C.; Severinov, K.; Darst, S. A. Crystal Structure of Thermus Aquaticus Core RNA Polymerase at 3.3 A Resolution. *Cell* **1999**, *98*, 811–24.

8. Hirata, A.; Klein, B. J.; Murakami, K. S. The X-Ray Crystal Structure of RNA Polymerase

from Archaea. *Nature* **2008**, *451*, 851–854.

9. Jun, S.-H. H.; Hirata, A.; Kanai, T.; Santangelo, T. J.; Imanaka, T.; Murakami, K. S. The X-Ray Crystal Structure of the Euryarchaeal RNA Polymerase in an Open-Clamp Configuration. *Nat Commun* **2014**, *5*, 5132.

10. Zhou, W.; Doetsch, P. W. Effects of Abasic Sites and DNA Single-Strand Breaks on Prokaryotic RNA Polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 6601–5.

11. Tornaletti, S.; Maeda, L. S.; Hanawalt, P. C. Transcription Arrest at an Abasic Site in the Transcribed Strand of Template DNA. *Chemical Research in Toxicology* **2006**, *19*, 1215–1220.

12. Chen, Y. H.; Bogenhagen, D. F. Effects of DNA Lesions on Transcription Elongation by T7 RNA Polymerase. *The Journal of biological chemistry* **1993**, *268*, 5849–55.

13. Tornaletti, S.; Maeda, L. S.; Kolodner, R. D.; Hanawalt, P. C. Effect of 8-Oxoguanine on Transcription Elongation by T7 RNA Polymerase and Mammalian RNA Polymerase II. *DNA repair* **2004**, *3*, 483–94.

14. Hatahet, Z.; Purmal, A. A.; Wallace, S. S. Oxidative DNA Lesions as Blocks to in Vitro Transcription by Phage T7 RNA Polymerase. *Annals of the New York Academy of Sciences* **1994**, *726*, 346–8.

15. Tornaletti, S.; Maeda, L. S.; Lloyd, D. R.; Reines, D.; Hanawalt, P. C. Effect of Thymine Glycol on Transcription Elongation by T7 RNA Polymerase and Mammalian RNA Polymerase II. *The Journal of biological chemistry* **2001**, *276*, 45367–71.

16. Clauson, C. L.; Oestreich, K. J.; Austin, J. W.; Doetsch, P. W. Abasic Sites and Strand Breaks in DNA Cause Transcriptional Mutagenesis in Escherichia Coli. *Proceedings of*

the National Academy of Sciences of the United States of America **2010**, *107*, 3657–62.

17. Viswanathan, A.; Doetsch, P. W. Effects of Nonbulky DNA Base Damages on Escherichia Coli RNA Polymerase-Mediated Elongation and Promoter Clearance. *Journal of Biological Chemistry* **1998**, *273*, 21276–21281.

18. Brégeon, D.; Doddridge, Z. A.; You, H. J.; Weiss, B.; Doetsch, P. W. Transcriptional Mutagenesis Induced by Uracil and 8-Oxoguanine in Escherichia Coli. *Molecular cell* **2003**, *12*, 959–70.

19. Smith, A. J.; Savery, N. J. Effects of the Bacterial Transcription-Repair Coupling Factor during Transcription of DNA Containing Non-Bulky Lesions. *DNA repair* **2008**, *7*, 1670–9.

20. Kuraoka, I.; Suzuki, K.; Ito, S.; Hayashida, M.; Kwei, J. S. M.; Ikegami, T.; Handa, H.; Nakabeppu, Y.; Tanaka, K. RNA Polymerase II Bypasses 8-Oxoguanine in the Presence of Transcription Elongation Factor TFIIS. *DNA repair* **2007**, *6*, 841–51.

21. Larsen, E.; Kwon, K.; Coin, F.; Egly, J.-M. M.; Klungland, A. Transcription Activities at 8-oxoG Lesions in DNA. *DNA Repair (Amst)* **2004**, *3*, 1457–1468.

22. Thorslund, T.; Sunesen, M.; Bohr, V. A.; Stevnsner, T. Repair of 8-oxoG Is Slower in Endogenous Nuclear Genes than in Mitochondrial DNA and Is without Strand Bias. *DNA repair* **2002**, *1*, 261–73.

23. Kathe, S. D.; Shen, G.-P.; Wallace, S. S. Single-Stranded Breaks in DNA but Not Oxidative DNA Base Damages Block Transcriptional Elongation by RNA Polymerase II in HeLa Cell Nuclear Extracts. *The Journal of biological chemistry* **2004**, *279*, 18511–20.

24. Kuraoka, I.; Endou, M.; Yamaguchi, Y.; Wada, T.; Handa, H.; Tanaka, K. Effects of Endogenous DNA Base Lesions on Transcription Elongation by Mammalian RNA Polymerase II. Implications for Transcription-Coupled DNA Repair and Transcriptional

Mutagenesis. *The Journal of biological chemistry* **2003**, *278*, 7294–9.

25. Gehring, A. M.; Walker, J. E.; Santangelo, T. J. Transcription Regulation in Archaea. *Journal of bacteriology* **2016**.

26. Grohmann, D.; Hirtreiter, A.; Werner, F. Molecular Mechanisms of Archaeal RNA Polymerase. *Biochem Soc Trans* **2009**, *37*, 12–17.

27. Kelman, Z.; White, M. F. Archaeal DNA Replication and Repair. *Curr Opin Microbiol* **2005**, *8*, 669–676.

28. Zivanovic, Y.; Armengaud, J.; Lagorce, A.; Leplat, C.; Guérin, P.; Dutertre, M.; Anthouard, V.; Forterre, P.; Wincker, P.; Confalonieri, F.; Chandler, M.; Barakat, M.; Dedieu, A.; Barbe, V.; Heulin, T.; Sommer, S.; Achouak, W.; Armengaud, J.; Song, J.; Thiessen, P.; Yamashita, R.; Yin, J.; Zhang, D.; Bryant, S.; Doolittle, W.; Duguet, M.; Gaasterland, T.; Garrett, R.; Ragan, M.; Sensen, C.; Oost, J. Van der Genome Analysis and Genome-Wide Proteomics of Thermococcus Gammatolerans, the Most Radioresistant Organism Known amongst the Archaea. *Genome Biology* **2009**, *10*, R70.

29. Grogan, D. W. Stability and Repair of DNA in Hyperthermophilic Archaea. *Curr Issues Mol Biol* **2004**, *6*, 137–144.

30. Hashimoto, H.; Inoue, T.; Nishioka, M.; Fujiwara, S.; Takagi, M.; Imanaka, T.; Kai, Y. Hyperthermostable Protein Structure Maintained by Intra and Inter-Helix Ion-Pairs in Archaeal O6-Methylguanine-DNA Methyltransferase. *Journal of Molecular Biology* **1999**, *292*, 707–716.

31. Constantinesco, F.; Forterre, P.; Koonin, E. V; Aravind, L.; Elie, C. A Bipolar DNA Helicase Gene, herA, Clusters with rad50, mre11 and nurA Genes in Thermophilic Archaea. *Nucleic Acids Research* **2004**, *32*, 1439–1447.

32. Wei, T.; Zhang, S.; Zhu, S.; Sheng, D.; Ni, J.; Shen, Y. Physical and Functional Interaction between Archaeal Single-Stranded DNA-Binding Protein and the 5′–3′ Nuclease NurA. *Biochemical and Biophysical Research Communications* **2008**, *367*, 523–529.

33. Makarova, K. S.; Aravind, L.; Grishin, N. V; Rogozin, I. B.; Koonin, E. V A DNA Repair System Specific for Thermophilic Archaea and Bacteria Predicted by Genomic Context Analysis. *Nucleic acids research* **2002**, *30*, 482–96.

34. Stantial, N.; Dumpe, J.; Pietrosimone, K.; Baltazar, F.; Crowley, D. J. Transcription-Coupled Repair of UV Damage in the Halophilic Archaea. *DNA repair* **2016**, *41*, 63–8.

35. Gehring, A. M.; Santangelo, T. J. Manipulating Archaeal Systems to Permit Analyses of Transcription Elongation-Termination Decisions in Vitro. *Methods in molecular biology (Clifton, N.J.)* **2015**, *1276*, 263–79.

36. Santangelo, T. J.; Reeve, J. N. Deletion of Switch 3 Results in an Archaeal RNA Polymerase That Is Defective in Transcript Elongation. *J Biol Chem* **2010**, *285*, 23908–23915.

37. Xie, Y.; Reeve, J. N. Transcription by an Archaeal RNA Polymerase Is Slowed but Not Blocked by an Archaeal Nucleosome. *J Bacteriol* **2004**, *186*, 3492–3498.

38. Kettenberger, H.; Armache, K.-J.; Cramer, P. Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Molecular Cell* **2004**, *16*, 955–965.

39. Barnes, C. O.; Calero, M.; Malik, I.; Graham, B. W.; Spahr, H.; Lin, G.; Cohen, A. E.; Brown, I. S.; Zhang, Q.; Pullara, F.; Trakselis, M. A.; Kaplan, C. D.; Calero, G. Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete

Transcription Bubble. *Molecular Cell* **2015**, *59*, 258–269.

40. Bae, B.; Feklistov, A.; Lass-Napiorkowska, A.; Landick, R.; Darst, S. A. Structure of a Bacterial RNA Polymerase Holoenzyme Open Promoter Complex. *eLife* **2015**, *4*.

41. Vassylyev, D. G.; Vassylyeva, M. N.; Perederina, A.; Tahirov, T. H.; Artsimovitch, I. Structural Basis for Transcription Elongation by Bacterial RNA Polymerase. *Nature* **2007**, *448*, 157–162.

42. Hirata, A.; Murakami, K. S. Archaeal RNA Polymerase. *Curr Opin Struct Biol* **2009**, *19*, 724–731.

43. 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* **2002**, *109*, 757–767.

44. Epshtein, V.; Kamarthapu, V.; McGary, K.; Svetlov, V.; Ueberheide, B.; Proshkin, S.; Mironov, A.; Nudler, E. UvrD Facilitates DNA Repair by Pulling RNA Polymerase Backwards. *Nature* **2014**, *505*, 372–377.

45. Spivak, G. Transcription-Coupled Repair: An Update. *Archives of toxicology* **2016**.

46. Santangelo, T. J.; Cubonova, L.; James, C. L.; Reeve, J. N.; 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* **2007**, *367*, 344–357.

47. Hein, P. P.; Kolb, K. E.; Windgassen, T.; Bellecourt, M. J.; Darst, S. A.; Mooney, R. A.; Landick, R. RNA Polymerase Pausing and Nascent-RNA Structure Formation Are Linked through Clamp-Domain Movement. *Nat Struct Mol Biol* **2014**, *21*, 794–802.

CHAPTER 3

GENOME REPLICATION IN *THERMOCOCCUS* KODAKARENSIS INDEPENDENT OF CDC6 AND AN ORIGIN OF REPLICATION[3]

3.1 INTRODUCTION

DNA replication is fundamental for cellular life and although there are differences in the details, the initiation of genome replication has common features in Bacteria, Archaea, and Eurkaryotes. An initiator protein or protein complex recognizes and assembles at one (all *Bacteria* and some *Archaea*) or multiple sites (some *Archaea* and all *Eukaryotes*) that function as origins of replication[1]. Under exceptional circumstances, initiator protein-independent genome replication, termed recombination-driven DNA replication initiation (RDR; also termed inducible and constitutive stable DNA replication) has been documented in *Bacteria,* but such mechanisms support - at best - only minimal cell growth[2–5]. It was surprising then when Hawkins *et al*[6] proposed that RDR not only facilitates genome replication but supports faster-than-wild type growth of a strain of the halophilic archaeon, *Haloferax volcanii* from which they had genetically deleted all four of the recognized origins of genome replication. Consistent with RDR initiation, the recombination factor RadA was essential for viability of the origin-less strain but could be deleted from the genome of the parental, origins-containing *H. volcanii*.

---

[3] This chapter is an accepted manuscript under the same title with the suggested reference below.

JNR, TJS, and ZK conceived the manuscript. RM, BWB, and I carried out the experiments. DPA and I analyzed the Marker Frequency data. TJS and I wrote the manuscript with input from JNR, ZK, and DPA.

Gehring, A. M.; Astling, D.P.; Matsumi, R.; Burkhart, B.W.; Kelman Z.; Reeve, J.N., Jones, K.L.; Santangelo, T. J. Genome replication in *Thermococcus kodakarensis* independent of Cdc6 and an origin of replication. *Frontiers in Microbiology*.

Given the established and convincingly large body of evidence that archaeal genomes have defined origins recognized and bound by initiator proteins[7–15], the proposal that RDR supports rapid growth in an archaeon[6] is unique and challenging. Most archaea encode replication initiator proteins that are homologous to eukaryotic initiation factors Orc1 and Cdc6, and one or more Cdc6-encoding genes are present in almost all sequenced archaeal genomes, usually located adjacent to a known or predicted origin(s) of replication[16–24]. An increase in the number of Cdc6 proteins is often positively correlated with the number of replication origins[20]; *H. volcanii* encodes fourteen Cdc6 proteins that function at three chromosomal origins and an integrated viral origin[9]. Some species are reliant on a single encoded, or only a single-functional Cdc6 protein to initiate replication, and the remaining Cdc6 isoforms are predicted to play roles in transcription regulation, recombination, replication restart or negative regulation of replication initiation[25].

To address the roles of Cdc6, presumptive origin sequences, and the potential of RDR to support rapid growth of archaeal strains, we took advantage of a procedure that permits the precise deletion of non-essential genome sequences and provides strong statistical evidence for essential genes in the hyperthermophilic archaeon *Thermococcus kodakarensis* [26]. Employing similar techniques, several essential and some surprisingly non-essential genes have already been identified revealing unanticipated features in archaeal DNA replication[27–31]. Bioinformatic analysis including GC-skew and Z-curve analysis predict only one origin of replication located directly adjacent to the gene encoding Cdc6 in *T. kodakarensis*[16,32,33]. Most members of the *Thermococcales*, including *T. kodakarensis*, encode only one identifiable Cdc6 protein[24]. We now report that Cdc6 and the adjacent previously-presumed origin of replication can be deleted from *T. kodakarensis* with no detectable consequences for viability, growth, genetic competence or plasmid maintenance. Data obtained by whole genome sequence and marker frequency analyses[34], coupled with the apparent essentiality of RadA and RadB, provide strong evidence

that *T. kodakarensis* normally employs a RDR mechanism for initiation that occurs at many sites around the genome.

3.2 RESULTS

Construction of T. kodakarensis Δcdc6

The procedure employed to delete genes from the *T. kodakarensis* genome permits a statistical definition of essentiality[26]. Plasmids are constructed and used to transform a parental strain (here *T. kodakarensis* TS559) so that the target locus is flanked by two sets of direct repeats. Spontaneous recombination in this intermediate strain between one set of the repeats results in the markerless deletion of the target locus, whereas an equally-probable recombination between the second set of repeats regenerates the parental strain. When only the parental strain is recovered, after screening >30 isolates generated from at least two independently-constructed intermediate strains, the target locus is defined operationally as essential for *T. kodakarensis* viability under our laboratory conditions.

TK1901-TK1902-TK1903 (encoding Cdc6, DNA polymerase D small and large subunits, respectively) form an operon[35] and essentiality has been previously established for TK1902 and TK1903 (Figure 3.1)[30]. Surprisingly, this was not true for TK1901, the only gene in *T. kodakarensis* that encodes a recognizable Cdc6 homologue. The design of the plasmid constructed to delete TK1901 ensured retention of the upstream promotor and so continued expression of TK1902-TK1903, and avoided deletion of any sequences in the adjacent ~900 bp region predicted to contain the origin of replication based on homology with the origin region in *Pyrococcus furiosus*, a related member of the *Thermococcales* (Figure 3.1)[16,36].

The presence of the TK1901 deletion was confirmed by diagnostic PCR and Southern blotting in two independent isolates (Figure 3.1). Amplicon sequencing confirmed that the 1,248 bp deletion extended precisely from the ATG-start codon to TGA-stop codon of TK1901, and this was subsequently re-confirmed by deep-sequencing (see below) of the entire genome of one isolate, designated *T. kodakarensis* Δ*cdc6*. This isolate was phenotypically

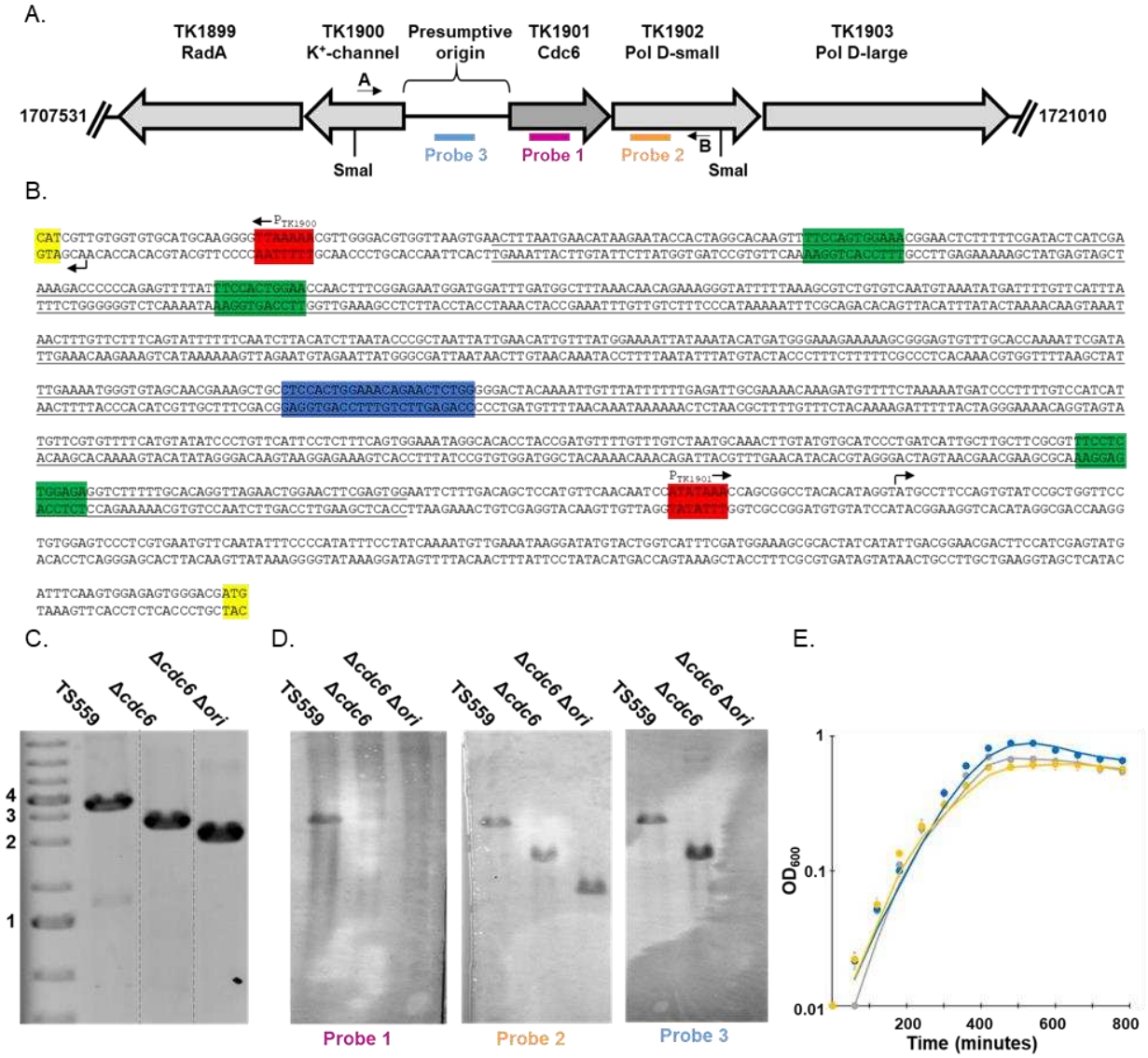

Figure 3.1. Deletion of TK1901 and the presumptive origin of replication from *T. kodakarensis* is non-phenotypic. (A) Organization of the *T. kodakarensis* genome surrounding TK1901. The locations of sequences used as primers in PCRs, probes in Southern blotting and SmaI recognition sites are shown. (B) Sequence of the presumptive origin region of *T. kodakarensis*. The full sequence of *cdc6* plus the underlined nucleotides were deleted from the genome in *T. kodakarensis* Δ*cdc6* Δ*ori*. (C) PCR generate amplicons confirm deletion of TK1901, as well as TK1901 and the presumptive origin from *T. kodakarensis* Δ*cdc6* and Δ*cdc6* Δ*ori*, respectively. (D) Southern blots of SmaI-digested genomic DNA from *T. kodakarensis* TS559, Δ*cdc6*, and Δ*cdc6 Δori* confirm deletion of *cdc6*, as well as *cdc6* and the presumptive origin, respectively. (E) Deletion of *cdc6* or *cdc6/ori* does not affect laboratory growth of *T. kodakarensis* TS559 (grey), Δ*cdc6* (blue), and Δ*cdc6* Δ*ori* (orange). Error bars report standard error of the mean of three biological replicates grown in triplicate.

indistinguishable from the parental *T. kodakarensis* TS559 strain: cultures grew at the same rate and reached the same final cell densities (Figure 3.1) and *T. kodakarensis* Δ*cdc6* was genetically competent and supported the autonomous replication of pTN1-based plasmids[37].

Marker frequency analysis of genomic DNA

Regions adjacent to an origin(s) are over-represented in growing cells, and marker frequency analyses comparing the number of sequencing reads across the genome has been used to identify replication origin(s) in many archaeal genomes[6,13,38]. Genomic DNA was therefore isolated, fragmented and deep-sequenced from growing and stationary phase cultures of *T. kodakarensis* TS559 and Δ*cdc6*. The sequences obtained confirmed the deletion of TK1901 but, in repeated experiments, all regions of the genome were equally represented in the DNA reads from both growing and stationary phase cells of both *T. kodakarensis* TS559 and Δ*cdc6* (Figure 3.2 and Figure 3.3). Given this unanticipated result, to provide confidence in the laboratory and computational procedures, the experiments were repeated with genomic DNAs from *Escherichia coli* MG1655 and *Pyrococcus furiosus* strain JFW02[39], species with established origins of replication. Quantification of the WGS reads clearly and correctly identified the origin loci established in the genomes of *E. coli* and *P. furiosus* (Figure 3.4)[39,40].

Construction of T. kodakarensis Δcdc6 Δori

Using the same markerless-genome modification techniques, the presumed origin sequences [32] were easily deleted in *T. kodakarensis* Δ*cdc6* (Figure 3.1). All the bioinformatically identified mini-origin recognition boxes (mini-ORBs; green) and one full ORB (blue) were deleted while retaining the promoters (red), transcription start sites (arrows), and translation start sites (yellow) for TK1902-1903, and for TK1900 (Figure 3.1). Both amplicon- and whole genome deep-sequencing confirmed the precision of the deletion. A representative isolate, designated *T. kodakarensis* Δ*cdc6* Δ*ori*, was phenotypically indistinguishable from *T. kodakarensis* TS559 and Δ*cdc6*. All three strains grew at the same rate, achieved the same final cell densities, were genetically competent and supported plasmid replication. Quantification of WGS reads also

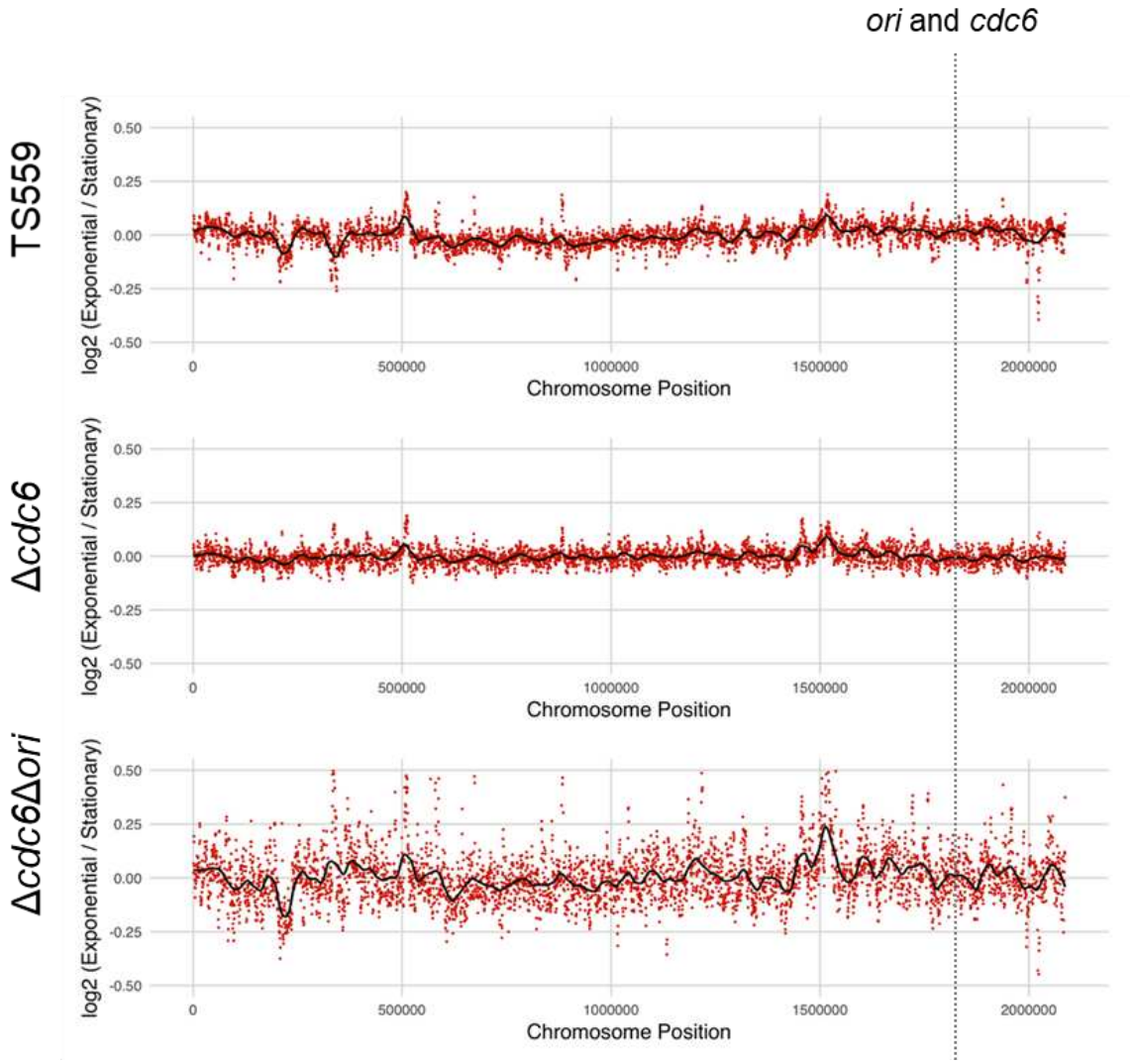

Figure 3.2. Marker frequency analysis of DNA sequence reads from *T. kodakarensis* strains fails to identify a defined origin(s) of replication. The $\log_2$ ratio of each nucleotide in sequences from exponentially growing cells divided by sequence from stationary phase cells is shown in each panel. Frequencies were calculated using 1 Kbp intervals (red dots) with a sliding window of 500 bp generating the average frequency shown in black. The location of the *ori-cdc6* region is indicated.

