

Key Concepts and Challenges in Archaeal Transcription

Fabian Blombach*, Dorota Matelska, Thomas Fouqueau, Gwenny Cackett, and Finn Werner*

Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, United Kingdom

*Correspondence: f.blombach@ucl.ac.uk and f.werner@ucl.ac.uk

Keywords: transcription, RNA polymerase, archaea, regulation, integrated systems biology

Abbreviations: CC – closed complex, IMR – initially melted region, OC – open complex, PIC – pre-initiation complex, RNAP – RNA polymerase, TCR – transcription-coupled repair, TEC – transcription elongation complex, TSS – transcription start site, TTS – transcription termination site, TU – transcription unit

Abstract

Transcription is enabled by RNA polymerase and general factors that allow its progress through the transcription cycle by facilitating initiation, elongation and termination. The transitions between specific stages of the transcription cycle provide opportunities for the global and gene-specific regulation of gene expression. The exact mechanisms and the extent to which the different steps of transcription are exploited for regulation vary between the domains of life, individual species and transcription units. Yet, a surprising degree of conservation is apparent. Similar key steps in the transcription cycle can be targeted by homologous or unrelated factors providing insights into the mechanisms of RNAP and the evolution of the transcription machinery. Archaea are bona fide prokaryotes but employ a eukaryote-like transcription system to express the information of bacteria-like genomes. Thus, archaea not only provide the means to study transcription mechanisms of interesting model systems, but also to test key concepts of regulation in this arena. In this review, we discuss key principles of archaeal transcription, new questions that still await experimental investigation, and how novel integrative approaches hold great promise to fill this gap in our knowledge.

Introduction

Advanced phylogenetic analyses suggest that life on earth is divided into two primary domains, bacteria and archaea and that a primordial archaeon closely related to the extant Asgård phylum was the likely ancestor of all eukaryotes [1, 2]. In this vein, the molecular machinery carrying out processing of biological information (i.e. replication, transcription and translation) in eukaryotes is closely related to the cognate systems in archaea [3]. All archaeal and eukaryotic

RNA polymerases (RNAPs) share a conserved core architecture in terms of subunit composition, structure and molecular mechanisms of nucleotide translocation, phosphodiester bond formation and cleavage [4]. In addition, they share a core set of basal transcription initiation (i.e. TBP, TFB and TFE) and elongation factors (including TFS, Spt4/5, Elf1) distinct from those used by bacteria with the sole exception of Spt5 that is homologous to bacterial elongation factor NusG [5-7]. Many archaea use histones to form a chromatinised template for transcription [8, 9]. Archaeal cells are prokaryotic and thus the two key processes of gene expression, transcription and translation are likely to be functionally coupled [10-12]. The transcription unit (TU) organisation and mRNAs of archaea and bacteria share important basic features: (i) many genes are organised in polycistronic operons, (ii) mRNAs are cap-less and intron-less, and (iii) translation initiation is reliant on a Shine-Dalgarno-dependent ribosome-binding mechanism or a leaderless mechanism with start codons located proximal to the mRNA 5'-terminus [13]. Archaeal genomes are relatively small (1.5-7 Mbp) and compact with very short intergenic regions, which limits the space for regulatory transcription factor binding sites compared to eukaryotic genomes. In essence, archaea run a gene expression program using a eukaryote-like transcription machinery on a bacterial-like genome. This unorthodox situation provides a unique opportunity to investigate (i) the molecular mechanisms of transcription in a highly tractable model system (see below), (ii) the evolution of transcription systems in archaea and eukaryotes, and (iii) the fundamental principles guiding the regulation of gene expression in prokaryotes and by inference eukaryotes.

The most common and well-studied archaeal model organisms for transcription exhibit an extremophilic lifestyle, including thermophiles such as *Sulfolobus*, *Thermococcus*, *Pyrococcus*, and *Methanocaldococcus* as well as the highly genetically tractable halophiles including