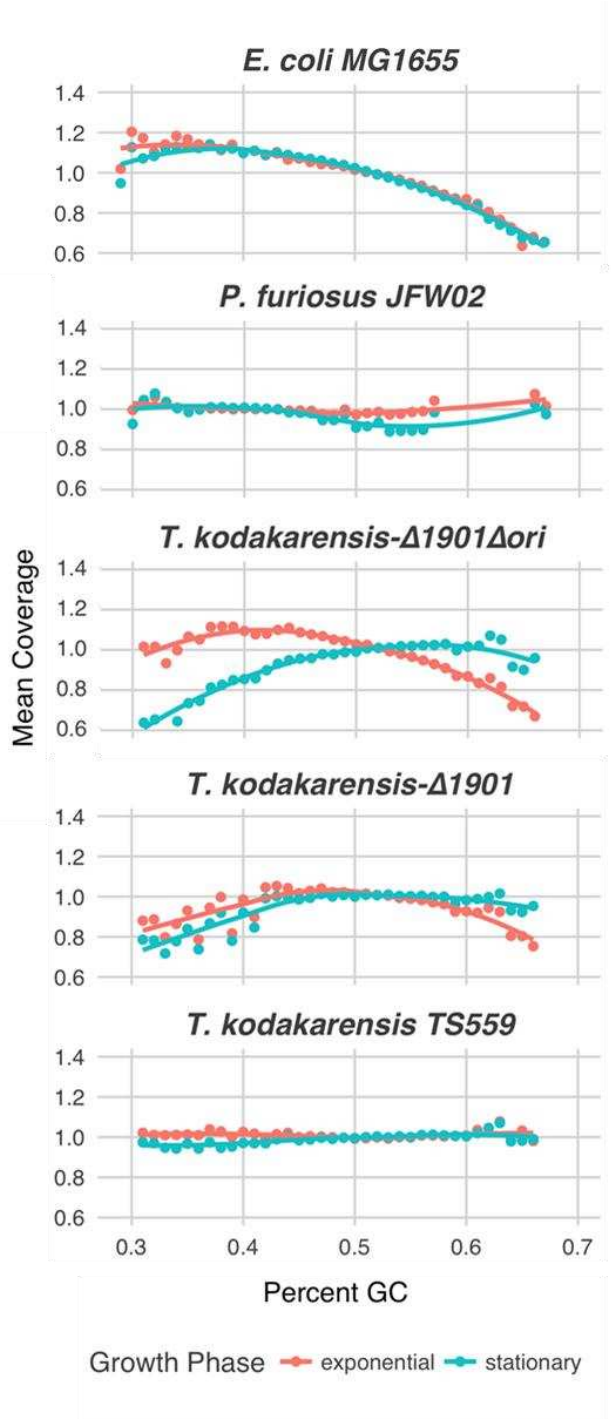

Figure 3.3. GC bias of WGS data of all strains used in the copy number analysis. The normalized coverage is plotted against the %GC content for each bin for the exponential (red) and stationary (blue) cultures.

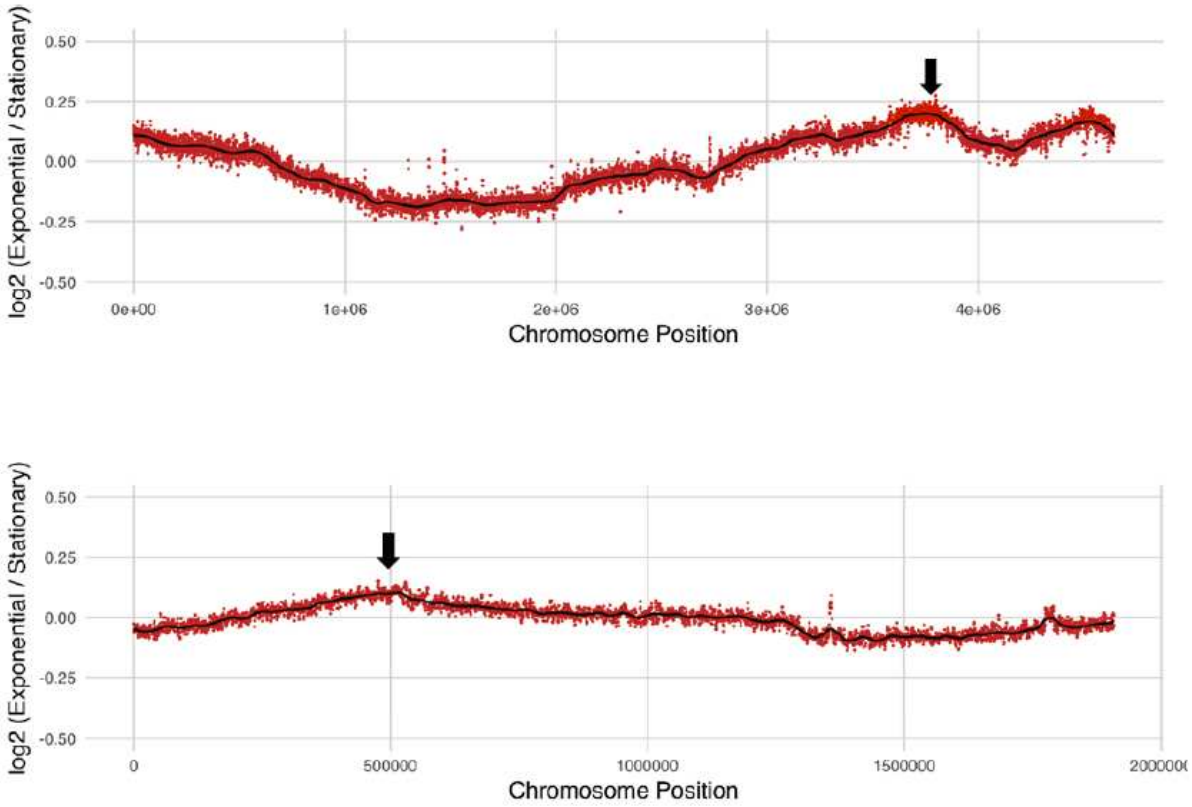

Figure 3.4. Copy number analysis for the *E. coli* MG1655 (top panel) and for *P. furiosus* JFW02 (bottom panel). The $\log_2$ frequency of each nucleotide in the WGS reads is plotted against the genome coordinates. The known, single replication origins (32, 33) were identified for both species.

failed to identify any preferred origin(s) sequences and indicated that replication was initiated at many sites around the genome of *T. kodakarensis* Δ*cdc6* Δ*ori* (Figure 3.2).

Spontaneous genome deletion and inversion

Although the WGS reads did not identify origins of replication, with >2000x genome coverage, they did identify spontaneous recombination events at two locations in subpopulations (<10%) of *T. kodakarensis* TS559 cells (Figure 3.5). The recombinations inverted an ~150 kbp region or excised ~100 kbp, and these events resulted in small spikes and dips in the marker frequency analyses at sites previously established to contain vestigial prophage TKV2 and TKV3 genomic sequences (Figure 3.5)[33,41]. Only a small number of sequences were obtained that extended across the sites of recombination, (Figure 3.5) but these were more prevalent in DNAs isolated from growing than from stationary phase cells. Based on PCR amplicons, these recombination events also occurred in *T. kodakarensis* Δ*cdc6* and Δ*cdc6* Δ*ori*. As deletion of TKV3 severely hinders growth [41] and several presumably essential genes are within the deleted sequences, these recombination events are likely lethal and prevent the effected cells from contributing to continued culture growth.

Mutations acquired by T. kodakarensis since isolation

T. kodakarensis KOD1 has been the focus of continuous research since its isolation in 1994[42]. It seemed possible therefore that the cdc6- and origin-independent replication of *T. kodakarensis* TS559 might result from mutations acquired and selected during laboratory culture. The genome sequence determined here for *T. kodakarensis* TS559 was therefore compared with that published for *T. kodakarensis* KOD1[33]. All the changes known to have been intentionally introduced to generate *T. kodakarensis* TS559 from *T. kodakarensis* KOD1 were present, and although there were no large genome rearrangements, an additional 35 single nucleotide differences were identified. Some of these changes are within open reading frames, but none would be predicted to radically change DNA replication or recombination (Table 3.1).

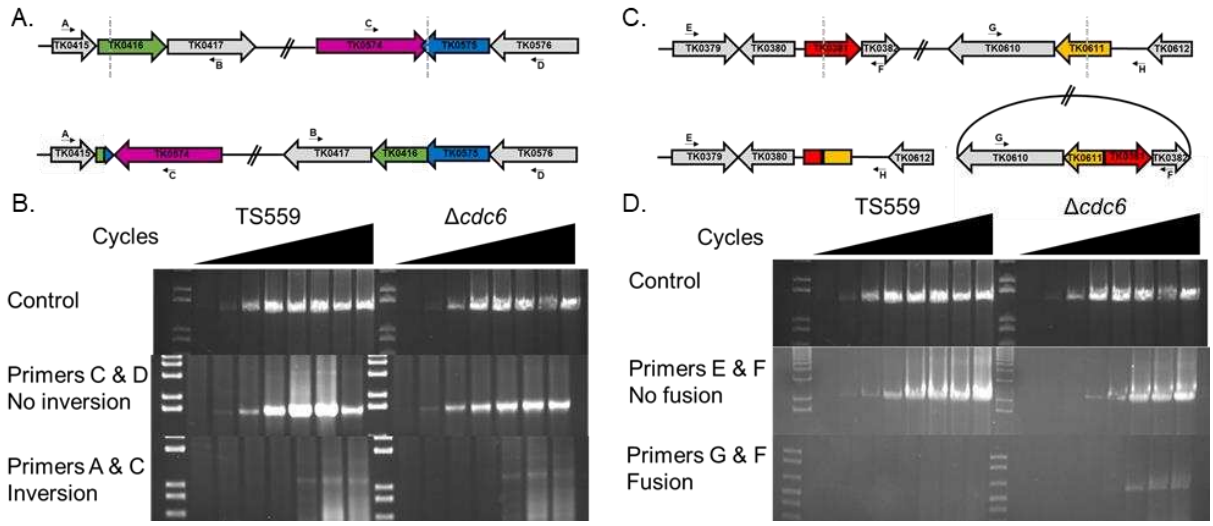

Figure 3.5. Semi-quantitative PCR supports large-scale genomic rearrangements. (A) Representations of TS559 (top) and inverted (bottom) genome structures. The end points of the inversion events are marked with dotted gray lines in the TS559 genome representation. (B) Semi-quantitative PCRs demonstrate the inversion genome rearrangements identified from the WGS data. The presence of the inversion was identified in < 10% of the samples. (C) Representations of TS559 (top) and fusion-event (bottom) genome structures. The end points of the fusion events are marked with dotted gray lines in the TS559 genome representation. (D) Semi-quantitative PCRs demonstrate the fusion event identified from the WGS data in < 10% of the samples.

Table 3.1. Differences in the genome sequences of the *T. kodakarensis* KOD1 and TS559

Position[a]	TK Gene[b]	Operon	DNA change	Protein change	Annotated/Putative function[b] (result of the mutation)
37106	0042	0038→0050	A→G	K185R	Flagellin
76583	0090	0086→0090	A→C	Q187H	Putative S-layer function
96733	0119	0119→0122	A→G	E183G	α-subunit proline dehydrogenase
201247	0238	0237→0241	A→G	E215G	nitrilase; C-H bond hydrolase
229773	0275	0279→0274	C→T	R351K	argD; acetyl-lysine amino transferase
327973	0392	0384→0393	A→G	N.C[c]	Hypothetical
327976	0392	0384→0393	C→A	N.C.	Hypothetical
343671	0415	0410→0419	G→A	G24E	Hypothetical
538367	0634	0631→0638	G→A	N.C.	chemotaxis methyl-acceptor
785924	Inter[d]		C→G	--	
785946	0901	0902→0901	ΔΔG	S115fs[e]	F-subunit RNA polymerase (frameshift extends the ORF –S* to –IDEYRPLE* at C-terminus)
898031	1021	1020→1021	ΔΔA	T800fs	Hef nuclease (frameshift extends the ORF –TGTLR* to –QAPYVEEEDKA* at C-terminus)
912171	1039	1039→1038	T→C	K342R	cyclic 2´3´-diphosphoglycerate synthetase
914113	1041	1041→1042	C→G	D132E	transcription regulator with H-T-H domain

1084046	1236	1236	C→G	A209P	AAA+ family ATPase
1124276	1285	1285	C→T	G102D	transcription regulator; LysR/AsnC with HTH domain
1124363	1285	1285	A→C	L73R	transcription regulator; LysR/AsnC with HTH domain
1127248	inter	tRNA	↑ᵗT	--	
1160792	1315	1315	A→C	F581V	phosphoadenosine phosphosulfate reductase
1160804	1315	1315	A→C	F577V	phosphoadenosine phosphosulfate reductase
1252468	1428	1429→1429	A→G	V20A	metal-dependent RNase with KH-domain
1361362	1554	1554	↑C	P412fs	cellulose synthetase; glycosyl transferase (frameshift changes CTSWFSSLRGLCTP* to - LYFMVFVLAGVVYTMRGLTKLLIGK LTWEKT QFRT* at C-terminus)
1524161	1729	1729→1730	↑A	L211fs	mannosyl transferase (frameshift results in in-frame ORF fusion with TK1730; TK1729 –NGEPATLC* to TK1729 –LKWGARYIV-TK1730)
1580984	1774	1770→1776	ΔΔA	T1069fs	amylopullanase (frameshift extends the ORF– NHHDYYNHIPRRRRKWQRIHHYQH LPRHRRW* to TTTTTTTTSPGGGGSGSGTTTSTS PGT GGGEEGGGICGPAFLVGLAVVPLL LRRRR* at C-terminus; does not overlap TK1775)
1585144	inter		ΔΔG	--	

1596662	1789	1789→1787	T→C	E108G	KaiC domain; recA-like ATPase
1743876	1932	1932→1930	↑T	N30fs	KaiC domain; ABC-family ATPase (frameshift results in 31 in-frame amino acids then *)
1824228	2030	2030	T→C	F70L	ACT-domain; amino acid metabolism regulator
1824230	2030	2030	T→A	F70L	ACT-domain; amino acid metabolism regulator
1824411	2030	2030	ΔΔG	R131fs	ACT-domain; amino acid metabolism regulator (frameshift changes GRNKQDLHSHRWNALNR DIWQNKDNQRLQEAHTPHT* to EETSKIYIVIDGTLSTETFGKIKTIRG FKRLILHTPEKDKEKFVCNYCEVKY CPKRVLLESLTTQR* at C-terminus
1828936	inter		↑G	--	
1860427	2069	2072→2066	↑C	V16fs	α-subunit of cytosolic NiFe hydrogenase (frameshift changes –GRGQGRR to -VEGKGGV* at aa 138)
2011218	2222	2222	T→G	T242P	ATPase
2050612	2262	2261→2263	G→T	R119L	PIN domain, likely VAPC toxin
2078803	2298	2298→2299	C→G	P247A	Anaerobic ribonucleoside reductase class III

a. Genome position 0 was defined in (27)
b. Numerical gene designations, TKxxxx, and annotated functions based on (27)
c. No change
d. Intergenic region
e. Frame shift
f. Nucleotide insertion

Why are Cdc6 and the origin-sequences retained?

T. kodakarensis is oligoploid (7 to 19 genomes/cell[43]) and this is consistent with the use of RDR. Nutrient-stress, the absence of defined DNA segregation strategies, and the potential for continued cell division without DNA replication may occasionally result in *T. kodakarensis* cells with only one genome. Such cells would be unable to restart growth by RDR initiation but could do so if an cdc6-oriC system of replication initiation was also available. To evaluate whether retention of *cdc6* or presumptive origin sequences promotes long-term viability, aliquots were taken from stationary phase cultures maintained at 85˚C for extended periods without nutrient addition, and assayed for viability. Extended (several months) incubation at elevated temperatures in nutrient poor conditions was predicted to deplete energy reserves, introduce stress into the genome, and potentially reduce ploidy as genomes were consumed to provide nutrients. Cells in cultures of *T. kodakarensis* TS559 and *Δcdc6*, strains with the presumed origin (oriC) region, were viable for ~40 days longer than cells in cultures of *T. kodakarensis* Δ*ori* (Figure 3.6). Retention of Cdc6, however, did not influence long-term viability.

TK1899 (RadA) and TK2231 (RadB) are essential genes

RadA was required for growth – presumably by RDR initiation – of the origin-deleted halophilic *H. volcanii* strains but non-essential in the parental, origins-containing strain[6]. Despite repeated attempts with different transforming DNA strategies, we were unable to generate *T. kodakarensis* strains with RadA (TK1899) or RadB (TK2231) deleted, regardless of the retention or absence of *cdc6* and/or origin sequences.

3.3 DISCUSSION

TK1901 encodes what appears to be a fully functional Cdc6 protein, with intact Walker A and Walker B motifs, DNA and ATP-binding domains. TK1901 is co-transcribed with TK1902 and TK1903, essential genes that encode the subunits of DNA polymerase D[35] and is located immediately adjacent to a region with sequences very similar to those of the origin of replication in *P. furiosus*. Nevertheless, the results reported establish conclusively that TK1901 and so

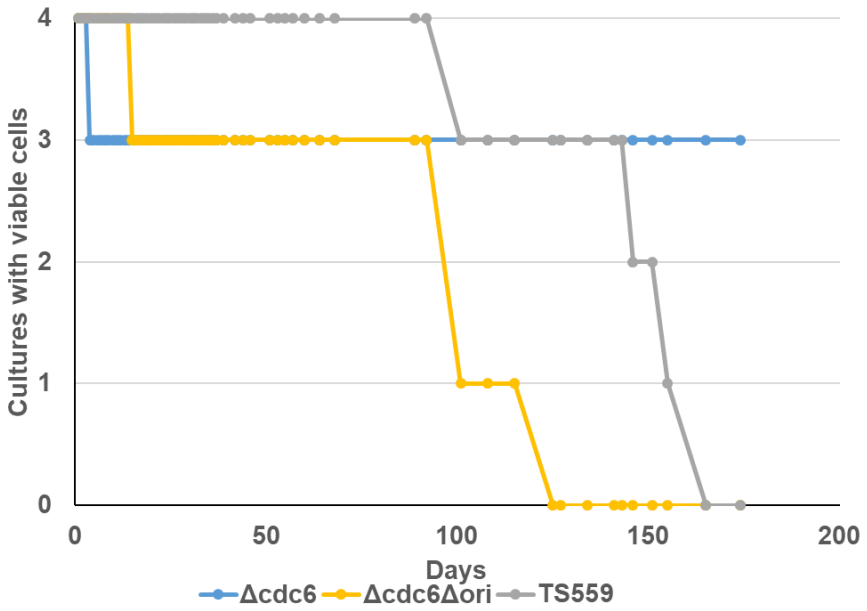

Figure 3.6. Presence of the origin region increases long-term viability. Four biological replicates of *T. kodakarensis* TS559 (blue), Δ*cdc6* (grey), and Δ*cdc6 Δori* (orange) were grown to stationary phase and incubation was continued without additions to the medium for > 170 days. Aliquots were removed at intervals and used to inoculate fresh growth media. The number of cultures with viable cells that generated progeny cultures is plotted against days of incubation at 85˚C.

Cdc6 are not required for *T. kodakarensis* viability and the absence of Cdc6 has no detectable effect on laboratory growth, genetic competence or the ability to support autonomous plasmid replication. Equally surprising, the previously presumed origin of replication can also be deleted without any detectable phenotypic consequence. Consistent with Cdc6 and oriC having no essential roles, marker frequency analyses of WGS data obtained from cultures of *T. kodakarensis* TS559, Δ*cdc6* and Δ*cdc6* Δ*ori* provided no evidence for origin-dependent replication initiation, even when the origin and the recognition protein were both present. Given the depth of the WGS, any regional over-representation of reads, consistent with initiation at as many as 5 separate location would have been detected, but this was not the case. In contrast, the WGS results argue for genome replication in *T. kodakarensis* TS559 being initiated at many sites distributed around the genome, consistent with the proposal for RDR-dependent genome replication in the *H. volcanii* strain with all origins of replication deleted[6]. We extend this proposal to suggest that RDR supports growth of *T. kodakarensis* despite the presence of a predicted origin. In support of this assertion, despite a considerable effort, we were unable to generate *T. kodakarensis* strains with TK1899 (RadA) or TK2231 (RadB) deleted. In *E. coli*, long-lived R-loops accumulate in strains lacking RNase H, and these R-loops can facilitate initiator protein (DnaA) independent constitutive stable DNA replication, but their growth is very slow[2,3,44]. The *T. kodakarensis* strains investigated here all express TK0805, the gene that encodes RNase HII[45] and thus R-loop accumulation is unlikely to be responsible for origin-independent genome replication in *T. kodakarensis*.

Employing RDR for genome replication could also explain why *T. kodakarensis* is atypically naturally competent and so amenable to genetic manipulation. Additional features of *T. kodakarensis* are consistent with RDR. In genomes replicated from distinct origins, highly expressed genes are transcribed predominantly in the same direction as replication fork movement[16,38,46,47]. But, if bidirectional replication was initiated from the previously presumed origin in *T. kodakarensis*, 628 transcripts would be transcribed with, and 626 would be

transcribed against the direction of the replisome movement[16,33,35]. A computational search also failed to identify any location, and so a putative origin, anywhere on the genome that would result in transcription and DNA replication occurring predominantly in the same direction[16]. Often, when cloned, an origin of genome replication will still function and can be used to construct self-replicating plasmids. This is the case for the origin of replication cloned from *P. furiosus*[36] and the generated plasmids replicate not only in the cytoplasm of *P. furiosus* but also in *T. kodakarensis*. In contrast, cloning many variants of the very similar presumed origin region from the *T. kodakarensis* genome did not result in a replicating plasmid.

RDR initiation requires the retention of more than one genome, and it is now well-established that members of the *Euryarchaeota* including *T. kodakarensis,* are naturally oligoploid[43,48,49]. To date, however, there is no evidence for precise genome segregation strategies suggesting that growing cultures will naturally produce cells with varying ploidy – including monoploid cells. If cells occur with only one genome, then the presence of an origin-dependent initiation module would provide a survival mechanism. With this in mind, we demonstrated that cultures of *T. kodakarensis* TS559 and Δ*cdc6* did retain viability longer than *T. kodakarensis* Δ*cdc6* Δ*ori*. Retention of Cdc6 did not influence survival under identical conditions, suggesting that Cdc6 may not be necessary for use of the presumptive origin sequences.

The presence of ≥ ~20 genomes per *T. kodakarensis* cell[43] raises challenging questions as to how they are all accommodated and replicated within a generation time of ~40 minutes. As established for the DNA clamp loader (PCNA1; >1,000 molecules/cell[28,50]), at minimum, the replisome components must be present at very high levels and maybe this also facilitates simultaneous replication from many sites around the *T. kodakarensis* genome. How the replicative apparatus is assembled and how simultaneous rounds of replication are prohibited or accommodated during rapid growth remain outstanding questions.

3.4 MATERIALS AND METHODS

Growth of microorganisms

T. kodakarensis strains were grown in artificial seawater (ASW) supplemented with 5 g/L of both yeast extract and tryptone (YT) and 2 g/L of sulfur (S°) or 5 g/L sodium pyruvate (Pyr) at 85°C. *P. furiosus* strain JFW02 [39] was grown at 95°C as described using maltose as a carbon source[51]. *E. coli* strain MG1655 was grown in Luria-Bertani (LB) broth at 37°C. The growth of cultures was measured by increases in optical density at 600 nm (OD_{600}). *T. kodakarensis* cultures were harvested at an OD_{600} of 0.2 (early exponential) and 0.6 (late-exponential) while the stationary phase cells were harvested at an OD_{600} of ~1.0. In Figure 3.6, four independent cultures of each *T. kodakarensis* strain were maintained at 85˚C in sealed vessels with no additions to the cultures over ~180 days. Loss of culture viability was defined as the inability of aliquots, sampled multiple times over three days, to initiate culture growth when inoculated into fresh medium. The inability of 9 or more individual aliquots, removed from cultures over 3 consecutive days, to support outgrowth confirmed that these cultures had lost all viable CFUs.

Strain construction of T. kodakarensis

Standard procedures[26] were used to construct plasmids, pOSUTK1901B and pJG4 respectively, that were used to delete TK1901 or TK1901 plus the origin sequences from *T. kodakarensis* TS559. In the resulting strains, *T. kodakarensis* Δ*cdc6* and Δ*cdc6* Δ*ori,* the Cdc6 encoding sequence (TK1901; 1,248 bp) was deleted but the promoter for the TK1901-TK1903 operon was retained to sure expression of TK1902-1903. Similarly, in *T. kodakarensis* Δ*cdc6* Δ*ori*, the putative origin (640 bp) with one origin recognition box (ORB) and three mini-ORBs were deleted, but the promoters for TK1901-1903 and TK1900 were retained (Figure 3.1). Use of the same procedures, and plasmids designed to precisely delete the TK1899 (RadA) and TK2231 (RadB) sequences, did not generate viable strains with the desired deletions.

Isolation of genomic DNA

Genomic DNA was isolated from all strains as described[52]. Cells pelleted from cultures at the designate OD_{600} were resuspended in 10% (w/v) sucrose, 20 mM Tris-HCl pH 8.0, and 5 mM EDTA. SDS (2% final concentration) and proteinase K (0.25 mg/ml) were added to the resulting lysate and the mixture was incubated for 1 hour at 55°C. NaCl (1 M final concentration) was then added, the mixture was chilled, centrifuged and an equal volume of isopropanol was added to the clarified supernatant. The precipitated nucleic acids were pelleted, resuspended with 10 mM Tris-HCl pH 8.0, 50 µg of RNase A added and incubation continued at 37°C for 30 minutes. The DNA remaining was further purified by repeated phenol/chloroform/isoamyl alcohol (25:24:1) extraction and an alcohol precipitation.

Southern Blotting

The procedure used has been previously described[30]. In Figure 3.1, TK1901 (Probe 1, pink) is only detectable in strain TS559 whereas origin sequences (Probe 3, blue) are detectable in strains TS559 and Δ*cdc6*, but not strain Δ*cdc6 Δori*; probe 3 highlights a smaller product in Δ*cdc6* that reflects deletion of TK1901. Probe 2 (orange) detected the sequences encoding TK1902 in all samples and the fragment lengths identified are appropriate for the corresponding strains.

Whole genome sequencing (WGS)

Sequencing libraries were prepared using TruSeq DNA library preparation kits (Illumina, San Diego, CA) and were subjected to WGS (pair-ended, 2 × 150 bp per read; 1 × 125 bp reads for *E. coli*) on an Illumina Hi-Seq 2000 platform (University of Colorado Denver Genomics and Microarray Core Facility). Individual genome coverages ranged from 3,300× to 7,800×.

Comparison of genome sequences

The reference genome for *Thermococcus kodakarensis* KOD1 (https://www.ncbi.nlm.nih.gov/nuccore/NC_006624.1) was downloaded from Genbank and manually edited to account for the laboratory manipulations mad in the lineage leading to *T.*

kodakarensis TS559. The reference genomes for *Pyrococcus furiosus* COM1 and *Escherichia coli* MG1655 were downloaded from Genbank here (https://www.ncbi.nlm.nih.gov/nuccore/CP003685 and https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3, respectively).

The *T. kodakarensis* TS559 and KOD1 genomes were aligned and differences identified by using Universal Genotyper (GATK version v2.1-8[53]) and MUMmer (version 3.1[54]). Low quality sequences, regions with <5x coverage and a small number of variants identified in DNA from only growing or stationary phase cells of the same isolate were not included. The coordinates of the RefSeq GFF file were updated to account for identified insertions and deletions, and the resulting variants annotated using snpEff (version 4.0e[55]).

Alignment and marker frequency analysis of whole genome sequences

An index was built for each reference genome with bowtie2-build (version 2.2.9[56]) using default settings. Illumina adaptor sequences and low quality bases (quality score < 10) were trimmed from the 3'-end of each read using cutadapt (version 1.11[57]) with reads discarded if more than half the bases were trimmed. The filtered reads were aligned to the reference genome using bowtie2 (version 2.2.9[56]) selecting the best alignment for each read. The alignment statistics are reported in Table 3.2. Reads that did not align as proper pairs were treated as single end reads. For copy number estimation, each reference genome was binned into 1 kb windows with a 500 bp sliding overlap between windows using bedtools (version 2.17.0[58]) For each sample, the coverage for each 1 kb was calculated as the number of sequenced bases that overlap with that window. The %G+C content of each window was calculated, and corrections made for potential bias in library preparation and sequencing due to GC content. The average coverage for each GC bin was plotted against the GC content and smoothed by a Lowess regression model. A correction factor, calculated by dividing the global mean coverage by the fitted model, was applied to each sample. Source code available at http://github.com/dpastling/plethora.

Supplementary Table 1: Alignment Statistics

Sample	Number of Reads	Filtered Reads	Aligned Reads	Percent Aligned
TS559 exponential	67,633,173	65,914,849	65,058,090	96.2
TS559 stationary	43,521,320	42,699,034	42,102,125	96.7
Δ*cdc6* exponential	66,437,434	64,584,318	63,846,669	96.1
Δ*cdc6* stationary	67,489,844	65,733,849	65,125,236	96.5
Δ*cdc6*Δ*ori* exponential	40,886,671	40,883,270	40,475,385	99.0
Δ*cdc6*Δ*ori* stationary	49,284,832	49,278,072	48,239,873	97.9
MG1655 exponential	130,627,142	129,951,009	125,558,107	96.1
MG1655 stationary	164,618,767	163,804,543	162,897,668	99.0
JFW02 exponential	46,985,782	46,982,363	46,396,801	98.7
JFW02 stationary	42,499,179	42,493,069	33,714,141	79.3

REFERENCES

1. Jacob, F.; Brenner, S.; Cuzin, F. On the Regulation of DNA Replication in Bacteria. *Cold Spring Harbor Symposia on Quantitative Biology* **1963**, *28*, 329–348.
2. Maduike, N. Z.; Tehranchi, A. K.; Wang, J. D.; Kreuzer, K. N. Replication of the Escherichia Coli Chromosome in RNase HI-Deficient Cells: Multiple Initiation Regions and Fork Dynamics. *Molecular microbiology* **2014**, *91*, 39–56.
3. Masai, H.; Asai, T.; Kubota, Y.; Arai, K.; Kogoma, T. Escherichia Coli PriA Protein Is Essential for Inducible and Constitutive Stable DNA Replication. *The EMBO journal* **1994**, *13*, 5338–45.
4. Kogoma, T. Stable DNA Replication: Interplay between DNA Replication, Homologous Recombination, and Transcription. *Microbiology and molecular biology reviews : MMBR* **1997**, *61*, 212–38.
5. Ogawa, T.; Pickett, G. G.; Kogoma, T.; Kornberg, A. RNase H Confers Specificity in the dnaA-Dependent Initiation of Replication at the Unique Origin of the Escherichia Coli Chromosome in Vivo and in Vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 1040–4.
6. Hawkins, M.; Malla, S.; Blythe, M. J.; Nieduszynski, C. A.; Allers, T. Accelerated Growth in the Absence of DNA Replication Origins. *Nature* **2013**, *503*, 544–547.
7. Scholefield, G.; Veening, J. W.; Murray, H. DnaA and ORC: More than DNA Replication Initiators. *Trends Cell Biol* **2011**, *21*, 188–194.
8. Kawakami, H.; Katayama, T. DnaA, ORC, and Cdc6: Similarity beyond the Domains of Life and Diversity. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **2010**, *88*, 49–62.
9. Norais, C.; Hawkins, M.; Hartman, A. L.; Eisen, J. A.; Myllykallio, H.; Allers, T. Genetic and Physical Mapping of DNA Replication Origins in Haloferax Volcanii. *PLoS genetics*

2007, *3*, e77.

10. Beattie, T. R.; Bell, S. D. Molecular Machines in Archaeal DNA Replication. *Current Opinion in Chemical Biology* **2011**, *15*, 614–619.
11. Matsunaga, F.; Takemura, K.; Akita, M.; Adachi, A.; Yamagami, T.; Ishino, Y. Localized Melting of Duplex DNA by Cdc6/Orc1 at the DNA Replication Origin in the Hyperthermophilic Archaeon Pyrococcus Furiosus. *Extremophiles* **2010**, *14*, 21–31.
12. Wigley, D. B. ORC Proteins: Marking the Start. *Current Opinion in Structural Biology* **2009**, *19*, 72–78.
13. Pelve, E. A.; Martens-Habbena, W.; Stahl, D. A.; Bernander, R. Mapping of Active Replication Origins in Vivo in Thaum- and Euryarchaeal Replicons. *Molecular microbiology* **2013**, *90*, 538–50.
14. Matsunaga, F.; Forterre, P.; Ishino, Y.; Myllykallio, H. In Vivo Interactions of Archaeal Cdc6/Orc1 and Minichromosome Maintenance Proteins with the Replication Origin. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 11152–7.
15. Matsunaga, F.; Glatigny, A.; Mucchielli-Giorgi, M.-H.; Agier, N.; Delacroix, H.; Marisa, L.; Durosay, P.; Ishino, Y.; Aggerbeck, L.; Forterre, P. Genomewide and Biochemical Analyses of DNA-Binding Activity of Cdc6/Orc1 and Mcm Proteins in Pyrococcus Sp. *Nucleic Acids Research* **2007**, *35*, 3214–3222.
16. Cossu, M.; Cunha, V. Da; Toffano-Nioche, C.; Forterre, P.; Oberto, J. Comparative Genomics Reveals Conserved Positioning of Essential Genomic Clusters in Highly Rearranged Thermococcales Chromosomes. *Biochimie* **2015**, *118*, 313–21.
17. Dueber, E. C.; Costa, A.; Corn, J. E.; Bell, S. D.; Berger, J. M. Molecular Determinants of Origin Discrimination by Orc1 Initiators in Archaea. *Nucleic Acids Res* **2011**, *39*, 3621–3631.
18. Bell, S. D. Archaeal orc1/cdc6 Proteins. *Subcell Biochem* **2012**, *62*, 59–69.

19. Arora, J.; Goswami, K.; Saha, S. Characterization of the Replication Initiator Orc1/Cdc6 from the Archaeon Picrophilus Torridus. *J Bacteriol* **2014**, *196*, 276–286.

20. Samson, R. Y.; Xu, Y.; Gadelha, C.; Stone, T. A.; Faqiri, J. N.; Li, D.; Qin, N.; Pu, F.; Liang, Y. X.; She, Q.; Bell, S. D. Specificity and Function of Archaeal DNA Replication Initiator Proteins. *Cell Rep* **2013**, *3*, 485–496.

21. Wu, Z.; Liu, J.; Yang, H.; Liu, H.; Xiang, H. Multiple Replication Origins with Diverse Control Mechanisms in Haloarcula Hispanica. *Nucleic Acids Res* **2014**, *42*, 2282–2294.

22. Barry, E. R.; Bell, S. D. DNA Replication in the Archaea. *Microbiology and Molecular Biology Reviews* **2006**, *70*, 876–887.

23. Robinson, N. P.; Bell, S. D. Origins of DNA Replication in the Three Domains of Life. *The FEBS journal* **2005**, *272*, 3757–66.

24. Makarova, K. S.; Koonin, E. V. Archaeology of Eukaryotic DNA Replication. *Cold Spring Harbor Perspectives in Biology* **2013**, *5*, a012963–a012963.

25. Ausiannikava, D.; Allers, T. Diversity of DNA Replication in the Archaea. *Genes* **2017**, *8*, 56.

26. Hileman, T. H.; Santangelo, T. J. Genetics Techniques for Thermococcus Kodakarensis. *Front Microbiol* **2012**, *3*, 195.

27. Pan, M.; Santangelo, T. J.; Li, Z.; Reeve, J. N.; Kelman, Z. Thermococcus Kodakarensis Encodes Three MCM Homologs but Only One Is Essential. *Nucleic Acids Res* **2011**, *39*, 9671–9680.

28. Pan, M.; Santangelo, T. J.; Čuboňová, L.; Li, Z.; Metangmo, H.; Ladner, J.; Hurwitz, J.; Reeve, J. N.; Kelman, Z.; Cubonova, 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 : life under extreme conditions* **2013**, *17*, 453–61.

29. Li, Z.; Santangelo, T. J.; Cubonova, L.; Reeve, J. N.; Kelman, Z. Affinity Purification of an

Archaeal DNA Replication Protein Network. *MBio* **2010**, *1*.

30. Cubonova, L.; Richardson, T.; Burkhart, B. W.; Kelman, Z.; Connolly, B. A.; Reeve, J. N.; Santangelo, T. J. Archaeal DNA Polymerase D but Not DNA Polymerase B Is Required for Genome Replication in Thermococcus Kodakarensis. *J Bacteriol* **2013**, *195*, 2322–2328.
31. Li, Z.; Huang, R. Y.-C.; Yopp, D. C.; Hileman, T. H.; Santangelo, T. J.; Hurwitz, J.; Hudgens, J. W.; Kelman, Z. A Novel Mechanism for Regulating the Activity of Proliferating Cell Nuclear Antigen by a Small Protein. *Nucleic Acids Research* **2014**, *42*, 5776–5789.
32. Ojha, K. K.; Swati, D. Mapping of Origin of Replication in Themococcales. *Bioinformation* **2010**, *5*, 213–218.
33. 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* **2005**, *15*, 352–363.
34. Xu, J.; Yanagisawa, Y.; Tsankov, A. M.; Hart, C.; Aoki, K.; Kommajosyula, N.; Steinmann, K. E.; Bochicchio, J.; Russ, C.; Regev, A.; Rando, O. J.; Nusbaum, C.; Niki, H.; Milos, P.; Weng, Z.; Rhind, N. Genome-Wide Identification and Characterization of Replication Origins by Deep Sequencing. *Genome Biology* **2012**, *13*, R27.
35. Jager, D.; Forstner, K. U.; Sharma, C. M.; Santangelo, T. J.; Reeve, J. N. Primary Transcriptome Map of the Hyperthermophilic Archaeon Thermococcus Kodakarensis. *BMC Genomics* **2014**, *15*, 684.
36. Farkas, J.; Chung, D.; DeBarry, M.; Adams, M. W. W.; Westpheling, J. Defining Components of the Chromosomal Origin of Replication of the Hyperthermophilic Archaeon Pyrococcus Furiosus Needed for Construction of a Stable Replicating Shuttle Vector. *Applied and environmental microbiology* **2011**, *77*, 6343–9.
37. Santangelo, T. J.; Cubonova, L.; Reeve, J. N. Shuttle Vector Expression in

Thermococcus Kodakaraensis: Contributions of Cis Elements to Protein Synthesis in a Hyperthermophilic Archaeon. *Appl Environ Microbiol* **2008**, *74*, 3099–3104.

38. Andersson, A. F.; Pelve, E. A.; Lindeberg, S.; Lundgren, M.; Nilsson, P.; Bernander, R. Replication-Biased Genome Organisation in the Crenarchaeon Sulfolobus. *BMC Genomics* **2010**, *11*, 454.
39. Farkas, J.; Stirrett, K.; Lipscomb, G. L.; Nixon, W.; Scott, R. A.; Adams, M. W. W.; Westpheling, J. Recombinogenic Properties of Pyrococcus Furiosus Strain COM1 Enable Rapid Selection of Targeted Mutants. *Applied and environmental microbiology* **2012**, *78*, 4669–76.
40. Burland, V.; Plunkett, G.; Daniels, D. L.; Blattner, F. R. DNA Sequence and Analysis of 136 Kilobases of the Escherichia Coli Genome: Organizational Symmetry around the Origin of Replication. *Genomics* **1993**, *16*, 551–61.
41. 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* **2013**, *17*, 153–160.
42. Morikawa, M.; Izawa, Y.; Rashid, N.; Hoaki, T.; Imanaka, T. Purification and Characterization of a Thermostable Thiol Protease from a Newly Isolated Hyperthermophilic Pyrococcus Sp. *Applied and environmental microbiology* **1994**, *60*, 4559–66.
43. Spaans, S. K.; Oost, J. van der; Kengen, S. W. M. The Chromosome Copy Number of the Hyperthermophilic Archaeon Thermococcus Kodakarensis KOD1. *Extremophiles : life under extreme conditions* **2015**, *19*, 741–50.
44. Masai, H.; Arai, K. Mechanisms of Primer RNA Synthesis and D-loop/R-Loop-Dependent DNA Replication in Escherichia Coli. *Biochimie* **1996**, *78*, 1109–17.
45. Heider, M. R.; Burkhart, B. W.; Santangelo, T. J.; Gardner, A. F. Defining the RNaseH2 Enzyme-Initiated Ribonucleotide Excision Repair Pathway in Archaea. *Journal of*

Biological Chemistry **2017**, *292*, 8835–8845.

46. Smith, C. E.; Llorente, B.; Symington, L. S. Template Switching during Break-Induced Replication. *Nature* **2007**, *447*, 102–5.

47. Paul, S.; Million-Weaver, S.; Chattopadhyay, S.; Sokurenko, E.; Merrikh, H. Accelerated Gene Evolution through Replication–transcription Conflicts. *Nature* **2013**, *495*, 512–515.

48. Hildenbrand, C.; Stock, T.; Lange, C.; Rother, M.; Soppa, J. Genome Copy Numbers and Gene Conversion in Methanogenic Archaea. *Journal of bacteriology* **2011**, *193*, 734–43.

49. Breuert, S.; Allers, T.; Spohn, G.; Soppa, J. Regulated Polyploidy in Halophilic Archaea. *PloS one* **2006**, *1*, e92.

50. Kuba, Y.; Ishino, S.; Yamagami, T.; Tokuhara, M.; Kanai, T.; Fujikane, R.; Daiyasu, H.; Atomi, H.; Ishino, Y. Comparative Analyses of the Two Proliferating Cell Nuclear Antigens from the Hyperthermophilic Archaeon, Thermococcus Kodakarensis. *Genes to cells : devoted to molecular & cellular mechanisms* **2012**, *17*, 923–37.

51. Adams, M. W. W.; Holden, J. F.; Menon, A. L.; Schut, G. J.; Grunden, A. M.; Hou, C.; Hutchins, A. M.; Jenney, F. E.; Kim, C.; Ma, K.; Pan, G.; Roy, R.; Sapra, R.; Story, S. V.; Verhagen, M. F. J. M. Key Role for Sulfur in Peptide Metabolism and in Regulation of Three Hydrogenases in the Hyperthermophilic Archaeon Pyrococcus Furiosus. *Journal of Bacteriology* **2001**, *183*, 716–724.

52. Santangelo, T. J.; Cubonova, L.; James, C. L.; Reeve, J. N.; 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* **2007**, *367*, 344–357.

53. Auwera, G. A. Van der; Carneiro, M. O.; Hartl, C.; Poplin, R.; Angel, G. del; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; Banks, E.; Garimella, K. V.; Altshuler, D.; Gabriel, S.; DePristo, M. A. From fastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* **2013**.

54. Kurtz, S.; Phillippy, A.; Delcher, A. L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S. L. Versatile and Open Software for Comparing Large Genomes. *Genome biology* **2004**, *5*, R12.

55. Cingolani, P.; Platts, A.; Wang, L. L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S. J.; Lu, X.; Ruden, D. M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92.

56. Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods* **2012**, *9*, 357–359.

57. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal* **2011**, *17*, 10–12.

58. Quinlan, A. R.; Hall, I. M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26*, 841–842.

CHAPTER 4

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

4.1 SUMMARY

The work presented in this dissertation fills significant gaps in our knowledge of archaeal information processing systems and lays the foundation for studying strategies in archaeal organisms that regulate gene expression and cellular growth. Preliminary studies presented here will provide evidence for experimental assays that will definitively determine if transcription-coupled DNA repair (TCR) is present in *Thermococcus kodakarensis*. All available evidence is in support of TCR[1–3], and thus the established procedures were designed to test the requirements for, and the activities of the likely unique archaeal-specific factors involved in TCR, transcription termination, and DNA repair. Additionally, the regulation of DNA replication, or lack thereof in *T. kodakarensis* was demonstrated. Origin-independent DNA replication is the only predominant mode of initiating cellular DNA replication in *Thermococcus kodakarensis*. Questions remain as to whether *T. kodakarensis* is unique in its replicative strategy or whether RDR is more broadly employed, and if so, how is RDR controlled and regulated? Studies to determine how other *Thermococcales* replicate their genomes will provide insight into the evolutionary history of DNA replication and the factors involved.

4.2 A SYSTEM FOR STUDYING TCR IN *T. KODAKARENSIS*

TCR has only been defined in two archaeal organisms, *Haloferax volcanii* and *Halobacterium* NRC-1[3]. Both species encode obvious homologs of the bacterial UvrA, UvrB, and UvrC proteins, likely acquired via horizontal gene transfer[4]. In Bacteria, the Uvr proteins are responsible for initiating TCR, and this is also the case for *Halobacterium* NRC-1[5]. Interestingly, TCR in *H. volcanii* is not dependent on the UvrA protein and thus TCR in this species is

therefore likely dependent on unknown archaeal specific proteins that recognize RNAP stalled at template strand lesions and initiate TCR[3].

A possible candidate for the archaeal-specific transcription-repair coupling factor (TRCF) may be the recently described transcription termination factor Eta[2]. *T. kodakarensis* strains lacking Eta exhibit an increased sensitivity to DNA damaging agents, including UV light, which suggests a role in DNA repair. The mechanism Eta uses to disrupt stalled transcription elongation complexes mimics Mfd-mediated transcription termination, and Mfd is the bacterial TRCF[5–8]. *In vitro* transcription reactions using purified *T. kodakarensis* RNAP and DNA templates with DNA lesions in either the template or the non-template strand demonstrated that the RNAP from *T. kodakarensis* is sensitive to DNA lesions only in the template strand of DNA, a hallmark of TCR[1]. Combined, the likely presence of an archaeal specific TRCF suggested by studies in *H. volcanii*, the similarities between Eta and Mfd, and the response of archaeal RNAP to DNA lesions only in the template strand of DNA suggest that TCR exists in *T. kodakarensis* and therefore likely other euryarchaeal organisms[1–3]. To facilitate studies investigating TCR in *T. kodakarensis*, I have set up system for demonstrating either the presence or absence of TCR following UV irradiation.

Construction of isogenic strains to study TCR in vivo

TCR is defined by a faster rate of DNA repair in the template strand of DNA compared to the non-template strand within transcribed regions of the genome. It is therefore necessary to monitor and report the rate of DNA repair in each strand in both the presence and absence of transcription to provide evidence for TCR. In most instances, the global rate of strand specific repair is not reported or even calculated, but rather the rate of a single gene or operon is reported and used as a proxy for the genome-wide rate of strand specific repair. In many bacterial and eukaryotic systems, easily controlled promoters can be switched “on” or “off”, and thus a single strain may provide a platform to monitor the rate of strand-specific DNA repair under transcription-permissive versus transcription-repressive conditions. Unfortunately, such

an easily controlled promoter is not yet available for hyperthermophilic archaeal systems, and thus I was forced to develop an alternative, two-strain mechanism to monitor the rate of strand specific DNA-repair in the absence and presence of transcription.