Halobacterium and *Haloferax* species. The thermophilic lifestyle of these archaea often translates into a high biochemical tractability of their proteins, which has made them favored targets for structure determination by crystallography [14-17] and has facilitated the *in vitro* reconstitution of RNAPs from their 12 individual recombinant subunits [18, 19], a feat that has not been successful for any eukaryotic RNAP. Reconstituted recombinant RNAPs have allowed a rigorous molecular genetics analysis of archaeal RNAPs unconstrained by cell viability, as well as the site-specific incorporation of molecular probes including fluorescent dyes and nitroxide spin labels into RNAP subunits. These derivatised RNAPs have been applied in a broad range of functional and biophysical studies that have furthered our understanding of RNAP mechanisms [19-25]. The extreme physical and chemical conditions including high temperature and low pH require special adaptations of biochemical assays and often rule out the use of conventional selective agents including antibiotics and chemical drugs, which rapidly degrade in the growth medium or reaction buffers. In addition, the archaeal RNAP appears insensitive to alpha-amanitin and rifampicin that are widely used as inhibitors of eukaryotic RNAPII and bacterial RNAP, respectively [26, 27]. Many archaeal proteins are modified by phosphorylation and methylation, but to date there is no direct evidence that components of the archaeal transcription machinery, including histones, are regulated by post-translational modifications [28-32]. In comparison, post-translational modifications play a major role in the regulation of RNAPII, including the extensive phosphorylation of the largest RNAPII subunit C-terminal domain that is not conserved in archaea [28, 33].

Archaeal model systems have provided many insights into the fundamental mechanisms of transcription. The key strategy to unlock the remaining mysteries is to integrate (i) mechanistic hypotheses derived from molecular structures and *in vitro* transcription experiments, (ii) classical

genetics and (iii) information provided by genome-wide mapping techniques, a field that has been rapidly developing in recent years. While current reviews have described recent progress in the field [6, 34], this manuscript focusses on some of the big questions that remain unanswered.

Mechanisms of transcription- lessons learned *in vitro*

The mechanisms of transcription initiation in archaea have been thoroughly characterised *in vitro*. Initiation proceeds via the stepwise assembly of a pre-initiation complex (PIC) that is directly comparable to the eukaryotic RNAPII PIC [35]. PIC formation is nucleated by the binding of the TATA-binding protein (TBP) and Transcription Factor B (TFB, TFIIB in eukaryotes) to the promoter, followed by the recruitment of RNAP and Transcription Factor E (TFE, TFIIE in eukaryotes) (Fig. 1AB). TBP and TFB recognize two promoter elements in a sequence-specific fashion, the TATA box and the B-Recognition Element (BRE), respectively. The binding of TBP distorts the promoter topology by inducing a $\sim 90^\circ$ bend in the DNA [14, 36]. TFB enhances the stability of the TBP-DNA complex [36], provides the correct orientation [37], and recruits the RNAP via multiple interactions with the RNAP dock domain and RNAP active site cleft [38, 39]. TFE facilitates the conversion of the PIC from the closed complex (CC) to the open complex (OC). During the conversion to the OC the DNA strands are locally separated upstream of the transcription start site (TSS) and the template strand is inserted into the active site of RNAP. The Initiator (Inr) promoter element is a dinucleotide motif that is important for the precise choice of TSS [40]. Transcription starts with the synthesis of very short (3-9 nt) 'abortive' transcripts, during which the PIC converts into the initially transcribing complex (ITC) that remains bound to the promoter via TBP and TFB. The fixed attachment of RNAP to the initiation factors upstream while downstream DNA is railed in likely results in DNA scrunching as has been observed for bacterial

RNAP [41]. In order to productively synthesize RNA transcripts, the RNAP has to be released from TBP and TFB. The underlying molecular mechanisms are complex and likely involve a concerted displacement of TFB domains by conformational changes in the RNAP and by the growing RNA chain similar to eukaryotic RNAPII [20, 38, 42]. During initiation, DNA melting and template strand loading can occur in the absence of TFE, since TBP and TFB are sufficient to initiate RNAP transcription in a TSS-specific fashion *in vitro* [18, 43]. To what extent transcription initiation depends on TFE *in vivo* remains to be determined. The essential nature of *tfeA*, the gene encoding the TFE alpha subunit, emphasizes the importance of TFE [44-46]. In most archaeal species, TFE is a heterodimer consisting of alpha and beta subunits, while several euryarchaeal lineages have lost the *tfeB* gene and use a monomeric TFE alpha variant with complete TFE functionality [19, 21, 39, 47]. The haloarchaea lineage within euryarchaea maintained the *tfeB* gene, but in at least in *Haloferax volcanii* *tfeB* is not essential [44]. In crenarchaea *tfeB* is strictly conserved and the gene is essential in *Sulfolobus acidocaldarius* [46, 48].