Two otherwise isogenic *T. kodakarensis* strains, termed here as AG1 and AG2, were constructed to modify the expression of the TK1761-3 operon using standard lab techniques[9,10] (Figure 4.1). TK1761 encodes a β-glycosidase while TK1762 and TK1763 encode proteins with unknown function[10,11]. Previous studies demonstrated that modifications of the promoter region for TK1761-3 could repress or highly activate transcription of this genomic locus, but that strains with differential expression of TK1761-3 did not display any phenotypic changes in growth rate or final densities[10]. Strain AG1 was constructed to place TK1761-3 under control of a constitutively expressed, exogenous promoter (P_{hmtB}) known to increase TK1761-3 expression well-above the expression levels observed under standard laboratory growth. Strain AG2, in contrast, was designed and constructed to abolish transcription of the TK1761-3 operon, and this was accomplished by deleting sequences upstream of TK1761 to remove the promoter elements necessary to recruit RNAP and the basal transcription factors. Once constructed and sequence-verified, these two otherwise isogenic strains provide a mechanism to monitor the rate of strand-specific DNA repair in the absence or presence of transcription.

To verify the desired consequences for TK1761-3 expression levels in strains AG1 and AG2, total RNA preparations were made from each strain, and the levels of TK1761-3 expression were quantified by qRT-PCR. A housekeeping gene encoding the S-layer protein (TK0895), with no known co-regulation or influence of TK1761-3, was used as a reference signal to quantify RNA levels of TK1761-3 in the parental strain (TS559) and both experimental strains (AG1 and AG2)[11,12]. The qRT-PCR results confirmed that, relative to TS559, deletion of promoter elements led to no detectable expression of TK1761-3 in strain AG2, whereas expression of TK1761-3 was highly increased in strain AG1 that employed the heterologous promoter to drive expression of the operon (Figure 4.2). The results obtained confirm that

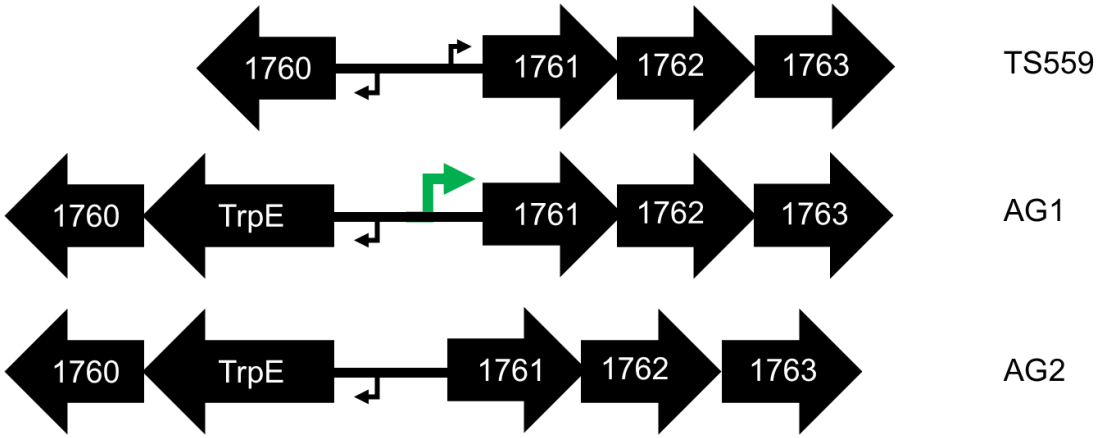

Figure 4.1. *T. kodakarensis* genome maps. TS559 is the parent genome. AG1 and AG2 are isogenic except for alterations in the promoter: Expression of TK1761-3 in TS559 is under control of the native promoter, AG1 a constitutive promoter, and AG2 a deleted promoter.

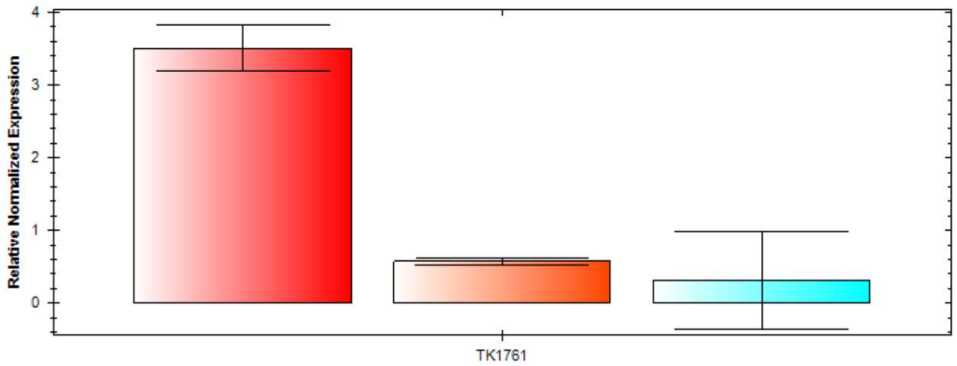

Figure 4.2. Relative expression levels of TK1761 in strains AG1 (red), AG2 (orange), and TS559 (aqua). Expression levels were normalized to TK0895, a constitutively expressed gene.

strains AG1 and AG2 constitutively express and block transcription, respectively, of TK1761-3 and thus provide a platform to monitor strand-specific DNA repair in *T. kodakarensis*.

Demonstrating the presence or absence of TCR in T. kodakarensis

There is no simple assay that can be used to monitor the rate of repair in both strands of DNA simultaneously. Instead, a multi-step, strand-specific Southern blot is traditionally used to determine the rate of repair in each strand of DNA[13]. The many steps of this Southern blot must be individually optimized (Figure 4.3).

The overall scheme relies on the introduction of DNA damage via UV irradiation, followed by periods of recovery that permit the cell to repair some (or all) of the introduced damage. The rate of repair is quantified by monitoring the repair of cyclopyrimidine dimers (CPDs) in each DNA strand of the reporter gene. Total genomic DNA is recovered from aliquots of cells that were allowed increasing periods of recovery, then digested with common restriction enzymes to release a small (~1 Kb) fragment that encodes the TK1761-3 reporter construct. The time-dependent retention or repair of CPDs in DNAs is estimated by the sensitivity of the released reporter-containing DNA to digestion by T4 Endonuclease V (TEV), an enzyme that will nick the DNA at the 5′ end of the CPDs[14,15]. TEV-nicking of the DNA reduces the size of the fragments, and when resolved in denaturing gels and observed via strand-specific Southern blotting, the rate of repair is calculated by the time required to generate DNA fragments that are resistant to TEV-digestion.

To ensure TEV was functional and specific for CPDs, pUC19 with ~1 CPD per plasmid was generated and used to test the enzyme and reaction conditions. TEV-induced plasmid nicking relaxes the plasmid DNA such that the cut vector resolves at a different position in an agarose gel than the original supercoiled plasmid (Figure 4.4). TEV is active under our experimental conditions, and thus is supportive to monitor archaeal TCR.

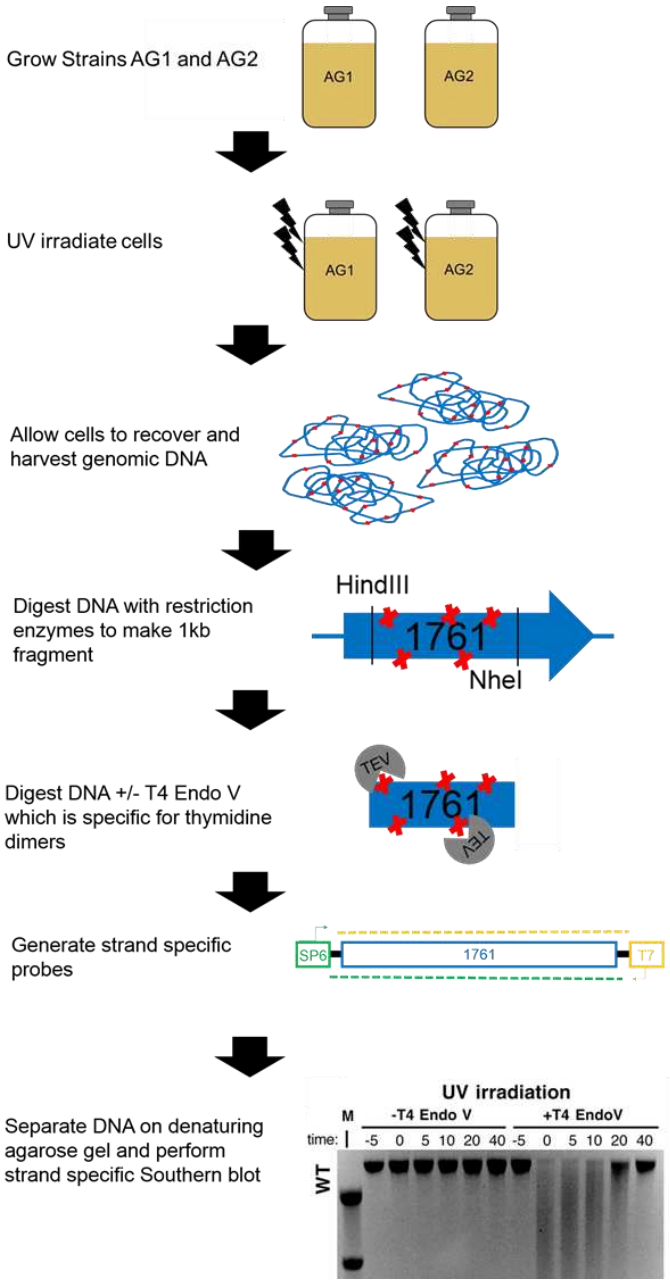

Figure 4.3. Schematic of the Southern blot experiment used to determine the presence of absence of TCR. *T. kodakarensis* strains expressing TK1761-3 (AG1) or not expressing TK1761-3 (AG2) are grown overnight and then UV irradiated. Follow UV irradiation, the cells are then allowed to recover for varying time points. Genomic DNA is harvested from cells and then digested with HindIII and NheI restriction enzyme to release a ~1 Kb fragment of DNA that will be probed in the Southern blot. This DNA is then digested with TEV or mock digested and resolved in a denaturing agarose gel and transferred to a positively charged nylon membrane. Simultaneously, radio-labeled RNA probes specific for either the template or the non-template strand are synthesized. The RNA probes are then hybridized to the membrane-bound denatured DNA. The hybridized and washed membrane is then imaged using a phosphorimager screen. If the resultant image looks similar to the bottom panel, TCR occurs in the cell. Reference gel from Schalow *et al.*, 2012.

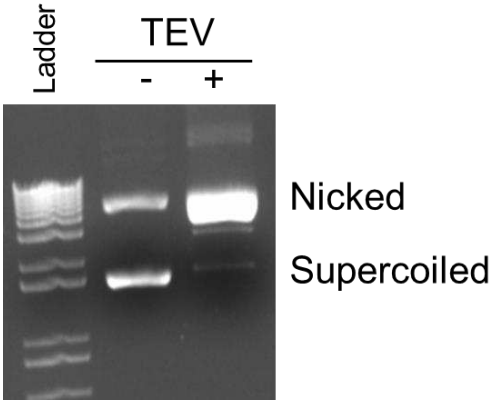

Figure 4.4. pUC19 with ~1 CPD per plasmid is incubated with or without TEV to determine if TEV was active under the reaction conditions. TEV will recognize a CPD and nick the DNA which relaxes the supercoil. This modifies the migration of the plasmid DNA in a standard agarose gel.

Synthesis of RNA probes specific for either the template or non-template strand of DNA

To determine the rate of DNA repair in the template strand of DNA compared to the non-template strand of DNA, two probes, each specific for only one strand of DNA, must be synthesized. The RNA probes are specific to the ~1 Kb region within TK1761-3 that is present after restriction digest of genomic DNA. To generate the RNA probes, the DNA fragment of interest was cloned into the pGEM-3Z vector, a standard *Escherichia coli* cloning vector used for RNA synthesis[16] resulting in the newly constructed pAMG1 vector. The region of interest in pAMG1 is flanked on either side by a SP6 or T7 phage RNAP promoter. (Figure 4.5). Using phage RNAP, purified pAMG1, ribonucleotides, and ^{32}P-CTP, a radiolabeled RNA probe is synthesized. Prior to use of the probe in the Southern blot, pAMG1 is removed using a DNase treatment. The probe synthesized using the SP6 RNAP is specific for the non-template strand whereas the probe synthesized by T7 RNAP is specific for the template strand.

Denaturing agarose gel and Southern blot conditions necessary for strand-specific probing

The rate of repair for each strand of DNA is determined by monitoring the amount of full-length (~ 1 Kb) restriction digested DNA fragments that are TEV-resistant following UV irradiation. Since TEV only nicks the DNA at the site of a CPD, a denaturing agarose gel must be run to separate the two strands of DNA. Traditionally, a NaOH denaturing gel is run[13], however the standard conditions have not yet yielded reproducible results that can be quantified. The only buffering solution in the NaOH denaturing agarose gels is NaOH which necessitates running condition of <30 volts and use of a recirculating buffer. Due to the obscure running conditions, it is common that the DNA does not resolve in tight bands and diffuses out of the gel overtime resulting in inconsistent results. As an alternative to the NaOH denaturing agarose gel, I have investigated protocols for formamide- or glyoxal-based denaturing agarose gels. Both of these protocols produce consistent results in the denaturation of the DNA and resolution of the gels, however the standard Southern blot hybridization techniques used in combination with either the

Figure 4.5. Close-up view of the region of pAMG2, the plasmid used for the generation of the strand specific RNA probes. The SP6 or T7 phage RNAP and corresponding promoters are used to synthesize the non-template strand or template strand probe, respectively.

formamide or glyoxal-based agarose produces images with large amounts of non-specific binding. This non-specific binding decreases the signal to noise ratio in the final image to such a great extent that the images cannot be analyzed. Continued work is focused on improving the resolution and the consistency of the NaOH denaturing gels as well as defining the Southern blot hybridization conditions necessary for reproducible results with both the formamide- and glyoxal-based denaturing gels.

Following denaturation of the DNA in the agarose gel, the DNA is then transferred and cross-linked to a positively charged nylon membrane. This membrane is then pre-incubated with a hybridization buffer consisting of 50% formamide. The radiolabeled RNA probe specific for either the template or the non-template strand of DNA is then added to the hybridization buffer and the membrane and incubated for at least 12 hours. To remove any non-specifically bound RNA probes, the membrane is washed in a series of SSC and SDS buffers. The washed membrane is ultimately exposed to a phosphorimager screen. My Southern blot conditions in combination with the NaOH denaturing agarose gels have been optimized to allow for as little as 3 µg of DNA in a single land to be detected (Figure 4.6). Hybridization and washing conditions will be adapted and modified as needed for the formamide- and glyoxal-based denaturing gels.

4.3 REGULATION OF ARCHAEAL TCR

TCR is still present in *H. volcanii* strains lacking UvrA, necessitating that an archaeal specific TCR pathway exists[3]. Eta may remove RNAP from the site of template-strand DNA lesions, but the remaining factors in the TCR pathway are unknown. Nucleotide excision repair (NER) has not yet been demonstrated in any archaeal clade, but if TCR does exist (and all evidence is in support of such), and TCR is normally a specialized version of NER, then what proteins are likely responsible for both NER and TCR in *T. kodakarensis*? The archaeal-encoded XP proteins, homologous to factors known to participate directly in eukaryotic TCR, possibly

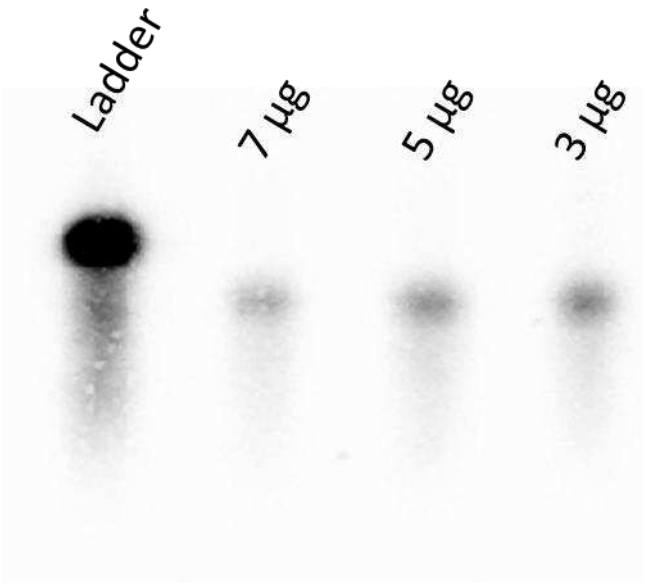

Figure 4.6. Southern blot using the RNA probe specific for the template strand of DNA to detect undamaged genomic DNA from *T.* kodakarensis resolved in a denaturing NaOH agarose gel.

have a role in archaeal TCR[17], and thus we also established mechanisms to directly test the role of XP proteins and Eta in archaeal TCR.

Role of XP Proteins

T. kodakarensis encodes three proteins homologous to factors involved in TCR in Eukarya, XPB, XPD, and XPF (TK0928, TK0784, TK1021, respectively)[11]. These three proteins are crucially important to TCR in Eukarya, and lack of XP protein activity results in Xeromderma Pigemtosum (XP). Patients with XP are exceptionally sensitive to UV light, and half of patients with XP develop skin cancer by age 10[18,19]. Previous studies in *T. kodakarensis* demonstrated strains lacking the XP proteins were sensitive to UV irradiation and other DNA damaging agents[17]. Strains lacking each the XP proteins are in hand and will be further modified to encode the promoter constructs upstream of TK1761-3 thus allowing for determination of their role, if any, in TCR in *T. kodakarensis*.

Role of the termination factor, Eta

Recently, the first archaeal transcription termination factor, Eta, was described in *T. kodakarensis*[2]. Eta is conserved across the euryarchaeal clade, the only archaeal clade in which TCR has been described. The termination mechanism of Eta is reminiscent of the bacterial termination factor and TRCF, Mfd[2]. Additionally, the deletion strain of *eta* is sensitive to UV irradiation, a commonality among TRCF[2,5,8,20]. Combined these results suggest that Eta may be a TRCF in *T. kodakarensis* and possibly all euryarchaeal organisms. The strain lacking Eta has been additionally modified to encode either the strong promoter or no promoter upstream of TK1761-3.

4.4 REPLICATION INITIATION IN THE *THERMOCOCCALES*

Studies of regulation of replication initiation have been onerous in archaeal organisms, and many questions about replication initiation remain, including the role of the Cdc6, the mechanism of Mcm loading, and the signals that regulate replication initiation. All archaea, except *Methanopyrus kandleri*, encode at least one Cdc6 protein[21]. It was predicted that at least

copy of Cdc6 was required for cell viability in all archaeal organisms. However, recent demonstrations of Cdc and origin-independent replication initiation in *T. kodakarensis* and origin-independent replication in *H. volcanii* have shifted this thinking[22].

In *H. volcanii* strains with all replication origins deleted, the cells still carried out DNA replication; however these cells also became dependent on the recombination proteins, RadA and RadB[22]. Further, the cells lacking all replication origins had an increased rate of growth compared to wildtype[22]. Interestingly, *T. kodakarensis* is dependent on recombination proteins under all conditions, and it has been demonstrated that under standard laboratory growth conditions the cells do not utilize the origin of replication and instead use origin-independent replication initiation. Experimental evidence from both *H. volcanii* and *T. kodakarensis* suggests replication is likely initiated via recombination dependent replication initiation (RDR), however it is unknown how or if this recombination is regulated in archaeal cells.

Role of recombination

It has been known for decades that recombination plays a role in DNA repair, and deficiencies in recombination increase susceptibility of a cell to DNA damaging agents. Additionally, recombination also plays a role in initiation of origin-independent DNA replication[23]. In prokaryotic cells, the initiation of replication randomly throughout the genome can be detrimental as replication forks are more likely to collide. One of the most intriguing questions about origin-independent replication initiation, specifically recombination driven replication initiation (RDR), concerns the cell's ability to discriminate between recombination events that are meant to repair the DNA verse those that are meant to initiate DNA replication.

In most cells, the expression of replication proteins is tightly linked to the cell cycle, however the euryarchaea do not have a canonical cell cycle and in some euryarchaeal organisms the DNA replication proteins are constitutively expressed[24–26]. Further complicating the lack of apparent regulation of the expression of DNA replication proteins, RDR has been hypothesized to be the mechanism by which *T. kodakarensis* initiates replication of its genome.

In *T. kodakarensis* is every recombination event a possible replication initiation event, and could this explain why the level of oligoploidy is so variable in *T. kodakarensis*[27]?

Retention of replication initiation proteins in T. kodakarensis

Origin-independent genome replication initiation, therefore also Cdc6 independent, is the preferred mode for initiating DNA replication in *T. kodakarensis* and origin-independent replication initiation occurs both in the presence and absence of Cdc6. What then is the evolutionary advantage of retention of Cdc6? Long-term starvation studies demonstrated that wild-type cells were viable over a longer period than cells that lacked the origin and Cdc6. It is hypothesized that retention of Cdc6 and the origin sequence is valuable during times of cell stress when the number of genomes retained by a cell drastically decreases and only one genome remains. When only 1 genome remains, RDR (dependent on >1 genome per cell) will no longer be feasible, and the cells must utilize origin-dependent replication initiation to survive. *T. kodakarensis* was initially isolated from a solfatara in the Pacific ocean where it is subject to ocean currents and could easily be swept away from its preferred environment[28,29]. Although *T. kodakarensis* has archeaellum which it can use to swim[11,30,31], individual cells are likely to go through periods of starvation where the presence of both Cdc6 and origin or replication would be crucial for cell survival.

REFERENCES

1. Gehring, A. M.; Santangelo, T. J. Archaeal RNA Polymerase Arrests Transcription at DNA Lesions. *Transcription* **2017**, e1324941.
2. Walker, J. E.; Luyties, O.; Santangelo, T. J. Factor-Dependent Archaeal Transcription Termination. *Proceedings of the National Academy of Sciences* **2017**, *114*.
3. Stantial, N.; Dumpe, J.; Pietrosimone, K.; Baltazar, F.; Crowley, D. J. Transcription-Coupled Repair of UV Damage in the Halophilic Archaea. *DNA repair* **2016**, *41*, 63–8.
4. Crowley, D. J.; Boubriak, I.; Berquist, B. R.; Clark, M.; Richard, E.; Sullivan, L.; DasSarma, S.; McCready, S. The uvrA, uvrB and uvrC Genes Are Required for Repair of Ultraviolet Light Induced DNA Photoproducts in Halobacterium Sp. NRC-1. *Saline systems* **2006**, *2*, 11.
5. Ganesan, A.; Spivak, G.; Hanawalt, P. C. Transcription-Coupled DNA Repair in Prokaryotes. *Progress in molecular biology and translational science* **2012**, *110*, 25–40.
6. 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* **2002**, *109*, 757–767.
7. Llorente, B.; Smith, C. E.; Symington, L. S. Break-Induced Replication: What Is It and What Is It For? *Cell cycle (Georgetown, Tex.)* **2008**, *7*, 859–64.
8. Schalow, B. J.; Courcelle, C. T.; Courcelle, J. Mfd Is Required for Rapid Recovery of Transcription Following UV-Induced DNA Damage but Not Oxidative DNA Damage in Escherichia Coli. *Journal of bacteriology* **2012**, *194*, 2637–45.
9. Hileman, T. H.; Santangelo, T. J. Genetics Techniques for Thermococcus Kodakarensis. *Front Microbiol* **2012**, *3*, 195.
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*

2008, *190*, 2244–2248.

11. 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* **2005**, *15*, 352–363.
12. Mattiroli, 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* **2017**, *357*, 609–612.
13. Spivak, G.; Pfeifer, G. P.; Hanawalt, P. In Vivo Assays for Transcription-Coupled Repair. *Methods Enzymol* **2006**, *408*, 223–246.
14. Yasuda, S.; Sekiguchi, M. T4 Endonuclease Involved in Repair of DNA. *Proceedings of the National Academy of Sciences of the United States of America* **1970**, *67*, 1839–45.
15. Radany, E. H.; Friedberg, E. C. A Pyrimidine Dimer-DNA Glycosylase Activity Associated with the v Gene Product of Bacterophage T4. *Nature* **1980**, *286*, 182–5.
16. Yanisch-Perron, C.; Vieira, J.; Messing, J. Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. *Gene* **1985**, *33*, 103–19.
17. Fujikane, R.; Ishino, S.; Ishino, Y.; Forterre, P. Genetic Analysis of DNA Repair in the Hyperthermophilic Archaeon, Thermococcus Kodakaraensis. *Genes Genet Syst* **2010**, *85*, 243–257.
18. Cleaver, J. E. Opinion: Cancer in Xeroderma Pigmentosum and Related Disorders of DNA Repair. *Nature Reviews Cancer* **2005**, *5*, 564–573.
19. Cleaver, J. E.; Lam, E. T.; Revet, I. Disorders of Nucleotide Excision Repair: The Genetic and Molecular Basis of Heterogeneity. *Nature reviews. Genetics* **2009**, *10*, 756–68.
20. Selby, C. P.; Sancar, A. Molecular Mechanism of Transcription-Repair Coupling. *Science (New York, N.Y.)* **1993**, *260*, 53–8.
21. Makarova, K. S.; Koonin, E. V. Archaeology of Eukaryotic DNA Replication. *Cold Spring*

Harbor Perspectives in Biology **2013**, *5*, a012963–a012963.

22. Hawkins, M.; Malla, S.; Blythe, M. J.; Nieduszynski, C. A.; Allers, T. Accelerated Growth in the Absence of DNA Replication Origins. *Nature* **2013**, *503*, 544–547.
23. Michel, B.; Boubakri, H.; Baharoglu, Z.; LeMasson, M.; Lestini, R. Recombination Proteins and Rescue of Arrested Replication Forks. *DNA Repair* **2007**, *6*, 967–980.
24. Jager, D.; Forstner, K. U.; Sharma, C. M.; Santangelo, T. J.; Reeve, J. N. Primary Transcriptome Map of the Hyperthermophilic Archaeon Thermococcus Kodakarensis. *BMC Genomics* **2014**, *15*, 684.
25. Lindås, A.-C.; Bernander, R. The Cell Cycle of Archaea. *Nature Reviews Microbiology* **2013**, *11*, 627–638.
26. Bernander, R. Chromosome Replication, Nucleoid Segregation and Cell Division in Archaea. *Trends in microbiology* **2000**, *8*, 278–83.
27. Spaans, S. K.; Oost, J. van der; Kengen, S. W. M. The Chromosome Copy Number of the Hyperthermophilic Archaeon Thermococcus Kodakarensis KOD1. *Extremophiles : life under extreme conditions* **2015**, *19*, 741–50.
28. Morikawa, M.; Izawa, Y.; Rashid, N.; Hoaki, T.; Imanaka, T. Purification and Characterization of a Thermostable Thiol Protease from a Newly Isolated Hyperthermophilic Pyrococcus Sp. *Applied and environmental microbiology* **1994**, *60*, 4559–66.
29. Atomi, H.; Fukui, T.; Kanai, T.; Morikawa, M.; Imanaka, T. Description of Thermococcus Kodakaraensis Sp. Nov., a Well Studied Hyperthermophilic Archaeon Previously Reported as Pyrococcus Sp. KOD1. *Archaea (Vancouver, B.C.)* **2004**, *1*, 263–7.
30. Hidese, R.; Nishikawa, R.; Gao, L.; Katano, M.; Imai, T.; Kato, S.; Kanai, T.; Atomi, H.; Imanaka, T.; Fujiwara, S. Different Roles of Two Transcription Factor B Proteins in the Hyperthermophilic Archaeon Thermococcus Kodakarensis. *Extremophiles : life under extreme conditions* **2014**, *18*, 573–88.

31. Briegel, A.; Oikonomou, C. M.; Chang, Y.; Kjær, A.; Huang, A. N.; Kim, K. W.; Ghosal, D.; Nguyen, H. H.; Kenny, D.; Ogorzalek Loo, R. R.; Gunsalus, R. P.; Jensen, G. J. Morphology of the Archaellar Motor and Associated Cytoplasmic Cone in *Thermococcus Kodakaraensis*. *EMBO reports* **2017**, *18*, 1660–1670.

APPENDIX 1

MANIPULATING ARCHAEAL SYSTEMS TO PERMIT ANALYSES OF TRANSCRIPTION ELONGATION-TERMINATION DECISIONS *IN VITRO*[4]

A1.1 INTRODUCTION

Processive transcription elongation adumbrates, and research findings support that, the decision to terminate transcript synthesis is tightly regulated[1–20]. It is therefore imperative that methods are available to discriminate delayed synthesis (paused, backtracked, or arrested complexes) from *bona vide* transcript release signaling transcription termination. Release of an RNA transcript from a fully formed elongation complex is the definition of transcription termination[21]. This definition discriminates true termination from abortive initiation wherein short transcripts are often repeatedly released from complexes transitioning from initiation to elongation[22–25]. Release of the nascent RNA may or may not occur simultaneously with recycling of RNAP from the DNA for another round of transcription[26].

Studies of elongation and termination often require positioning elongation complexes at discrete template positions *in vitro*. By non-covalently linking template DNAs to a solid – and often magnetic – support, stable transcription complexes can be generated by limiting the NTP substrates provided to RNAP for synthesis[27,28]. Transcription of a template attached to a solid-

[4] This appendix was previously published as a methods article with the same title, "Manipulating Archaeal Systems to Permit Analyses of Transcription Elongation-Termination Decisions *In Vitro*", in January 2015.

TJS and I conceived the content and co-wrote the manuscript.

Gehring, A. M.; Santangelo, T. J. Manipulating Archaeal Systems to Permint Analyses of Transcription Elongation-Termination Deciscions In Vitro. *Methods in Molecular Biology* **2015**, 1276, 263-279. PMID: 25665569.

support permits walking RNAP to discrete positions and also provides a simple technique to distinguish those transcripts associated with stalled but stable elongation complexes from true termination events.

Herein we describe methods to expand this technology to the study of archaeal RNAPs. The simplified transcription system employed by all archaea most closely mimics the eukaryotic RNA polymerase II (Pol II) system rather than the bacterial or eukaryotic Pol I or Pol III machinery[29–36]. Archaeal transcription systems necessitate that at least two additional considerations must be addressed. First, essentially all current *in vitro* transcription systems with archaeal components are derived from hyperthermophiles[37–42], and transcription at high temperatures often requires two compatible solid-support matrixes to facilitate multiple rounds of walking and to distinguish true termination events. Secondly, the archaeal transcription apparatus is not sensitive to commonly used RNAP inhibitors (i.e. rifampicin or α-amanitin)[43], and thus transcription of a template linked to a solid support is often necessary to obtain a single elongation complex per DNA template to analyze single-round transcription elongation and termination *in vitro.* The methods presented here detail the promoter directed *in vitro* transcription derived from *Thermococcus kodakarensis* and the steps necessary to generate stalled elongation complexes and monitor transcription termination.

A1.2 Materials

Bacterial cell growth and purification of recombinant TBP, TFB1, and TFB2.

1. LB media (10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl). 1 L autoclave sterilized within a 4 L baffled Erlenmeyer flask.
2. Kanamycin
3. Chloramphenicol
4. Rosetta 2 (DE3) cells (Novagen) (see note 1).
5. Isopropyl-β-D-1 thiogalactopyranoside (IPTG)
6. Sorbitol

7. Lysozyme

8. TEN buffers: 20 mM Tris-HCl pH 8, 0.1 mM EDTA, x mM NaCl. Buffers are named based on [NaCl]. For example, TEN-100 represents a solution with 100 mM NaCl.

9. Protein storage buffer: 20 mM Tris–HCl pH 8, 0.1 mM EDTA, 100 mM NaCl, 50% glycerol, 5 mM β-ME.

10. 6X SDS-loading buffer (0.375M Tris pH 6.8, 12% SDS, 60% glycerol, 0.6M DTT, 0.06% bromophenol blue)

11. 5X SDS-running buffer (94 g glycine, 15 g Tris-base, and 5 g sodium dodecyl sulfate (SDS) per liter).

12. Dialysis tubing (10 kD and 100 kD molecular weight cut-off).

Thermococcus cell growth and purification of RNAP

1. Anaerobic chamber (Coy Laboratories) (see note 2).

2. Artificial sea-water medium supplemented with 0.5% (w/v) tryptone, 0.5% (w/v) yeast extract, 1X trace mineral solution, and 1X vitamin mixture. 1X artificial sea-water contains 20 g NaCl, 3 g $MgCl_2 \cdot 6H2O$, 6 g $MgSO_4 \cdot 7H_2O$, 1 g $(NH4)_2SO_4$, 200 mg $NaHCO_3$, 300 mg $CaCl_2 \cdot 2H_2O$, 0.5 g KCl, 420 mg KH_2PO_4, 50 mg NaBr, 20 mg $SrCl_2 \cdot 6H2O$, and 10 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ sulfate per L .

3. Trace mineral solution (1000X): 0.5 g $MnSO_4 \cdot H_2O$, 0.1 g $CoCl_2 \cdot 6 H_2O$, 0.1 g $ZnSO_4 \cdot 7 H_2O$, 0.01 g $CuSO_4 \cdot 5 H_2O$, 0.01 g $AlK(SO_4)_2 \cdot 12 H_2O$, 0.01 g H_3BO_3, 0.01 g $Na_2MoO_4 \cdot 2 H_2O$ per L.

4. Vitamin mixture (200X): 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-amionbenzoic acid, 0.2 g thiamine, 0.2 g riboflavin, 0.2 g pyridoxine, 0.2 g cobalmin per L.

5. Elemental sulfur (flowers of sulfur).

In vitro transcription

1. rNTPs (see note 3).

2. 10X Transcription Buffer: 155 mM Tris-HCl pH 8.0, 27.5 mM $MgCl_2$.

3. SA Buffer: 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mg/mL bovine serum albumin (BSA).

4. Wash Buffer: 20 mM Tris-HCl, 0.1 mM EDTA, 250 mM KCl, 4 mM $MgCl_2$, 20 µg/mL BSA.

5. 1.2X Stop buffer: 0.6 M Tris-HCl pH 8.0, 12 mM EDTA.

6. 1.0X Stop buffer: 0.5 M Tris-HCl pH 8.0, 10 mM EDTA.

7. Formamide loading buffer: 95% formamide, 1X TBE containing 0.1% bromophenol blue and 0.1% xylene cyanol.

8. 1M KCl.

9. Streptavidin coated magnetic beads (We use Streptavidin MagneSphere Paramagnetic Particles from Promega).

10. Anti-digoxigenen magnetic beads (We use Anti-digoxigenen magnetic particles from Roche).

11. 25:24:1 Phenol : Chloroform : Isoamyl alcohol mixture; Tris-saturated at pH 8.0.

12. Dinucleotide (ApC; see note 3).

13. OmniFlex 200 µl Gel-Load pipette tips (Life Science Products).

14. ^{32}P-□-ATP.

15. T4 polynulceotide kinase (We use PNK purchased from New England Biolabs).

16. 10X T4 PNK buffer (We use the buffer supplied from New England Biolabs).

Additional instrumentation

1. Top heated thermocycler (see note 4).

2. 40 cm vertical gel electrophoresis apparatus and an appropriate power supply.

3. Sonicator and appropriate horn.

4. Large capacity high-speed centrifuge. We use an Avanti-J26 XP1 equipped with JA25.50, JLA10.5, and J20/1 rotors; Beckman Coulter.

5. Liquid chromatography system equipped with fraction collector. We use an AKTA platform; GE Healthcare.

6. Chromatography columns, including Hi Trap Q-FF, Superdex 200 16/60, Hi Trap SP-FF, Mono Q, and Hi Trap Chelating. We obtain all of our columns from GE Healthcare, although any high quality columns with similar matrixes will suffice.

7. Centrifugal concentrators (10, 30, and 100 kD molecular weight cut-off; We use units from Ambion).

8. Discontinuous SDS-PAGE system. We generate all gels in-house and resolve gels using a Minimax protein apparatus; Aquebogue Machine Shop.

9. Coomassie stain. (1 g coomassie per liter; 50% methanol, 10% acetic acid, 40% H_2O.

10. Magnetic separation stand (Although any magnet will suffice, we typically rely on commercial stands from Promega).

11. pH and conductivity meter.

A1.3 METHODS

Transcription using archaeal components requires only three protein complexes for accurate promoter initiated transcription, namely RNAP, TATA-binding protein (TBP), and Transcription Factor B (TFB)[37,44]. Procedures for the purification of the necessary protein complexes are provided first, followed by details regarding the formation of elongation complexes. Separation of intact transcription elongation complexes from RNAs released into solution is then detailed.

Purification of recombinant archaeal TBP

1. Grow Rosetta2 (DE3) cells carrying a pET30b-derived vector expressing recombinant TBP (TK0132) from *T. kodakarensis* in LB supplemented with 34 µg/ml chloramphenicol and 40 µg/ml kanamycin to an A_{600} of 0.4 at 37˚C. Induce expression by addition of IPTG to a final concentration of 0.25 mM and sorbitol to final concentration of 1% (w/v) (see note 5). Allow 12 hours at 22˚C for induction. (see note 6).

2. Harvest biomass (5,000 x *g*), and discard supernatant. Resuspend biomass (3 ml per gram) in TEN-0 with 0.2 mg lysozyme/ml and lyse by repeated sonication (see note 7).

3. Clarify mixture by centrifugation (10,000 x *g*). Discard pelleted debris.

4. Load the clarified supernatant over a 5 ml Hi-Trap Q-FF column (pre-equilibrated with TEN-100) attached to a chromatography system. Discard flowthrough.

5. Flush the column with minimally 20 column volumes TEN-100 (continue to flush with TEN-100 if substantial protein elution can be detected by monitoring the UV absorbance of the eluent). Elute bound proteins, including TBP, with a ~30 column volume linear gradient of NaCl from TEN-100 to TEN-2000.

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

7. Fractions containing TBP (molecular weight ~ 21.5 Kd) should be identified by SDS-PAGE and coomassie staining.

8. Fractions containing TBP should be pooled and concentrated (at 4˚C) to < 1 ml using 10 kDA molecular-weight cut-off centrifugal concentrators following the manufacturers recommendations.

9. Concentrated fractions containing TBP should be loaded onto and resolved through a Superdex 200 16/60 column equilibrated with TEN-500 at 0.5 ml/min. Collect 1-2 ml fractions during the elution.

10. Fractions containing TBP should be identified by SDS-PAGE and coomassie staining (as in sections 3.1.6. and 3.1.7.) These fractions should be pooled and concentrated (as in section 3.1.7.).

11. Concentrated fractions containing TBP should be reapplied onto a 5 ml HiTrap Q-FF column pre-equilibrated with TEN-200. The column should be washed with TEN-200 until no changes in UV absorbance are noted in the eluent. Bound proteins are eluted with a linear ~30 CV gradient from TEN-200 to TEN-900.

12. Fractions containing TBP should be identified by SDS-PAGE and coomassie staining. These fractions should be pooled and concentrated (if necessary). Secure samples in 10kD dialysis tubing and dialyze samples (twice, for at least 6 hours) against minimally 1000-volumes protein storage buffer. Recover purified protein from the dialysis tube and store at -80˚C.

Purification of recombinant archaeal TFB

T. kodakarensis encodes two isoforms of transcription factor B [(TFB); TFB1 = TK1280; TFB2 = TK2287], both of which are full functional *in vivo* and *in vitro* for all transcription reactions[38]. Details are provided that are generally applicable for purification of either TFB isoform tagged with an N-terminal His_6-sequence.

1. Grow Rosetta2 (DE3) cells carrying a pET28a-derived vector expressing a recombinant TFB from *T. kodakarensis* in LB supplemented with 34 µg/ml chloramphenicol and 40 µg/ml kanamycin to an A_{600} of 0.5 at 37˚C. Induce expression by addition of IPTG to a final concentration of 0.25 mM and sorbitol to final concentration of 1% (w/v). Allow 24 hours at 22˚C for induction (see note 8).

2. Harvest biomass (5,000 x *g*) and discard supernatant. Resuspend biomass in TEN-100 (3 ml per gram) containing 0.2 mg lysozyme/ml and lyse by repeated sonication.

3. Clarify mixture by centrifugation (10,000 x *g*). Discard pelleted debris.

4. Load the clarified supernatant over a 5 ml Hi-Trap SP-FF column (pre-equilibrated with TNE-100) attached to a chromatography system. Discard flowthrough.

5. Flush the column with minimally 20 CV TEN-100 (continue to flush with TEN-100 if substantial protein elution can be detected by monitoring the UV absorbance of the eluent). Elute the bound proteins, including TFB, with a ~30 CV linear gradient of TEN-100 to TEN-2000.

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

7. Fractions containing TFB (molecular weight ~34 Kd) should be identified by SDS-PAGE and coomassie staining.

8. Fractions containing TFB should be pooled, and loaded directly onto a 5 ml Ni^{2+}-charged chelating column (see note 9). After extensive washing with 20 ml Tris-HCl pH 8.0 containing 500 mM NaCl and 5 mM imidazole, bound proteins should be eluted with a linear gradient of 500 mM NaCl and 5 mM imidazole to 100 mM NaCl and 500 mM imidazole in 20 mM Tris HCl pH 8.0.

9. Fractions containing TFB should be identified by SDS-PAGE and coomassie staining. These fractions should be pooled and concentrated (if necessary). Secure samples in 10kD dialysis tubing and dialyze samples (twice, for at least 6 hours) against minimally 1000-volumes protein storage buffer. Recover purified protein from the dialysis tube and store at -80˚C.

Purification of archaeal RNAP

Many archaeal RNAPs contain Fe-S centers[45,46], however, there is no evidence for such clusters in *T. kodakarensis* RNAP and the enzyme can be purified aerobically without concern. Strains wherein a gene encoding a single subunit of RNA polymerase is modified to encode a protein with a His_6-tag are now routinely used for purification of RNAP[38].

1. Prepare sterile artificial sea water media supplemented with yeast extract, tryptone, vitamins, and trace minerals within a Coy anaerobic chamber (see note 2). Add sulfur to 2 g per L. Media should be inoculated with an appropriate *T. kodakarensis* culture (1:100) and be placed at 85˚C until growth reaches mid-exponential phase (A_{600} ~ 0.6).

2. Harvest biomass (aerobically, 5000 x *g*; see note 10) and discard supernatant. Resuspend biomass (3 ml per gram) in 25 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol and lyse by repeated freeze-thawing employing liquid N_2 (see note 11).

3. Clarify mixture by centrifugation (15,000 x *g*). Discard pelleted debris.

4. Load the clarified supernatant over a 5 ml Ni^{2+}-charged Hi-Trap chelating column (pre-equilibrated with 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol) attached to a chromatography system. Discard flowthrough.

5. Flush the column with minimally 20 CV 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol (continue to flush with 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol if substantial protein elution can be detected by monitoring the UV absorbance of the eluent). Elute the bound proteins, including RNAP, with a ~30 CV linear gradient from 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol to 25 ml Tris-HCl pH 8.0 containing 0.1M NaCl, 10% (v/v) glycerol, and 100 mM imidazole.

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

7. Fractions containing RNAP (molecular weight ~380000 Kd; 12 subunits) should be identified by SDS-PAGE and coomassie staining. The two largest subunits of RNAP run as an easily identifiable doublet near the top of the gel.

8. Fractions containing RNAP should be pooled and concentrated to < 1 ml using 100 kDA molecular-weight cut-off centrifugal concentrators.

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

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

11. Fractions containing RNAP (molecular weight ~380000 Kd; 12 subunits) should be identified by SDS-PAGE and coomassie staining.

12. Fractions containing RNAP should be pooled and concentrated to < 1 ml using 100 kDA molecular-weight cut-off centrifugal concentrators. Concentrated material should be loaded onto and resolved through a Superdex 200 16/60 column equilibrated with 20 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 100 mM NaCl at 0.5 ml/min.

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

14. Fractions containing RNAP (molecular weight ~380000 Kd; 12 subunits) should be identified by SDS-PAGE and coomassie staining. Pool appropriate fractions.

15. Pooled fractions containing RNAP should be reapplied to a 5 ml Ni^{2+}-charged HiTrap chelating column (pre-equilibrated with 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol) attached to a chromatography system. Discard flowthrough.
Flush the column with minimally 20 CV 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol (continue to flush with 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol if substantial protein elution can be detected by monitoring the UV absorbance of the eluent). Elute the bound proteins, including RNAP, with a ~30 CV linear gradient from 25 ml Tris-HCl pH 8.0 containing 1 M NaCl and 10% (v/v) glycerol to 25 ml Tris-HCl pH 8.0 containing 0.1M NaCl, 10% (v/v) glycerol, and 100 mM imidazole.

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

17. Fractions containing RNAP should be identified by SDS-PAGE and coomassie staining. These fractions should be pooled and concentrated (if necessary). Secure samples in 100kD dialysis tubing and dialyze samples (twice, for at least 6 hours) against minimally 1000-volumes protein storage buffer. Recover purified protein from the dialysis tube and store at -80˚C.

Design of DNA template

We employ DNA templates containing the modified promoter from *Methanothermobacter sp.* used to initiate transcription of histone B (5′ - GCGATATATTTATATAGGGATATAGTAATAGATAATATC**A** - 3′)[47]. This short promoter sequence (P_{hmtB}) contains both BRE (underlined) and TATA box (double underlined) sequences to aid transcription factor binding. The transcription start site (bold) is defined and uniform, and > 70% of template DNAs contain elongation complexes when initiation uses equimolar RNAP and templates concentrations; TBP and TFB are provided in 4-fold molar excess[38]. The sequence downstream from the transcription initiation site can be designed to suit the particular needs of an experiment; for illustrative purposes, we will detail elongation on a DNA template permitting elongation to +116 with the initially transcribed sequence of +1-ACGGTAACCGG (Figure A1.1).

It is important to note that the DNA templates employed in transcription assays can be any length, and can include the use of even supercoiled plasmid templates. We typically employ templates wherein the 5′ end of the non-template strand of DNA is biotinylated or digoxigenin-labeled such that elongation complexes can be captured and washed during the experiment. For experiments requiring multiple walking steps interspersed with heating steps, the template strand should also be labeled at the 5’ end with either digoxigenin or biotin; note that on DNA templates where both 5’ bases are modified, one strand should be labeled with biotin and the complement with digoxigenin. For all of our DNAs used in transcription reactions, non-template sequences are fully complementary to template strand sequences. Double stranded templates can be generated via PCR (with or without 5’-modified primers), or each strand of the DNA can be chemical synthesized and oligonucleotides paired to generate double stranded templates for transcription.

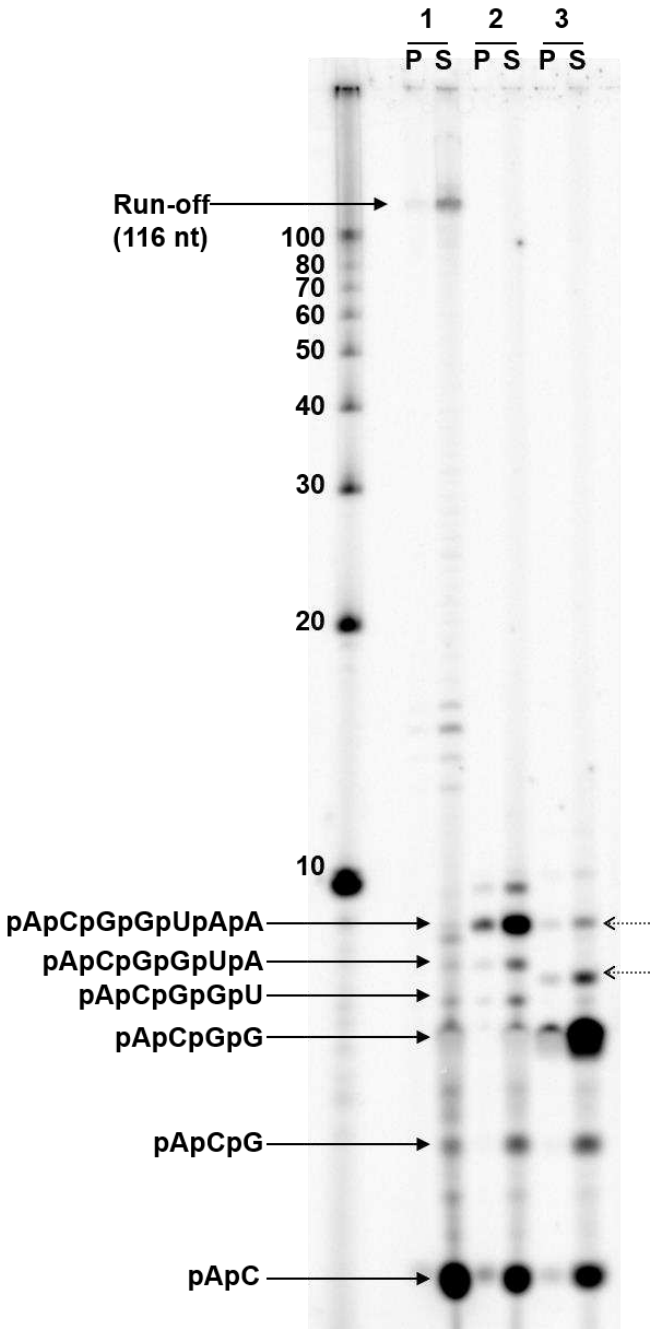

Figure A1.1. Initial transcription and separation of nascent- versus released-transcripts. Denaturing electrophoresis of 5′-ApC dinucleotide radiolabeled RNAs generated from in vitro transcription reactions with archaeal components permits identification of specific transcripts (identified with solid arrows on the left). Two misincorporation products are visible in reaction #3 (identified with dashed arrows on the right). Reactions were separated into pellet (P) and supernatant (S) fractions to identify transcripts associated with the transcription apparatus (pellet fraction) from those transcripts released to solution (supernatant fraction). +116 nt transcripts represent run-off products. Radiolabeled ssDNA 10-bp markers (far left lane) serve as approximate size standards. Reactions resolved in lanes 1–3 each contain 5′- 32 P-ApC and were supplemented with 200 μM: (1) ATP, GTP, CTP, andUTP; (2) only ATP, GTP, and UTP; (3) only GTP.