Transcription elongation is a discontinuous process that is frequently interrupted by pausing or arrest caused by the sequence context (e.g. poly-A tracts), DNA-associated proteins that present a road block to elongating RNAPs (including histone-based chromatin), or chemical DNA adducts (e.g. pyrimidine dimers) typical for radiation damage. A spectrum of elongation factors assists RNAP in overcoming these impediments, including the transcript cleavage factor TFS, the processivity factors Spt4/5 and Elf1 (Fig. 1A), and the termination factor Eta [21, 49, 50]. TFS is homologous to TFIIIS in the eukaryotic RNAPII system, and both factors stimulate the intrinsic transcript cleavage activity of RNAP to resolve stalled and backtracked TECs [49, 51]. Spt4/5 is the only RNAP-associated factor that is universally conserved throughout all domains of life [52].

Spt4/5 binds to the RNAP clamp and across the active site cleft forming a tunnel through which the template DNA is threaded towards the catalytic site [15, 50, 53] (Fig. 1C). The smaller Spt4 subunit, specific to archaea and eukaryotes, is entirely comprised of a compact zinc-ribbon domain. Spt4 in conjunction with the C-terminal KOW domain of Spt5 bind close to the reannealed upstream DNA in the RNAP transcription elongation complex (TEC) [15, 53, 54]. The incorporation of Spt4/5 stabilizes the TEC and improves the processivity [22, 52]. Spt4/5 and TFS both facilitate transcription elongation through naked as well as chromatinised and histone-bound DNA [22, 55]. Based on *in vitro* studies, TFE and Spt4/5 interact with RNAP in the context of transcription initiation and elongation, respectively. The binding of TFE and Spt4/5 to RNAP is mutually exclusive, and hence TFE and Spt4/5 have to be swapped during the transition from initiation to elongation (ITC to TEC) [21]. Whether Spt4/5 actively displaces TFE during promoter escape, or whether the release of TFE unmask the binding site of Spt4/5 and that triggers its recruitment remains to be determined. Well aligned with the factor swapping hypothesis, the interference with Spt4/5 (aka DSIF) recruitment to RNAPII using cdk7 inhibitors leads to the retention of TFIIE in early elongation complexes [56].

Can mechanisms determined *in vitro* be extrapolated to live cells?

The widely reported view on archaeal transcription as summarized above is mainly the outcome of *in vitro* experimentation following a reductionist approach using reconstituted transcription systems and thus ignoring the potential diversity of transcription mechanisms that occur *in vivo*. Moreover, these studies were often limited to a relatively small number of strong model promoters controlling viral genes (SSV-T6), stable RNAs (tRNA and ribosomal RNA) or promoters of highly expressed metabolic genes with excellent signal-to-noise ratios. To truly

appreciate the complexities of transcription in living cells and to advance our understanding of specific processes including RNAP recruitment, promoter escape, transcription pausing and processivity, it is useful to characterize the genome-wide occupancy of RNAP and basal transcription factors and the dynamic changes thereof in response to the environment, e.g. during different growth phases and stresses. The most widely used technique to this end is Chromatin Immunoprecipitation followed by next generation sequencing (ChIP-seq). For proteins that translocate along the DNA such as RNAP and elongation factors, the genomic occupancy is a function of the initiation rate, the elongation rate, and processivity. This makes the interpretation of their occupancy profiles more complicated than the localized or even DNA sequence-specific binding of transcription initiation factors including TBP, TFB and TFE [57]. Complementary mapping approaches make use of the RNA signal to map RNAP distribution. Techniques such as Nascent Elongating Transcript (NET)-seq [58], GRO- [59] and PRO-seq [60] all essentially determine the position of the nascent RNA 3'-ends within the TEC with nucleotide- or near-nucleotide resolution. This allows to infer the genome-wide distribution of RNAP at high resolution and with strand-specificity. Importantly, these methods are able to provide insights into the dynamic aspects of actively transcribing RNAP. Additional sequencing methods such as ChIP-exo [61] and PIP-seq [62] can increase the spatial resolution by mapping protein-DNA complex boundaries and the single stranded regions of the transcription bubble at near-nucleotide resolution, respectively .