Labeling dinucleotide primers with ^{32}P

1. For each transcription reaction, combine 1.5 µmole ApC dinucleotide with ~50µC ^{32}P-γ-ATP, in a 5 µl reaction containing 1X T4 PNK buffer and 10 U T4 PNK.

2. Incubate the reaction at 37˚C for 1 hour.

3. Incubate the reaction at 85˚C for 10 minutes to inactivate T4 PNK.

Transcription initiation with subsets of NTPs permits walking of RNAP

1. Combine a double stranded DNA template (10 nM final) with 40 nM RNAP, 80 nM TFB, 80nM TBP, in 1X Transcription buffer supplemented with 250 mM KCl, 5 mM DTT and 5 mM $MgCl_2$. Add the entirety of the reaction wherein the dinucleotide was labeled (from section 3.4.). Reactions are typically prepared in sterile, 0.6 ml plastic tubes with flat caps that can be fitted into a thermocycler block equipped with a heated lid. This mixture is heated to 85˚C for 5 minutes to allow formation of promoter-bound, open complex (see note 12).

2. Add a subset of NTPs (each NTP to [200 µM] final, dissolved in 1X transcription buffer, 250 mM KCl, 5 mM $MgCl_2$, 5 mM DTT) that allows for elongation to a discrete, predetermined position in the sequence is added to the reaction. As one example, addition of GTP alone to a reaction employing a DNA substrate with the non-template sequence (+1-ACGGTAACC) permits elongation to generate a RNA that is +4 nucleotides in length. Addition of just ATP, GTP, and UTP would permit elongation to +7, but transcription elongation complexes (TECs) containing 8 or fewer nucleotides are generally not stable and typically do not survive washing steps. Formation of a promoter-proximal stalled complex sterically inhibits the formation of a second TEC on the same template, thus limiting transcription initiation to a single-round. Washing the complexes (see below), removes excess RNAP, TBP, and TFB, eliminating any future TEC formation.

3. The reaction is placed on ice and transferred to pre-equilibrated dry streptavidin coated paramagnetic particles (see note 13). The streptavidin coated paramagnetic particles and the TECs formed on the biotinylated DNA substrates are allowed to incubate on ice for 10 minutes.

4. Using the Magnetic Separation Stand, the magnetic particles and the bound TECs are washed 3 times using 100µL of wash buffer. These washes remove essentially all unincorporated NTPs. The particles are then resuspened in 1X transcription buffer, supplemented with 5mM DTT, and 250mM KCl.

5. A new subset of NTPs can be added to allow walking to a different position and this process can be repeated until all desired walking steps are performed (see note 14). To continue the example from above, TECs containing 7 nucleotide RNAs can be supplemented with CTP to permit continued elongation to +9, or all NTPs to generate TECs with RNAs that are +116 nucleotides long.

Stopping of reaction and preparation of samples

1. Samples are typically removed in 20 µl volumes and added to 100 µl 1.2x stop buffer or 120 µl of 1.0x stop buffer if TECs bound to particles are not resuspended after washing. Additionally, 8 µg of carrier tRNA is added to each sample.

2. A volume of phenol:chloroform:isoamyl alcohol equal to the total volume of the stopped reaction is added, and each reaction is mixed extensively.

3. The extracted samples are centrifuged (3 minutes, 14,000 x *g*) to separate the aqueous and organic phases. The aqueous layer is removed to a clean tube containing 2.6 volumes of 100% ethanol. The reactions are vigorously mixed, quickly centrifuged to collect material from the side of the tubes, then placed at -20˚C for minimally 1 hour (see note 15).

4. The cold samples are centrifuged in a refrigerated bench top centrifuge (4˚C, 14,000 x *g*) for 30 minutes. A small visible pellet should be seen at the end of the centrifugation.

5. The ethanol is then removed from the samples and appropriately discarded

6. Each pellet is then resuspended in 4µL formamide loading buffer. Samples are heated to 99˚C for 3 minutes, then immediately transferred to ice prior to loading on denaturing gels.

Resolving of samples

The samples are resolved in a 40cm polyacrylamide gel on an aluminum backed gel apparatus.

1. Each total sample (4 µL) is transferred via OmniFlex tips into unique wells in the gel.

3. The gel is then immediately resolved by applying ~1500-2500V until desired resolution is achieved. For separation of transcripts differing by a few or even a single nucleotide, we resolve samples until the bromophenol blue migrates at least 30 cm in ≤ 15% denaturing gels.

Analysis

When the gel is sufficiently resolved, the apparatus is broken down and the glass plates containing the gel are separated. The specific activity of the nascent transcripts is typically sufficient such that drying of the gels is unnecessary. The gel is covered with a single layer of plastic-wrap and is exposed to a phosphoimager screen. Use of a radiolabeled dinucleotide ensures that all transcripts, regardless of length, have the same specific activity permitting immediate and easy quantification of molar ratios of different transcripts present in the reactions.

A1.4 NOTES

1: Archaeal and bacterial genomes use the same genetic code, but substantial differences in codon bias between the two domains generally necessitates use of an *E.coli* strain wherein all rare tRNAs are overexpressed to facilitate high-level translation of archaeal transcripts for recombinant protein production.

2: *T. kodakarensis* is an obligate anaerobe. Facilities for anaerobic microbiology are critical for the proper passage and growth of *Thermococcus* strains. We use an anaerobic chamber

manufactured by Coy Laboratories, although any unit capable of maintaining an anaerobic environment would be suitable. All media preparations should be carried out within the chamber, sealed within the chamber, and then removed from the chamber for sterilization. Extreme caution is warranted when autoclaving sealed media bottles. 1 L of media is typically prepared in 2 L pyrex bottles with high temperature closures and anaerobic septums.

3: When employing subsets of NTPs to allow elongation to a specific position, it is critical that the purity of the NTPs be extremely high. The quality and purity of NTP preparations from suppliers differs significantly. We routinely use NTPs purchased from GE Healthcare, although any sufficiently pure NTPs will suffice. Transcription initiation with a dinucleotide is often preferable to initiation using two NTPs, as initiation is more uniform and the dinucleotide permits radiolabeling of the dinucleotide with ^{32}P-□-ATP. Use of radiolabeled dinucleotides results in transcripts with a single radiolabel and facilitates rapid quantification of RNA products on a molar ratio.

4: Small volume transcription reactions (< 100 µl) heated to 85˚C in a traditional wet- or dry-bath suffer too great a volumetric loss due to evaporation to control reaction conditions for reproducible results. We rely on top-heated thermocyclers to limit evaporative loss. The reaction tubes used for the experimentation should be sized appropriately to fit into the thermocycler and allow the top of the thermocycler to close tightly. We utilize 0.65 ml tubes with flat caps.

5: The concentrations of IPTG used here have been empirically determined to maximize protein expression. Alternate expression constructs are likely to require different conditions. The addition of sorbitol aids in expression and limits degradation of recombinant proteins.

6: TBP expression peaks at 12 hours post-induction and solubility is greater when cultures are shifted to 22˚C. Purification is possible from cultures maintained at 37˚C and from cultures induced for as short as 1 hour.

7: Sonication is typically completed with a series of 10 second pulses, on ice, with 30-40 seconds between pulses to limit heating the sample. Sonication is continued until the mixture has a viscosity approximately equal to water.

8: Expression of each TFB peaks at least 24 hours after induction. Lowering the culture temperature to 22˚C aids in total yield and solubility of each TFB isoform.

9: Many Ni^{2+}-affinity matrixes are available. We routinely use pre-packed chelating columns that we charge with $NiSO_4$ prior to use. Columns are flushed, in order, with 5 CV H_2O, 1 CV 0.5M EDTA (to remove all bound metals), 5 CV H_2O, 1 CV 0.1 M $NiSO_4$ (to charge the column with Ni^{2+}), and 5 CV H_2O before each use.

10: The metabolism of *T. kodakarensis* typically uses elemental sulfur as the terminal electron acceptor, thus generating copious amount of poisonous H_2S gas and building substantial pressures within the sealed growth vessel. Extreme care should be taken to properly vent the pressurized gases prior to harvesting biomass. All spent media should be disposed within a fume hood.

11: Biomass is flash frozen by immersion in liquid N_2 then immediately thawed by incubation at 85˚C. Repeated rounds of freezing and thawing effectively lyses *T. kodakarensis* cultures.

12: These procedures have been optimized at the normal growth temperature of *Thermococcus kodakarensis*. Transcription initiation and elongation are possible at lower temperatures. Note that sequences that direct intrinsic termination at physiological temperature only function to direct RNAP to pause at reduced temperatures[6], and thus it is critical to evaluate elongation-termination decisions at the physiological temperature

13: Streptavidin Magnasphere Paramagnetic Particles must be equilibrated prior to use. The particles are equilibrated by washing 10µL of particles (1 mg/mL) with 100 µL of SA buffer three times. To separate the particles from the buffer, a MagneSphere Technology Magnetic Separation Stand (Promega) is used. The buffer is removed from the particles prior to the

reaction being added to the particles. For each reaction that particles will be used, 10µL of particles must be equilibrated.

The streptavidin molecules denature at temperatures above 55˚C, and thus we cool complexes only to ensure capture to the solid support. Complexes must be returned to physiological temperatures before elongation-termination decisions can be properly monitored following NTP addition.

DNAs containing digoxigenin moieties can be employed when coupled with solid-supports facilitating capture of digoxigenin-labeled DNAs. We often employ templates with separate digoxigenin and biotin moieties at the 5' ends of the substrates.

14: The temperature constraints of the archaeal transcription apparatus necessitate the addition of unheated streptavidin-coated particles after each elongation cycle at high temperature. Elongation at room temperature is possible, but is facilitated by incubation at 85˚C. Room temperature elongation is utilized when the physiological state of the TEC is not under study during the elongation, however elongation at 85˚ permits analysis at the optimal reaction conditions.

15: We rely on a high-concentration of Tris to aid in the precipitation of *in vitro* synthesized RNAs. Use of tRNA and Tris as carriers results in samples that do not contain high concentrations of NaCl or KCl that typically lower the resolution and clarity of samples during electrophoresis. It is critical to extract and precipitate the RNAs away from RNAP, as RNAP will non-specifically associate with transcripts and retard their migration during electrophoresis.

REFERENCES

1. Yarnell, W. S.; Roberts, J. W. Mechanism of Intrinsic Transcription Termination and Antitermination. *Science* **1999**, *284*, 611–615.
2. Santangelo, T. J.; Roberts, J. W. Forward Translocation Is the Natural Pathway of RNA Release at an Intrinsic Terminator. *Mol Cell* **2004**, *14*, 117–126.
3. Santangelo, T. J.; Artsimovitch, I. Termination and Antitermination: RNA Polymerase Runs a Stop Sign. *Nat Rev Microbiol* **2011**, *9*, 319–329.
4. Toulokhonov, I.; Artsimovitch, I.; Landick, R. Allosteric Control of RNA Polymerase by a Site That Contacts Nascent RNA Hairpins. *Science* **2001**, *292*, 730–733.
5. White, E.; Kamieniarz-Gdula, K.; Dye, M. J.; Proudfoot, N. J. AT-Rich Sequence Elements Promote Nascent Transcript Cleavage Leading to RNA Polymerase II Termination. *Nucleic Acids Res* **2013**, *41*, 1797–1806.
6. Santangelo, T. J.; Reeve, J. N. Archaeal RNA Polymerase Is Sensitive to Intrinsic Termination Directed by Transcribed and Remote Sequences. *J Mol Biol* **2006**, *355*, 196–210.
7. Santangelo, T. J.; Cubonova, L.; Skinner, K. M.; Reeve, J. N. Archaeal Intrinsic Transcription Termination in Vivo. *J Bacteriol* **2009**, *191*, 7102–7108.
8. Hollands, K.; Sevostiyanova, A.; Groisman, E. A. Unusually Long-Lived Pause Required for Regulation of a Rho-Dependent Transcription Terminator. *Proc Natl Acad Sci U S A* **2014**.
9. Guo, J.; Turek, M. E.; Price, D. H. Regulation of RNA Polymerase II Termination by Phosphorylation of Gdown1. *J Biol Chem* **2014**, *289*, 12657–12665.
10. Webb, S.; Hector, R. D.; Kudla, G.; Granneman, S. PAR-CLIP Data Indicate That Nrd1-Nab3-Dependent Transcription Termination Regulates Expression of Hundreds of Protein Coding Genes in Yeast. *Genome Biol* **2014**, *15*, R8.

11. Schulz, D.; Schwalb, B.; Kiesel, A.; Baejen, C.; Torkler, P.; Gagneur, J.; Soeding, J.; Cramer, P. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell* **2013**, *155*, 1075–1087.

12. Contreras, X.; Benkirane, M.; Kiernan, R. Premature Termination of Transcription by RNAP II: The Beginning of the End. *Transcription* **2013**, *4*, 72–76.

13. Peters, J. M.; Mooney, R. A.; Grass, J. A.; Jessen, E. D.; Tran, F.; Landick, R. Rho and NusG Suppress Pervasive Antisense Transcription in Escherichia Coli. *Genes Dev* **2012**, *26*, 2621–2633.

14. Conaway, R. C.; Conaway, J. W. The Mediator Complex and Transcription Elongation. *Biochim Biophys Acta* **2013**, *1829*, 69–75.

15. Wagschal, A.; Rousset, E.; Basavarajaiah, P.; Contreras, X.; Harwig, A.; Laurent-Chabalier, S.; Nakamura, M.; Chen, X.; Zhang, K.; Meziane, O.; Boyer, F.; Parrinello, H.; Berkhout, B.; Terzian, C.; Benkirane, M.; Kiernan, R. Microprocessor, Setx, Xrn2, and Rrp6 Co-Operate to Induce Premature Termination of Transcription by RNAPII. *Cell* **2012**, *150*, 1147–1157.

16. Peters, J. M.; Vangeloff, A. D.; Landick, R. Bacterial Transcription Terminators: The RNA 3'-end Chronicles. *J Mol Biol* **2011**, *412*, 793–813.

17. Mischo, H. E.; Proudfoot, N. J. Disengaging Polymerase: Terminating RNA Polymerase II Transcription in Budding Yeast. *Biochim Biophys Acta* **2013**, *1829*, 174–185.

18. Spitalny, P.; Thomm, M. A Polymerase III-like Reinitiation Mechanism Is Operating in Regulation of Histone Expression in Archaea. *Mol Microbiol* **2008**, *67*, 958–970.

19. Werner, F. Molecular Mechanisms of Transcription Elongation in Archaea. *Chem Rev* **2013**, *113*, 8331–8349.

20. Grohmann, D.; Werner, F. Hold on!: RNA Polymerase Interactions with the Nascent RNA Modulate Transcription Elongation and Termination. *RNA Biol* **2010**, *7*, 310–315.

21. Adhya, S.; Gottesman, M. Control of Transcription Termination. *Annu Rev Biochem* **1978**,

47, 967–996.

22. Goldman, S. R.; Ebright, R. H.; Nickels, B. E. Direct Detection of Abortive RNA Transcripts in Vivo. *Science* **2009**, *324*, 927–928.
23. Hsu, L. M. Promoter Clearance and Escape in Prokaryotes. *Biochim Biophys Acta* **2002**, *1577*, 191–207.
24. Carpousis, A. J.; Gralla, J. D. Cycling of Ribonucleic Acid Polymerase to Produce Oligonucleotides during Initiation in Vitro at the Lac UV5 Promoter. *Biochemistry* **1980**, *19*, 3245–3253.
25. Gralla, J. D.; Carpousis, A. J.; Stefano, J. E. Productive and Abortive Initiation of Transcription in Vitro at the Lac UV5 Promoter. *Biochemistry* **1980**, *19*, 5864–5869.
26. Epshtein, V.; Dutta, D.; Wade, J.; Nudler, E. An Allosteric Mechanism of Rho-Dependent Transcription Termination. *Nature* **2010**, *463*, 245–249.
27. Nudler, E.; Gusarov, I. Analysis of the Intrinsic Transcription Termination Mechanism and Its Control. *Methods Enzymol* **2003**, *371*, 369–382.
28. Nudler, E.; Gusarov, I.; Bar-Nahum, G. Methods of Walking with the RNA Polymerase. *Methods Enzymol* **2003**, *371*, 160–169.
29. Hirata, A.; Murakami, K. S. Archaeal RNA Polymerase. *Curr Opin Struct Biol* **2009**, *19*, 724–731.
30. 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* **2009**, *71*, 989–1002.
31. Langer, D.; Hain, J.; Thuriaux, P.; Zillig, W. Transcription in Archaea: Similarity to That in Eucarya. *Proc Natl Acad Sci U S A* **1995**, *92*, 5768–5772.
32. Werner, F. A Nexus for Gene Expression-Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life. *J Mol Biol* **2012**, *417*, 13–27.
33. Grohmann, D.; Werner, F. Recent Advances in the Understanding of Archaeal

Transcription. *Curr Opin Microbiol* **2011**, *14*, 328–334.

34. Werner, F.; Grohmann, D. Evolution of Multisubunit RNA Polymerases in the Three Domains of Life. *Nat Rev Microbiol* **2011**, *9*, 85–98.

35. Kostrewa, D.; Zeller, M. E.; Armache, K. J.; Seizl, M.; Leike, K.; Thomm, M.; Cramer, P. RNA Polymerase II-TFIIB Structure and Mechanism of Transcription Initiation. *Nature* **2009**, *462*, 323–330.

36. Kusser, A. G.; Bertero, M. G.; Naji, S.; Becker, T.; Thomm, M.; Beckmann, R.; Cramer, P. Structure of an Archaeal RNA Polymerase. *J Mol Biol* **2008**, *376*, 303–307.

37. Santangelo, T. J.; Cubonova, L.; James, C. L.; Reeve, J. N.; 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* **2007**, *367*, 344–357.

38. Thomm, M.; Stetter, K. O. Transcription in Methanogens. Evidence for Specific in Vitro Transcription of the Purified DNA-Dependent RNA Polymerase of Methanococcus Thermolithotrophicus. *Eur J Biochem* **1985**, *149*, 345–351.

39. Hudepohl, U.; Reiter, W. D.; Zillig, W. In Vitro Transcription of Two rRNA Genes of the Archaebacterium Sulfolobus Sp. B12 Indicates a Factor Requirement for Specific Initiation. *Proc Natl Acad Sci U S A* **1990**, *87*, 5851–5855.

40. Hethke, C.; Geerling, A. C.; Hausner, W.; Vos, W. M. de; Thomm, M. A Cell-Free Transcription System for the Hyperthermophilic Archaeon Pyrococcus Furiosus. *Nucleic Acids Res* **1996**, *24*, 2369–2376.

41. Werner, F.; Weinzierl, R. O. A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription. *Mol Cell* **2002**, *10*, 635–646.

42. Qureshi, S. A.; Bell, S. D.; Jackson, S. P. Factor Requirements for Transcription in the Archaeon Sulfolobus Shibatae. *EMBO J* **1997**, *16*, 2927–2936.

43. Madon, J.; Leser, U.; Zillig, W. DNA-Dependent RNA Polymerase from the Extremely Halophilic Archaebacterium Halococcus Morrhuae. *Eur J Biochem* **1983**, *135*, 279–283.

44. Bell, S. D.; Magill, C. P.; Jackson, S. P. Basal and Regulated Transcription in Archaea. *Biochem Soc Trans* **2001**, *29*, 392–395.

45. Hirata, A.; Klein, B. J.; Murakami, K. S. The X-Ray Crystal Structure of RNA Polymerase from Archaea. *Nature* **2008**, *451*, 851–854.

46. Lessner, F. H.; Jennings, M. E.; Hirata, A.; Duin, E. C.; Lessner, D. J. Subunit D of RNA Polymerase from Methanosarcina Acetivorans Contains Two Oxygen-Labile [4Fe-4S] Clusters: Implications for Oxidant-Dependent Regulation of Transcription. *J Biol Chem* **2012**, *287*, 18510–18523.

47. Xie, Y.; Reeve, J. N. In Vitro Transcription Assays Using Components from Methanothermobacter Thermautotrophicus. *Methods Enzymol* **2003**, *370*, 66–72.

APPENDIX 2

MARKERLESS GENE EDITING IN THE HYPERTHERMOPHILIC ARCHAEON *THERMOCOCCUS KODAKARENSIS*[5]

A2.1 INTRODUCTION

Archaea often thrive in seemingly inhospitable and rapidly changing environments. Analyses of archaeal genomes reveal a plethora of metabolic strategies, predict sophisticated and highly interdependent regulatory networks underlying gene expression and reveal many genes whose protein–and increasingly often stable RNA–products lack a defined function. The ability to challenge existing, and define new pathways through genetic manipulation has assisted in deconvoluting archaeal physiology and information processing systems, and has more recently opened archaeal species to synthetic- and systems-level approaches to define intra- and intercellular networks.

Thermococcus kodakarensis is a hyperthermophilic, anaerobic, marine archaeon for which a genetic system has been developed over the last decade[1–6]. The ability to genetically modify *T. kodakarensis* has allowed for the study of individual gene function in metabolism, replication, transcription and translation. Using a recombination based system and both selective and counter-selective markers, individual genes are deleted from the *T. kodakarensis* genome in a markerless manner (Figure A2.1). This markerless deletion strategy allows the consecutive deletion of multiple genes in a single strain using the same strategy for each gene.

---

[5] This appendix is an accepted protocol under the same title with the suggested reference below.

TS and I conceived and wrote the content with input from TJS.

Gehring, A.M., Sanders, T. and Santangelo, T. J. Markerless gene editing in the hyperthermophilic archaeon T*hermococcus kodakarensis*. *Bio-protocol*.

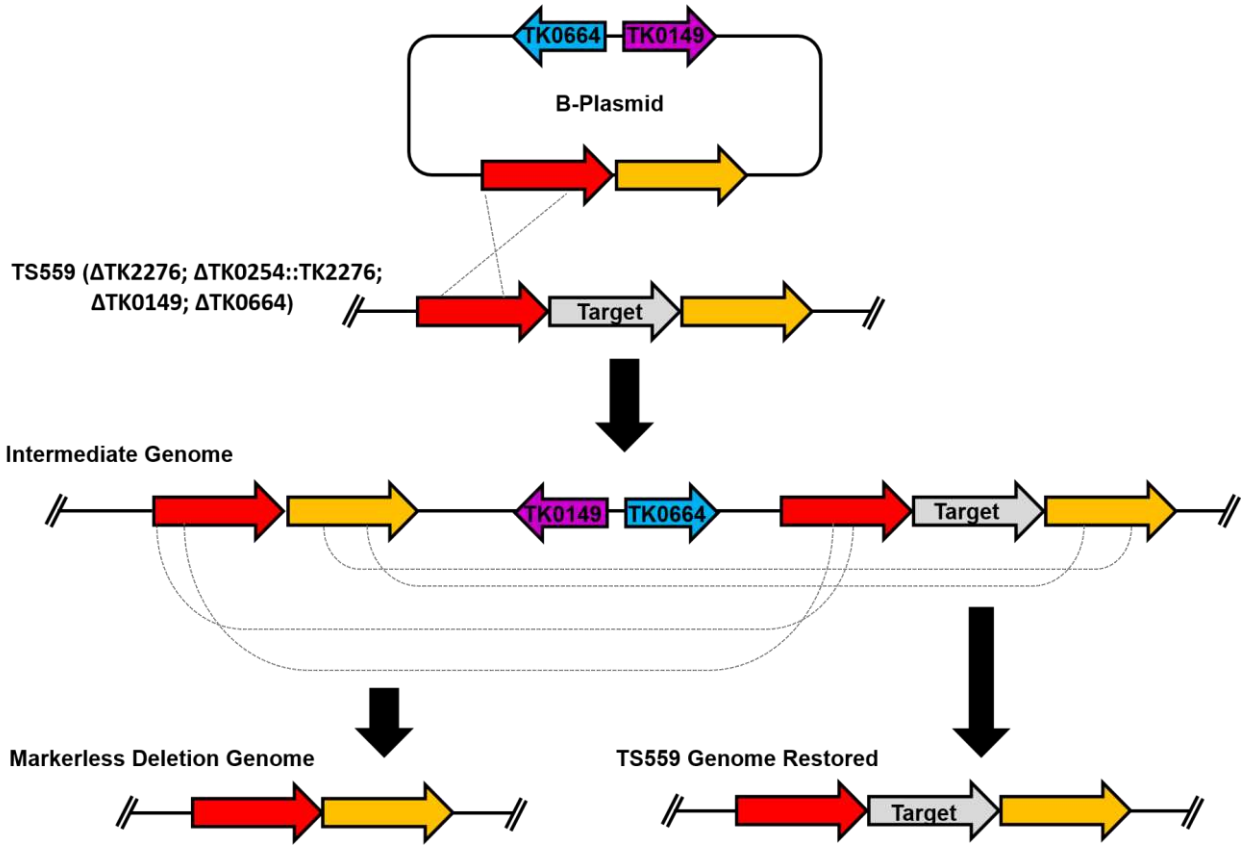

Figure A2.1. Overview of the markerless deletion scheme used in *T. kodakarensis*. At the top of the figure is the B-plasmid used to delete the target gene from the genome. The plasmid recombines into the genome providing agmatine prototrophy to recipient cells and yields an intermediate genome. Two intermediate genomes are possible; however only one is depicted here. A second spontaneous recombination event excises plasmid sequences and permits survival in the presence of cytotoxic 6-MP. This second recombination event will result in the desired deletion genome (left) or the restoration of the TS559 genome (right).

T. kodakarensis strain TS559 (ΔTK2276; ΔTK0254::TK2276; ΔTK0149; ΔTK0664) requires the presence of agmatine and tryptophan for cellular growth[7]. The deletion strategy presented here utilizes the selectable and counter-selectable markers TK0149 and TK0664, respectively. TK0149 encodes a pyruvoyl-dependent arginine decarboxylase, an enzyme necessary in the conversion of arginine to agmatine which is then converted to putrescine. Cells lacking TK0149 are dependent on the addition of agmatine to the media for viability. TK0664 encodes a hypoxanthine guanine phosphoribosyltransferase, an enzyme involved in a ribonucleotide scavenging pathway. Cells encoding TK0664 can metabolize 6-methylpurine (6-MP), a cytotoxic purine derivative, and thus perish in environments containing 6-MP.

To assist others in implementing this technology, here we outline a procedure to delete a gene [as one example, we delete TK0566[8]] from the *T. kodakarensis* TS559 genome.

A2.2 MATERIALS AND REAGENTS

1. 1 ml TB syringe (BD, catalog number: 309624)
2. 1.7 ml microcentrifuge tubes (VWR, catalog number: 490004-444)
3. 0.2 ml PCR tubes (VWR, catalog number: 20170-012)
4. Polystyrene Petri plates (Fisher Scientific, catalog number: S33580A)
5. Split rubber stopper (Wheaton, catalog number: W224100-282)
6. 20 mm aluminum seals (Wheaton, catalog number: 224178-01)
7. 20 mm E-Z Crimper, Standard Seal (Wheaton, catalog number: W225303)
8. 20 mm E-Z Decapper (Wheaton, catalog number: W225353)
9. Polycarbonate centrifuge tubes (Beckman Coulter, catalog number: 361690)
10. Glass Petri plates (VWR, catalog number: 89000-304)

    Note: Glass petri plates are used here, as plastic petri plates will melt at T. kodakarensis incubation temperature (85°C).
11. Cell spreader (Fisher Scientific, catalog number: 08-100-10)
12. 10 ml serum bottles (Wheaton, catalog number: 223739)

13. *T. kodakarensis* strain TS559[7]
14. DH5α *E. coli* competent cells (Fisher Scientific, catalog number: 18258012)
15. XL1-Blue *E. coli* competent cells (Agilent Technologies, catalog number: 200228)
16. 700Forward Primer (5′ CGCCGCAATAGCGGTCGTCGTCATGTTCCC 3′)
17. 700Reverse Primer (5′ AACAATTTCACACAGGAAACAGCTATGACC 3′)
18. pTS700[5]

    Note: Please contact corresponding author to obtain plasmid.
19. 10 mM Tris-HCl pH 8.0 (VWR, catalog number: 97061-258)
20. Isopropanol (Sigma-Aldrich, catalog number: 190764)
21. Chloroform (Sigma-Aldrich, catalog number: C2432)
22. Phenol (VWR, catalog number: 0945)
23. Isoamyl Alcohol (Sigma-Aldrich, catalog number: 1009791000)
24. Phusion DNA polymerase (New England Biolabs, catalog number: M0530)
25. LE Quick dissolve agarose (VWR, catalog number: 490000-004)
26. Ethidium bromide (Sigma-Aldrich, catalog number: E1510)
27. AMPure XP (Beckman Coulter, catalog number: AG3881)
28. Nucleospin Gel and PCR Clean-up Kit (Machery-Nagel, catalog number: 740609)
29. ZR Plasmid Miniprep kit (Zymo Research, catalog number: D4015)
30. SwaI restriction enzyme (New England Biolabs, catalog number: R0604)
31. T4 DNA Polymerase (New England Biolabs, catalog number: M0203)
32. dCTP (Fisher Scientific, catalog number: 10297018)
33. dGTP (Fisher Scientific, catalog number: 10297018)
34. NEBuffer 2.1 (New England Biolabs, catalog number: B7202)
35. Ampicillin (Sigma-Aldrich, catalog number: A0166)
36. Taq DNA Polymerase (New England Biolabs, catalog number: M0267)
37. dNTPs (Fisher Scientific, catalog number: 10297018)

38. Yeast extract (VWR, catalog number: 97063)

    Note: For E. coli media, any yeast extract is suitable, however T. kodakarensis requires this source of yeast extract.

39. Tryptone (EMD Millipore, catalog number: 1072131000)

    Note: T. kodakarensis requires casein peptone that is enzymatically digested using pancreatic enzymes. Other sources of tryptone are suitable for E. coli media.

40. Sodium chloride (NaCl) (Sigma-Aldrich, catalog number: 793566)
41. Quikchange II (Agilent Technologies, catalog number: 200523)
42. Agmatine (Sigma-Aldrich, catalog number: A7127)
43. Elemental sulfur (VWR, catalog number: 101224-218)
44. Gelzan (Sigma-Aldrich, catalog number: G1910)
45. 6-methylpurine (Sigma-Aldrich, catalog number: M1256)
46. Niacin (Sigma-Aldrich, catalog number: PHR1276)
47. Biotin (Amresco, catalog number: 0340)
48. Pantothenate (Sigma-Aldrich, catalog number: 259721)
49. Lipoic acid (Fisher Scientific, catalog number: BP2682)
50. Folic acid (Sigma-Aldrich, catalog number: F7876)
51. P-aminobenzoic acid (Acros Organics, catalog number: 146210010)
52. Thiamine (Fisher Scientific, catalog number: BP892)
53. Riboflavin (Sigma-Aldrich, catalog number: R1706)
54. Pyridoxine (Sigma-Aldrich, catalog number: P9755)
55. Cobalamin (Sigma-Aldrich, catalog number: V6629)
56. Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$) (Sigma-Aldrich, catalog number: M9272)

57. Magnesium sulfate heptahydrate ($MgSO_4 \cdot 7H_2O$) (EMD Millipore, catalog number: MX0070)
58. Ammonium sulfate ($(NH4)_2SO_4$) (VWR, catalog number: BDH9216)
59. Sodium bicarbonate ($NaHCO_3$) (Sigma-Aldrich, catalog number: S6014)
60. Calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$) (EMD Millipore, catalog number: CX0130)
61. Potassium chloride (KCl) (Sigma-Aldrich, catalog number: P3911)
62. Potassium phosphate monobasic (K_2HPO_4) (Sigma-Aldrich, catalog number: P0662)
63. Sodium bromide (NaBr) (Fisher Scientific, catalog number: S255)
64. Strontium chloride hexahydrate ($SrCl_2 \cdot 6H_2O$) (Sigma-Aldrich, catalog number: 13909)
65. Ammonium iron(II) sulfate hexahydrate ($Fe(NH4)_2(SO_4)_2 \cdot 6H_2O$) (Sigma-Aldrich, catalog number: F3754)
66. Manganese(II) Sulfate Monohydrate ($MnSO_4 \cdot H_2O$) (Fisher Scientific, catalog number: M10337)
67. Cobalt (II) chloride hexahydrate ($CoCl_2 \cdot 6H_2O$) (Sigma-Aldrich, catalog number: 202185)
68. Zinc sulfate heptahydrate ($ZnSO_4 \cdot 7H_2O$) (Ricca Chemical Company, catalog number: RDCZ0200)
69. Copper (II) sulfate pentahydrate ($CuSO_4 \cdot 5H_2O$) (Sigma-Aldrich, catalog number: C8027)
70. Aluminum potassium sulfate dodecahydrate ($AlK(SO_4)_2 \cdot 12H_2O$) (Fisher Scientific, catalog number: S70459)
71. Boric acid (H_3BO_3) (Fisher Scientific, catalog number: A73)
72. Sodium molybdate dehydrate ($Na_2MoO_4 \cdot 2H_2O$) (Sigma-Aldrich, catalog number: S6646
73. Sodium sulfide nonahydrate ($Na_2S \cdot 9H_2O$) (Sigma-Aldrich, catalog number: 431648)
74. Cysteine (Fisher Scientific, catalog number: BP377)
75. Glutamic acid (Amresco, catalog number: 0421)
76. Glycine (Sigma-Aldrich, catalog number: W328707)
77. Arginine (Sigma-Aldrich, catalog number: A5131)

78. Proline (Sigma-Aldrich, catalog number: W331902)

79. Asparagine (Amresco, catalog number: 94341)

80. Histidine (Sigma-Aldrich, catalog number: H800)

81. Isoleucine (Sigma-Aldrich, catalog number: W527602)

82. Leucine (Sigma-Aldrich, catalog number: L8000)

83. Lysine (Sigma-Aldrich, catalog number: L5626)

84. Threonine (Sigma-Aldrich, catalog number: T8625)

85. Tyrosine (Sigma-Aldrich, catalog number: T3754)

86. Alanine (Sigma-Aldrich, catalog number: W381829)

87. Methionine (Sigma-Aldrich, catalog number: M9625)

88. Phenylalanine (Sigma-Aldrich, catalog number: P2126)

89. Serine (Sigma-Aldrich, catalog number: S8407)

90. Tryptophan (Amresco, catalog number: E800)

91. Aspartic acid (Amresco, catalog number: 0192)

92. Glutamine (Sigma-Aldrich, catalog number: G3126)

93. Valine (Sigma-Aldrich, catalog number: V0500)

94. Phenol:Chlorofrom:Isoamyl Alcohol (see Recipes)

95. LB plates (see Recipes)

96. LB media (see Recipes)

97. KOD vitamins (see Recipes)

98. ASW-YT media (see Recipes)

99. 2x ASW solution (see Recipes)

100. Trace minerals solution (see Recipes)

101.Polysulfides (see Recipes)

102. 0.8x ASW solution (see Recipes)

103. 20 amino acid solution (see Recipes)

A2.3 EQUIPMENT

1. Pipettes (Gilson, catalog numbers: F123600, F123615, F123602)
2. 125 ml serum bottles (Wheaton, catalog number: 223748)
3. JLA10.500 rotor (Beckman Coulter, catalog number: 369681)
4. Enzyme Cooler, Isotherm System (Eppendorf, catalog number: 3880000011)
5. GasPak EZ Anaerobe Container System (BD, catalog number: 260678)
6. Eppendorf Microcentrifuge 5424 (Eppendorf, model: 5424)
7. Eppendorf Mastercycler Nexus Thermal cycler (Eppendorf, catalog number: 6333000022)
8. VWR forced air incubator (37 °C and 85 °C)
9. Thermo MaxQ 4000 benchtop orbital shaker (Thermo Fisher Scientific, model: MaxQ™ 4000)
10. Anaerobic Chamber (Coy Labs)
11. Beckman Avanti J Series centrifuge system
12. Autoclave
13. Dry block heater (VWR, catalog number: 12621-090)

A2.4 SOFTWARE

1. Primer3[9]

A2.5 PROCEDURE

Primer design

1. The first step in constructing a deletion strain is generating a plasmid construct that will facilitate the deletion of the gene of interest, referred to here as your favorite gene (YFG). A specific amplicon must be inserted into a common plasmid backbone (pTS700, see below), and the amplicon of choice must be generated with specific 5' and 3' sequences (added to the primers that generate the amplicon) to facilitate construction of the desired plasmid. Primer design starts by identifying the sequence of your favorite

gene (YFG) as well as the sequence of adjacent regions in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[10]. The *T. kodakarensis* reference genome is entry T00226.

2. Copy the positive-strand sequences encoding YFG plus 700 nucleotides upstream and 700 nucleotides downstream of YFG. End-join the 700 nucleotide sequence upstream and the 700 nucleotide sequence downstream of YFG by replacing the sequence of YFG with a single 'N'. The resulting 1,401 nucleotide sequence serves as an input sequence for the primer design software program Primer3[9].
3. Within Primer3, set the parameters to demand selection of a pair of 25 nucleotide primers that will yield an amplicon of ≥ 1,200 nucleotides from the inputted 1401 nucleotide sequence. Each primer should have a minimum T_m of 57 °C, no maximum T_m, 40%-75% G/C, and a 2 nucleotide GC-3' end clamp. Primer3 typically returns multiple primer pairs, and for convenience only, we select the first pair. These algorithm derived primers are sufficient to amplify the amplicon of choice, but must be amended to add sequences that facilitate insertion into pTS700.
4. To the upstream primer, add the sequence 5′ GGAGGTGAATTTC. This new primer (38 nucleotides) will be designated 001-XXXX, where XXXX is the gene number. To the downstream primer, 5′ GGTGAAGGATTTC will be added, and this primer will be designated 002-XXXX.

   Note: The additional sequence added to 001-XXXX and 002-XXXX primers facilitates the use of Ligation Independent Cloning (LIC), our preferred cloning method, and may be omitted if a different cloning method is used.

A. A-plasmid construction

Note: Construction of an A-plasmid is not necessary to delete a gene from the T. kodakarensis genome. The A-plasmid serves as a building block from which multiple gene modification constructs can be built, including the addition of an affinity tag sequence,

truncated versions of the gene, or allelic changes within the gene. Thus, if the deletion of the gene is the only desired result, B-plasmid construction (see 'Mutagenic PCR for B-plasmid construction) is the only necessary plasmid and may be constructed using an alternative cloning strategy. For example, ordering of a synthetic DNA fragment containing only the upstream and downstream regions of homology to YFG for use in cloning will suffice.

1. *T. kodakarensis* strain TS559 genomic DNA is prepared for amplification of the gene of interest using the primers designed above (001-XXXX and 002-XXXX). To properly inoculate a fresh culture of *T. kodakarensis*, 1 ml culture from a stock, 100 µl 1 M agmatine, 100 µl KOD vitamins (see Recipes), and 0.2 g of sulfur are added to 100 ml of ASW-YT media (see Recipes) under anaerobic conditions. The newly inoculated culture is incubated at 85 °C for 12 h prior to the transformation. Using a 1 ml syringe, remove 1 ml of *T. kodakarensis* TS559 culture and transfer to a 1.7 ml microcentrifuge tube, and collect the cells via centrifugation (9,000 *x g* for 5 min in a tabletop centrifuge). Decant the supernatant, resuspend the harvested cells in 100 µl of 10 mM Tris-HCl pH 8.0, and then add 50 µl of Phenol:Chloroform:Isoamyl Alcohol (see Recipes). After vigorously mixing, separate organic and aqueous phases via centrifugation (9,000 *x g*, 5 min in the tabletop centrifuge), aliquot 50 µl of the upper aqueous layer to a fresh 1.7 ml microcentrifuge tube that contains 50 µl 10 mM Trisi-HCl pH 8.0. Precipitate the nucleic acids by adding 100 µl of 100% isopropanol. Following a 30 minute spin (9,000 *x g*, tabletop centrifuge), carefully remove the isopropanol and allow the near-colorless, small nucleic acid pellet to dry for 10 min. Resuspend the dried pellet in 30 µl 10 mM Tris-HCl pH 8.0. The DNA is not quantified prior to use in PCR reactions.
2. PCR amplification is typically achieved using 001-XXXX and 002-XXXX primers in a mixture (50-100 µl) containing 3 µl of the genomic DNA (~300 ng), as prepared above. Any high-fidelity DNA polymerase can be used for the PCR amplification and should be used following the manufacturers recommendations.

3. To determine if the amplicon of interest was generated, ~5 µl of the amplification reaction is loaded into a 1% agarose gel, resolved, stained with ethidium bromide (EtBr), and imaged. If the correct product was amplified with no alternative products, the resultant amplicon is purified using AMPure XP beads following the manufacturer's directions. If a mixture of desired and alternative products were identified, the total reaction should be loaded into the gel, resolved, stained, imaged and the desired amplicon should be excised out of the gel and purified using a commercially available gel purification kit such as Nucleospin Gel and PCR Clean-up Kit by Machery-Nagel following the manufacturer's instructions.

   Note: A gel purification kit or PCR clean-up kit can be used as an alternative to AMPure XP beads.

4. Simultaneous to generating the desired amplicon for YFG, the plasmid pTS700 should be prepared for accepting the amplicon. pTS700 is the vector used for all gene deletion constructs and contains an *E. coli* origin of replication (*oriC*), Amp^R, and *T. kodakarensis* selectable and counter-selectable markers, TK0149 and TK0664, respectively. pTS700 also features a SwaI restriction enzyme cut site that is used to insert, via LIC, the YFG amplicon. pTS700 can be maintained in any standard *E. coli* strain using ampicillin as the selectable marker, is easily recovered from cultures via commercially available miniprep kits (we typically use the ZR Plasmid Miniprep Kit and follow manufacturer's instructions), and the concentration determined by fluorometric or spectroscopic techniques.
5. At this point, both the YFG desired amplicon and pTS700 should be purified. LIC will be used to build the new plasmid, pCSUXXXXA, again where XXXX is the gene number of YFG. LIC uses 12 nucleotide complimentary overhangs between the plasmid and PCR product to drive incorporation of the amplicon sequences into the vector. The

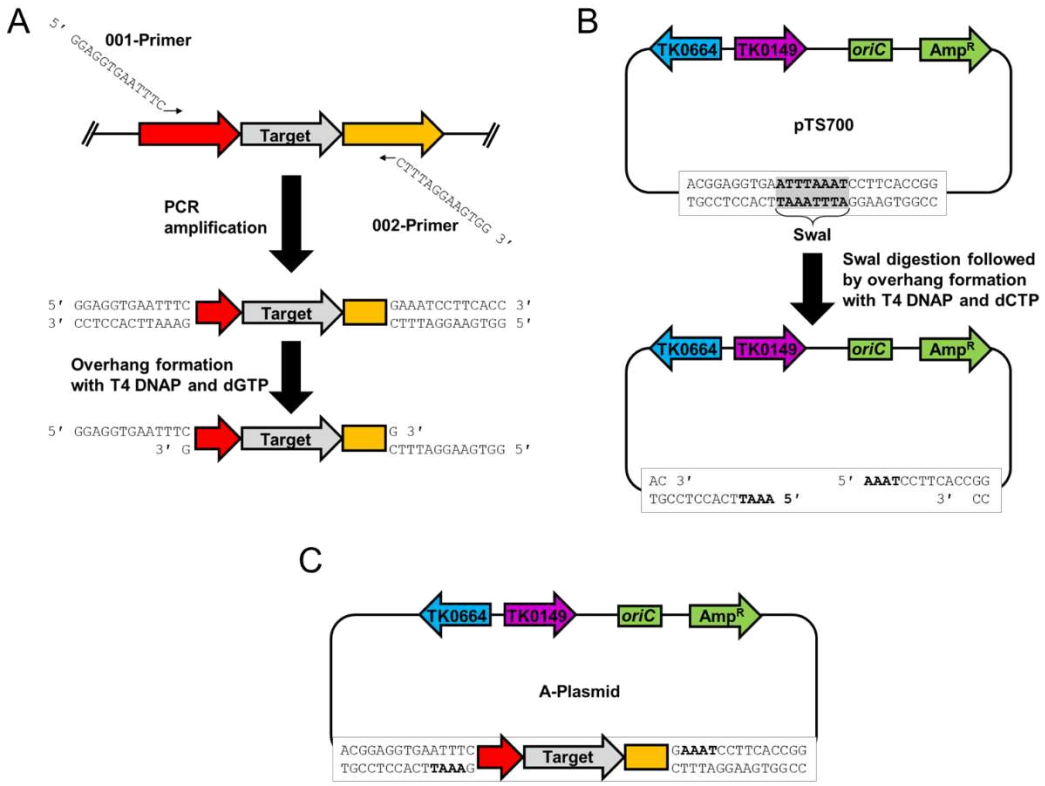

Figure A2.2. Construction of the A-plasmid using ligation independent cloning. A. The target gene as well as the upstream and downstream sequences are amplified using 001-XXXX and 002-XXXX primers that include the necessary 5'-tail sequences. The purified amplicon is incubated with T4 DNAP and dGTP to generate the 12 nucleotide overhangs. B. pTS700 is digested using SwaI and then incubated with T4 DNAP and dCTP to generate amplicon-complementary 12 nucleotide overhangs. C. The T4 DNAP treated amplicon and plasmid are incubated together and transformed into *E. coli* to generate the A-plasmid.

complementarity is generated by T4 DNA polymerase (DNAP)-mediated exonuclease activity on the vector and amplicon sequences (Figure A2.2).

Note: LIC is not essential in A-plasmid construction. We utilize LIC for plasmid construction because it is the most cost effective-option for constructing deletionplasmids for all 2,306 genes in the T. kodakarensis genome. LIC is not dependent on a DNA ligase or phosphatase, and the only enzyme required is T4 DNAP[11]*. Any standard cloning method may be used in place of LIC.*

Digest 500 ng of purified pTS700 with SwaI in a 20 µl reaction following manufacturer's instructions. Following SwaI digestion, 1 µl of T4 DNAP and 1 µl of 100 mM dCTP are added, incubation is continued at 37 °C for 30 min, then 20 min at 85 °C to heat inactivate all enzymes, and finally the reaction is placed on ice. Simultaneously, a second reaction will prepare the YFG amplicon for LIC. Mix 500 ng desired amplicon, 0.9 µl NEBuffer 2.1, 0.5 µl 100 mM dGTP, and 0.5 µl T4 DNAP in a final volume of 9 µl, incubate at 37 °C for 30 min, then 20 min at 85 °C to heat inactivate the enzymes. 1 µl aliquots of each reaction are combined in a single tube, heated to 85 °C for 3 min, and the mixtures allowed to slowly cool to room temperature. This slow cooling permits the 12 nucleotide complementary overhangs to anneal. Following the cooling step, the 2 µl reaction is transformed into 50 µl competent cells of any standard *E. coli* strain, and spread onto LB-Amp plates (see Recipes).

1. After overnight incubation, colony PCR is used to determine if resultant transformants contain the newly constructed, desired A-plasmid. At least 10 distinct colonies should be checked using the pTS700 specific primers (700Forward and 700Reverse), although in some cases additional colonies may need to be screened. Each colony is picked from the plate into 6 µl of dH_2O in a 0.2 ml PCR tube and resuspended. 4 µl is removed and spotted onto a LB-Amp plate leaving 2 µl behind in the PCR tube (ensure that the label

on the PCR tube matches that on the spot plate). Once all colonies are spotted, incubate the LB + Amp plate at 37 °C to allow overnight growth. The remaining 2 µl of cells will be used as the DNA component in an amplification reaction using Taq DNAP. Taq DNAP yields the best results as it is not sensitive to the cellular and media components carried over into the reactions.

Colony PCR Conditions (20 µl)

Taq DNAP (20 U/µl)	0.1 µl
dNTPs (2.5 mM)	1.6 µl
10x ThermoPol buffer	2 µl
700Forward primer (100 µM)	0.1 µl
700Reverse primer (100 µM)	0.1 µl
Resuspended Cells	2 µl
H_2O	13.9 µl

Following PCR, the total reaction is loaded into a 1% agarose gel, resolved, stained, and imaged. If the LIC was unsuccessful, a product of 150 bp will be generated. If LIC was successful, a product of 150 bp plus the size of your initial amplicon (YFG + 1,400 bp flanking regions) will be present. Identify one of the successful LIC generated plasmids, return to the spot plate, and pick the corresponding colony into 5 ml of LB with ampicillin and grow overnight at 37 °C while shaking. Purify the plasmid from *E. coli* using a preferred plasmid Miniprep kit, quantify, and sequence the entire amplicon using the 700Forward and 700Reverse primers in two separate reactions. If the sequencing confirms the YFG along with the upstream and downstream sequences have been inserted into pTS700, the plasmid is now designated pCSUXXXXA and is referred to as the A-plasmid.

B. Mutagenic PCR for B-plasmid construction

1. The B-plasmid is used for gene deletions and lacks the YFG but retains the upstream and downstream sequences. To construct the B-plasmid, a Quikchange-II reaction is used with the A-plasmid as DNA template. 60 nucleotide primers are employed, and the primers bind to 30 nucleotides on either side of the gene of interest thus deleting the sequences encoding YFG from the final product. The first 30 nucleotides should be identical to the 30 nucleotides immediately upstream of YFG, while the 3' terminal 30 nucleotides should be identical to the 30 nucleotides immediately downstream of YFG. These long primers ensure the primers are of the adequate melting temperature (> 78 °C). The 60 nucleotide primers used in the Quikchange are designated 016-XXXX and 017-XXXX and, as Quikchange demands, the primers are reverse complements of each other.
2. The conditions for the Quikchange reaction follow exactly the Quikchange-II kit protocol including the DpnI treatment to degrade the *E. coli* methylated A-plasmid. After the DpnI treatment, 2 µl of the total reaction is used to transform 50 µl XL1-Blue *E. coli* cells, which are then plated on LB-Amp agar plates, and incubated overnight at 37 °C. Colony PCR is performed, again using primers 700Forward and 700Reverse, to determine if the sequences encoding YFG were deleted from the A-plasmid leaving only the now-fused upstream and downstream sequences. The product size for all deletions should be ~1,400 bp, and the colony corresponding to the deletion is picked into 5 ml LB-Amp, incubated at 37 °C overnight, and newly constructed B-plasmid DNA purified using the preferred plasmid Miniprep Kit. The resulting purified B-plasmid is sequenced using the 700Forward and 700Reverse primers in two separate reactions. If the sequencing demonstrates that YFG has been deleted, this new plasmid is now designated pCSU-XXXXB and will be used to construct the *T. kodakarensis* deletion strain.

C. *Thermococcus kodakarensis* transformation

1. The B-plasmid must be transformed into *T. kodakarensis* strain TS559. The first step of the transformation is to inoculate a fresh culture of TS559 as described in step B1.
2. To complete the transformation, the cells will be spread onto specialized plates that will remain solid at 85 °C. The media used to make the plates must be autoclaved immediately before pouring. Plates should be poured no more than 24 h prior to use. Two 100 ml serum bottles with unique ingredients are autoclaved separately and mixed immediately before pouring. Each set of serum bottles will yield four plates, and each transformation requires two plates. For all plates, one of the bottles will remain consistent and contains 1 g of Gelzan in 50 ml of distilled H_2O. Once the Gelzan and dH_2O are added, cap the serum bottle with a rubber stopper and aluminum seal and shake to mix well. Note that the Gelzan will not completely dissolve prior to autoclaving. The contents of the second bottle will vary depending on the composition of the desired solid media. The media required for the initial transformation is ASW-YT-S, and thus the second serum bottle will contain 50 ml of 2x ASW (see Recipes) with 0.5 g yeast extract, 0.5 g tryptone, and 500 µl trace minerals (see Recipes). After all reagents are in the bottle, cap with a rubber stopper and seal with an aluminum seal.

   Note: Place both bottles in an autoclave-safe vessel with a lid and add a small amount of water to the bottom. For this, we used a large noodle pot, however any large vessel with a lid will work. The sealed bottles present an explosion hazard in the autoclave. Placing the bottles in a large noodle pot ensures that on the rare occasions an explosion occurs it will be contained. When working with sealed vials and the autoclave, proper safety precautions should be taken including the use of face shields, lab coats, and autoclave gloves.

   A standard liquid autoclave cycle with a 20-min sterilization at 121 °C is sufficient to sterilize the media and dissolve the Gelzan. As soon as the autoclave cycle has finished,

remove the noodle pot, and bring the bottles into the anaerobic chamber. The Gelzan will begin to solidify at any temperature below 85 °C and will remain solidified so you must pour the plates immediately after the autoclave cycle has completed. For every set of two bottles, four, previously autoclaved, glass Petri plates should be arranged in the chamber to allow for quick pouring. Once the bottles are in the chamber, work quickly to uncap the bottle containing the 2x ASW and add to this bottle 100 µl vitamins and 200 µl polysulfides (see Recipes). Swirl to mix well, careful not to spill. Now, uncap the bottle containing Gelzan and pour the 2x ASW bottle into the Gelzan bottle, again swirl to mix well. Pour approximately ¼ of the bottle into each plate. The plates should set in about 10 sec. Allow the plates to cool for ~10 min then flip the plates over until ready for use.

3. After 12 h, the TS559 liquid culture (100 ml) is taken into the anaerobic chamber and poured into a polycarbonate centrifuge tube suited for use in a high-speed centrifuge. The centrifuge tube is sealed, removed from the chamber, and spun in a JLA10.500 rotor at 18,000 *x g* for 10 min at 4 °C to pellet the cells. After centrifugation, immediately remove and invert the centrifuge bottle so that the cell pellet is not disturbed by the supernatant while moving the tube back into the anaerobic chamber. Once the tube is back in the chamber, carefully remove the supernatant without disturbing the pellet.
   Note: Polycarbonate tubes must be used as they can withstand the high temperatures of the cultures.
4. The cell pellet is resuspended in 3 ml of 0.8x ASW (see Recipes). Each transformation only requires 200 µl of resuspended cells therefore a 100 ml of culture yields enough cells for ~15 transformations. Following resuspension, remove the cells from the centrifuge bottle and place them in 1.7 ml microcentrifuge tubes on ice.
   Note: To prevent the chambers from becoming humid, use a small ice bucket with a lid or an enzyme cooler. Allow the cells to incubate on ice for 30 min.
   Note: Cells cannot be stored for future experiments.

5. While the cells incubating on ice, the previously purified B-plasmid(s) are prepared for the transformation. For each transformation, aliquot 3 µg of purified B-plasmid into a 1.7 ml microcentrifuge tube and bring into the chamber.
6. After the cells have incubated for 30 min on ice, add 200 µl of cells to the 3 µg of plasmid. Continue incubating on ice for an additional 60 min.
7. Although *T. kodakarensis* is naturally competent, a heat shock step is used in the transformation protocol to increase the efficiency of the transformation. Heat the cell/plasmid mixture to 85 °C for 1 min followed by a 5 min recovery on ice. Each transformation will be spread on two plates, one with a high volume of cells (160 µl) and one with a low volume of cells (40 µl) using a cell spreader. After spreading, the plates are flipped upside down and placed in a metal cylinder, packed with paper towels to absorb moisture and a GasPak EZ Anaerobe Container System packet to maintain anaerobic conditions during growth. The cylinder is sealed in the chamber, removed from the chamber, and placed in an 85 °C incubator for 48-72 h.
8. After allowing colony formation, remove the cylinder from the incubator and bring into the anaerobic chamber. Remove the plates from the cylinder and identify *T. kodakarensis* colonies. The colonies are small, clear, and can be difficult to identify. For this reason, it is difficult to pick single colonies directly from the transformation plates and it is therefore often necessary to spot the colonies. At least 10 colonies are spotted onto media of identical composition to the transformation plates. To spot single colonies, a colony from the transformation plate is picked into 6 µl of 0.8x ASW and serial dilutions are spotted (spots are ~6 µl). Using 0.8x ASW and a 10-fold dilution at each step, spot the colony for 10 dilutions onto freshly poured plates. The plates are then placed in a metal cylinder with an anaerobe pack and incubated at 85 °C for 48-72 h.

9. Following the incubation, the plates are removed from the incubator and brought into the anaerobic chamber. One colony from each series of serial dilutions is used to inoculate 5 ml of ASW-YT supplemented with 0.1 g of elemental sulfur and 5 µl KOD vitamins. These freshly inoculated bottles are sealed in the anaerobic chamber, removed, and incubated at 85°C overnight. These cultures will be referred to as potential intermediate strains. After overnight growth, total genomic DNA will be extracted from these cultures as described in section B1.

D. Selection of intermediate strains

1. For the transformation to be successful, indicated by the presence of viable colonies, the B-plasmid must have integrated into the genome via homologous recombination. The plasmid has the potential to recombine at one of two loci, either upstream or downstream of the gene of interest (Figure A2.3).

It is necessary to confirm via PCR that the B-plasmid integrated into the genome and determine the locus of this recombination. Purified genomic DNA from potential intermediate strains (as specified in section D9 above) will be used in PCR reactions with two separate primer combinations: 700Forward is used in combination with 002-XXXY while 700Reverse will be used 001-XXXW. PCR is performed using 3 µl (~300 ng) of genomic DNA with Phusion DNAP following manufacturer's instructions. The PCR reactions are resolved in a 1% agarose gel, stained, and imaged. Ideally, both the upstream and downstream recombination events will be identified in individual colonies, however in some cases only one of the recombination events is identified.

Note: The 001 and 002 primers used in this PCR are not for YFG, but instead for the genes immediately upstream and downstream of YFG, respectively. If the 001 and 002 primers are not available for the adjacent genes, any primer should suffice as long as it

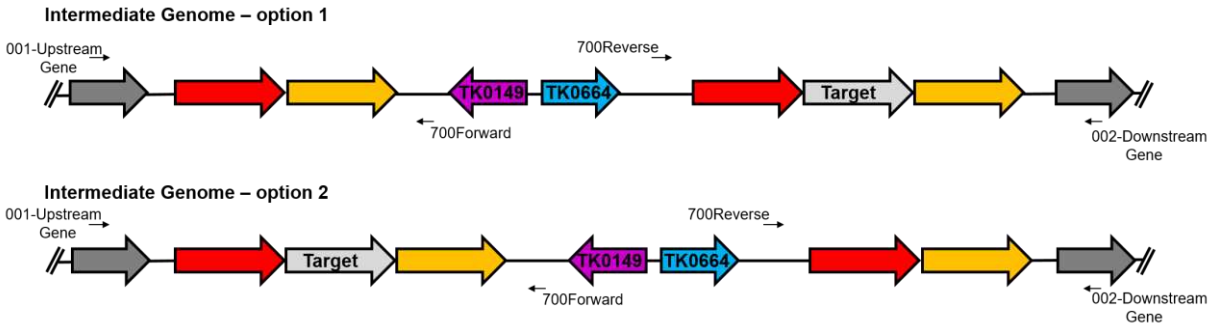

Figure A2.3. Identification of the intermediate genomes using PCR. The intermediate genomes are identified using primers specific to the B-plasmid (700Forward and 700Reverse) with primers specific to the genome sequence. For intermediate PCR, the 001 and 002 primers for the target gene cannot be used; instead the 001 and 002 primers for the upstream and downstream genes, respectively, should be used.

is external to the 001 and 002 primers for the gene of interest. The 001 and 002 primers for the gene of interest cannot be used because they will bind both the plasmid and genomic DNA sequences.

2. Using the results from the intermediate PCR, two distinct intermediates are selected, ideally one each of the 2 possible recombination events. Each of the selected intermediate strains is used to inoculate 5 ml of ASW-YT supplemented with 5 µl KOD vitamins, elemental sulfur, and 5 µl 1 M agmatine. Agmatine is added at this point so that the cells no longer have selective pressure to retain the plasmid in the genome and to allow for the second recombination event to occur. These cultures are allowed to grow for 12 h at 85 °C.
3. To select for cells that have spontaneously excised the plasmid sequences and potentially generated the deletion strain of YFG, overnight cultures of the intermediate strains are spread on plates containing both agmatine and 6-methylpurine in addition to KOD vitamins, polysulfides, and 20 amino acids solution (see Recipes). For each intermediate culture, two plates will be poured. As done in step D4, two 50 ml serum bottles are used. As before, one serum bottle will contain 1 g of Gelzan in 50 ml of H_2O. The second serum bottle will now contain only 50 ml of 2x ASW and 500 µl trace minerals. The same autoclaving procedure is used. After removing the serum bottles from the autoclave and bringing each into the anaerobic chamber, the serum bottle containing 2x ASW is opened and 100 µl of 1 M agmatine, 100 µl 100 µM 6-methylpurine, 200 µl polysulfides, and 5 ml 20 amino acid solution are added and mixed well. This solution is then added to the Gelzan containing bottle, mixed well, and poured into 4 glass Petri plates.
4. Following 12 h incubation, the agmatine-containing confirmed intermediate cultures are brought into the chamber where they will be spread onto the counter-selective plates. The two plates for each culture will be used to spread either high (160 µl) or low (40 µl)

volumes of cells. For each plate, the proper number of cells are pipetted onto the surface and a cell spreader is used to spread the cells around the plate. Following this, the plates are put into a metal cylinder with an anaerobe pack and sealed. The cylinder is removed from the chamber and placed in the 85 °C incubator for 48-120 h.

5. After 85 °C incubation, at least 10 colonies from the spread plates must be spotted onto newly poured plates made in the exact same manner as the spread plates to obtain single colonies. To spot single colonies, a colony from the spread the plate is picked into 6 µl of 0.8x ASW and serial dilution spot plating is performed using 0.8x ASW and a 10-fold dilution at each step for 10 steps (spots are ~6 µl). Following serial-dilution spotting, the plates are placed in a metal cylinder with an anaerobe pack and incubated at 85 °C for 48-120 h.
6. After these spot plates have been allowed to grow, one colony from each serial dilution series is used to inoculate 5 ml of ASW-YT supplemented with 5 µl 1M agmatine, 5 µl KOD vitamins, and 0.1 g elemental sulfur. After overnight growth, total genomic DNA can be extracted from these cultures as described in section B1.

E. Confirmation of deletion strain

1. There are two possible recombination events that could occur when the plasmid recombines out of the genome. One recombination will return the genotype to the parental strain that retains YFG, while the other will yield the genome with the desired deletion of YFG. If after checking 30 individual colonies from at least 2 intermediates, ideally one from each of the possible recombination events, only the wildtype genome is observed, then the gene is deemed statistically essential. A series of different PCR reactions are performed to determine if YFG was deleted (Figure A2.4).

TS559 genomic DNA is always used as a positive control for the final PCR reactions. The first PCR uses the 001-XXXX and 002-XXXX primers to determine if YFG is deleted, and a second reaction uses either 001-XXXX or 002-XXXX and a primer

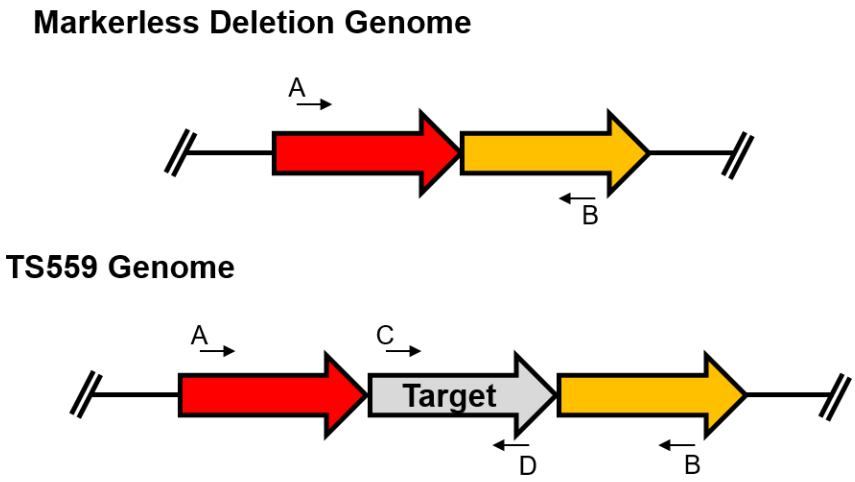

Figure A2.4. Confirmation of the desired deletion strain. The primer designated 'A' is the 001-XXXX and the primer designated 'B' is 002-XXXX. The C and D primers are any primers internal to the gene of interest. To confirm the deletion of the gene of interest, at least 3 combinations of primers must be used, A/B, C/D, and A/D or B/C. If the gene of interest has been deleted from the genome, no product will be synthesized using the C and D primers.

internal to YFG. The third PCR reaction utilizes a pair of primers that are completely internal to YFG. The lack of a PCR product for the second and third reactions signifies the deletion occurred as long as the product was present using the control TS559 genomic DNA. These PCR reactions are performed using a high fidelity DNAP and 3 µl genomic DNA (~300 ng). Following PCR, the products are separated using a 1% agarose gel, stained with EtBr, and imaged.

2. If the desired deletion genome is identified within the final PCR amplifications, further amplifications are performed on that strain to extend confidence that YFG is truly deleted. The external PCR using the 001-XXXX and 002-XXX primers is repeated in a 50 µl reaction using Phusion DNAP, and 3 µl of purified genomic DNA (~300 ng). Following PCR, only 5 µl of this reaction is used in an agarose gel to confirm the PCR worked. The remaining reaction is sent to a DNA sequencing facility for purification and sequencing using the 001-XXXX and 002-XXXX primers in two separate sequence reactions.
3. Once the sequencing has confirmed the deletion of YFG, the deletion strain has been constructed. If desired, a Southern blot or whole genome sequencing can be performed to further confirm the deletion.

A2.6 DATA ANALYSIS

1. Plasmid construction confirmation

   Construction of both the A and B plasmid is typically a facile procedure with few problems. Complications can result from primer design errors or primer incompatibility, and when such occurs, the use of the second set of primers selected by the program is typically sufficient to resolve any issues. Infusion cloning can be used (following manufacturer's instructions) in instances where LIC is unsuccessful. Construction of pCSU-0556A and pCSU-0556B were typical and non-problematic using the described protocol.

2. Intermediate strain PCR

   PCR for the intermediate strains may only give a product for one of the primer pairs (001-XXXX /700Reverse or 002-XXXX/700Forward). In most cases, modification of the PCR conditions, annealing temperature or elongation time, will allow for amplification using both pairs of PCR primers. For the deletion of TK0556, both intermediate strains were identified using standard PCR conditions (Figure A2.5).

3. Final deletion strain PCR

   Ideally, four PCR reactions, in addition to DNA sequencing, will be used to determine if the gene of interest was deleted from the *T. kodakarensis* TS559 genome (Figure A2.6). The PCR reaction using two primers internal to the gene of interest is crucial in ensuring that the gene is deleted from the genome, and has not moved to a different locus via an off-target recombination event.

A2.7 RECIPES

1. Phenol:Chlorofrom:Isoamyl Alcohol (25:24:1)

   25 ml Tris-saturated Phenol

   24 ml chloroform

   1 ml isoamyl alcohol

2. LB media (1 L)

   10 g tryptone

   5 g yeast extract

   10 g NaCl

3. LB-Amp plates (1 L)

   10 g tryptone

   5 g yeast extract

   10 g NaCl

   15 g agar

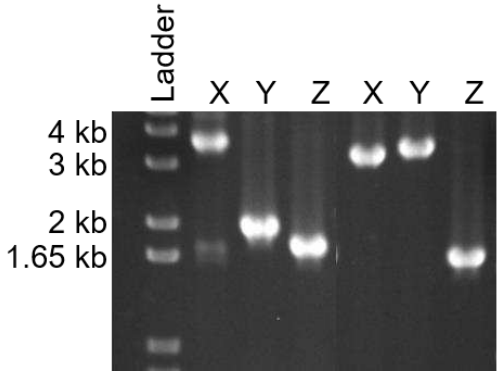

Figure A2.5. Identification of both intermediate genomes for the deletion of TK0566 using diagnostic PCR. X and Y lanes are used to determine the orientation of the intermediate genome. Primers 001-0565 and 700Reverse are used in the X reaction while 002-0567 and 700Forward are used in the Y reaction. If the genome is the first intermediate, the X product should be ~3,600 bp while the Y products should be ~1,600 bp. For the second intermediate, the X product should be ~3,600 bp and the Y product ~3,650 bp. A third, control PCR reaction (Z) was performed using 001-1418 and 002-1418 to ensure that the genomic preps yielded DNA suitable for PCR. The expected Z product was ~1,600 bp.

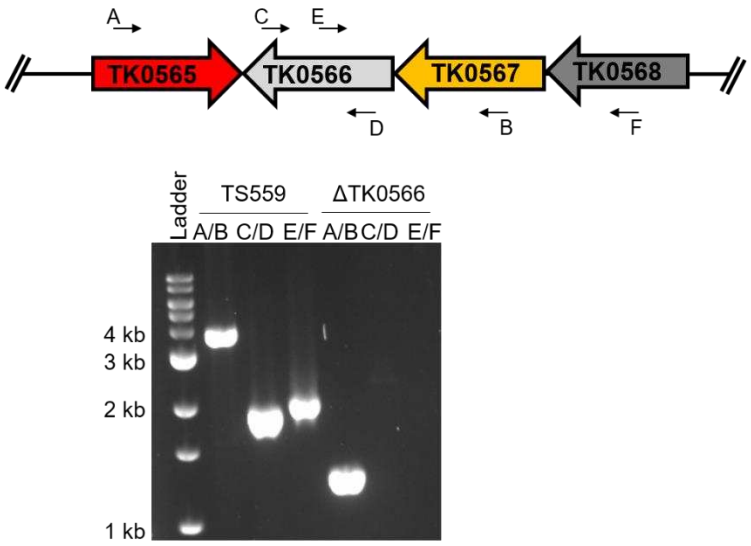

Figure A2.6. Diagnostic PCR confirms the deletion of TK0566 from the *T. kodakarensis* genome. The same primer pairs are used on the TS559 genome as the deletion genome. The size shift using primers A/B corresponds to the deletion of TK0566 while the absence of products for the reactions using C/D and E/F confirms the gene has been deleted from its native locus. The absence of products using primers C/D demonstrates that TK0566 is not present anywhere in the genome of the deletion strain.

4. KOD Vitamins (200x) (1 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

   Note: This solution is light sensitive and should be protected.

5. ASW-YT media

   Artificial sea-water medium supplemented with:

   0.5% (w/v) tryptone

   0.5% (w/v) yeast extract

   1x trace mineral solution

   1x vitamin mixture

   1x artificial sea-water contains (1 L):

   20 g NaCl

   3 g $MgCl_2 \cdot 6H_2O$

   6 g $MgSO_4 \cdot 7H_2O$

   1 g $(NH4)_2SO_4$

   200 mg $NaHCO_3$

   300 mg $CaCl_2 \cdot 2H_2O$

   0.5 g KCl

420 mg KH_2PO_4

50 mg NaBr

20 mg $SrCl_2 \cdot 6H_2O$

10 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$

6. 2x ASW (1 L)

   40 g NaCl

   6 g $MgCl_2 \cdot 6H_2O$

   12 g $MgSO_4 \cdot 7H_2O$

   2 g $(NH4)_2SO_4$

   400 mg $NaHCO_3$

   600 mg $CaCl_2 \cdot 2H_2O$

   1 g KCl

   840 mg KH_2PO_4

   100 mg NaBr

   40 mg $SrCl_2 \cdot 6H_2O$

   20 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$

7. Trace minerals (1,000x) (1 L)

   0.5 g $MnSO_4 \cdot H_2O$

   0.1 g $CoCl_2 \cdot 6H_2O$

   0.1 g $ZnSO_4 \cdot 7H_2O$

   0.01 g $CuSO_4 \cdot 5H_2O$

   0.01 g $AlK(SO_4)_2 \cdot 12H_2O$

   0.01 g H_3BO_3

   0.01 g $Na_2MoO_4 \cdot 2H_2O$

8. Polysulfides

   10 g $Na_2S \cdot 9H_2O$, 3 g Sulfur per 15 ml

   Dissolve the mixture using heat

   Note: It should be a deep red color when complete.

9. 0.8x ASW (1 L)

   16 g NaCl

   2.4 g $MgCl_2 \cdot 6H_2O$

   4.8 g $MgSO_4 \cdot 7H_2O$

   800 mg $(NH_4)_2SO_4$

   160 mg $NaHCO_3$

   240 mg $CaCl_2 \cdot 2H_2O$

   400 mg KCl

   336 mg KH_2PO_4

   40 mg NaBr

   16 mg $SrCl_2 \cdot 6H2O$

   8 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$

   Note: must be autoclaved

10. 20 amino acid solution (1 L)

    1 g cysteine

    1 g glutamic acid

    1 g glycine

    500 mg arginine

    500 mg proline

    400 mg asparagine

    400 mg histidine

    400 mg isoleucine

400 mg leucine

400 mg lysine

400 mg threonine

400 mg tyrosine

300 mg alanine

300 mg methionine

300 mg phenylalanine

300 mg serine

300 mg tryptophan

200 mg aspartic acid

200 mg glutamine

200 mg valine

REFERENCES

1. 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* **2005**, *15*, 352–363.
2. Sato, T.; Fukui, T.; Atomi, H.; Imanaka, T. Targeted Gene Disruption by Homologous Recombination in the Hyperthermophilic Archaeon Thermococcus Kodakaraensis KOD1. *J Bacteriol* **2003**, *185*, 210–220.
3. Sato, T.; Fukui, T.; Atomi, H.; Imanaka, T. Improved and Versatile Transformation System Allowing Multiple Genetic Manipulations of the Hyperthermophilic Archaeon Thermococcus Kodakaraensis. *Appl Environ Microbiol* **2005**, *71*, 3889–3899.
4. Santangelo, T. J.; Cubonova, L.; Reeve, J. N. Shuttle Vector Expression in Thermococcus Kodakaraensis: Contributions of Cis Elements to Protein Synthesis in a Hyperthermophilic Archaeon. *Appl Environ Microbiol* **2008**, *74*, 3099–3104.
5. Hileman, T. H.; Santangelo, T. J. Genetics Techniques for Thermococcus Kodakarensis. *Front Microbiol* **2012**, *3*, 195.
6. Santangelo, T. J.; Reeve, J. N. Genetic Tools and Manipulations of the Hyperthermophilic Heterotrophic Archaeon Thermococcus kodakarensis. In *Extremophiles Handbook*; Springer Japan: Tokyo, 2011; pp. 567–582.
7. 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. *Applied and environmental microbiology* **2010**, *76*, 1044–52.
8. Walker, J. E.; Luyties, O.; Santangelo, T. J. Factor-Dependent Archaeal Transcription Termination. *Proceedings of the National Academy of Sciences* **2017**, *114*.
9. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J. A. M.

Primer3Plus, an Enhanced Web Interface to Primer3. *Nucleic Acids Research* **2007**, *35*, W71–W74.

10. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Research* **2017**, *45*, D353–D361.

11. Aslanidis, C.; Jong, P. J. de Ligation-Independent Cloning of PCR Products (LIC-PCR). *Nucleic acids research* **1990**, *18*, 6069–74.