Genome-wide mapping methods have been extensively applied to the basal transcription machineries of eukaryotes, and comprehensive studies have been published for some bacterial species in *Escherichia coli* and *Bacillus subtilis* [63-65]. Archaea have remained much less explored [40, 66]. Recently, a pioneering multi-omics approach reported on the global TSS

map, genome-wide occupancy profiles of TBP, TFB, Spt4/5 and RNAP, and a transcriptome analysis in the euryarchaeon *M. jannaschii* [40]. The overarching results were encouraging, in as much as many mechanisms based on in vitro studies using a few model promoters were indeed highly relevant for promoters systems-wide, and in vivo. Firstly, the combination of TATA and BRE elements were confirmed to be the principal promoter elements in archaea [40, 67-69]. Secondly, the similarity of the promoter to the TATA/BRE consensus was only weakly correlated to TBP/TFB occupancy at these promoters, which suggest that gene-specific regulators (e.g. Ptr2 [70-72]) and chromatin [9] likely influence TBP/TFB recruitment and thereby contribute to core promoter strength. Thirdly, the study identified Spt4/5 as a general elongation factor that correlated very strongly with RNAP on protein-encoding as well as non-coding RNA TUs. While archaeal and eukaryotic Spt4/5 are recruited to the RNAP proximal to the promoter, the recruitment of its homologue NusG in bacteria occurs later in a stochastic fashion throughout the gene body [63]. In contrast to NusG, its paralogue RfaH in *E. coli* is recruited promoter-proximal via a specific sequence motif termed *ops* [73].

The reason for this discrepancy concerning their recruitment of NusG and Spt4/5 is not understood, but it is tempting to speculate that the Spt4 subunit -which is not conserved in bacteria - enhances the binding of Spt4/5 to the TEC and ensures its early recruitment, possibly by interacting physically with the upstream duplex DNA [54]. The early recruitment of archaeal Spt4/5 supports the notion of a mechanism by which the swapping between TFE and Spt4/5 on the majority of TUs operates genome-wide. A specific subset of promoters enriched in noncoding RNA genes revealed a delayed TSS off-set Spt4/5 recruitment indicative of an alternative modus operandi of escape on some TUs [40]. The study also uncovered potential

gene-internal and antisense promoters which reflects the complex picture of transcription provided by RNA-seq data.

In summary, global mapping analyses have both the power to test whether mechanisms based on reductionist *in vitro* experimentation apply on a wider scale *in vivo* to identify exceptions that serve as generators of hypotheses that can be rigorously tested *in vitro*. However, methods relying on immunoprecipitation of targets are limited to components that are known to the investigator, and leave us with the urgent need to identify regulatory factors *de novo*.

Identification and characterization of novel archaeal transcription factors

To date, all functionally characterized general transcription factors in archaea were identified based on sequence homology with their eukaryotic and bacterial counterparts, including TBP and TFB [74-78], TFE alpha and beta [48, 77-84], Spt4/5 [22, 83], TFS [49, 51, 84]. Sequence searches have identified two further candidates that await experimental vetting and characterisation, archaeal homologues of the RNAPII elongation factor Elf1 [85] and the bacterial elongation factor NusA [6]. Structures of RNAPII TECs show that eukaryotic Elf1 binds proximal to Spt5 NGN domain, which in conjunction seal the downstream region of the DNA-binding channel (Fig. 1C). *In vitro* transcription experiments suggest that yeast Elf1 can negatively or positively influence transcription, dependent whether the DNA is naked, or nucleosome-bound, respectively [54, 86]. Whole genome occupancy profiles demonstrate that Elf1 follows RNAPII through the genome in similar manner to Spt4/5 [87]. In Archaea, homologues of Elf1 were identified in members of TACK and ASGARD superphyla, but its role during transcription remains to be characterised [6, 85]. Bacterial NusA modulates transcription elongation by interacting with RNAP, the nascent RNA and other Nus factors including NusB, E

and G, as well as bacteriophage-encoded proteins including Q and N, which gave the Nus factors their name (N-utilisation substances). Bacterial NusA contains an N-terminal RNAP-binding domain and at least three RNA-binding domains, one S1 (OB fold) and two KH domains, while the archaeal variant encompasses only the two KH domains. Interestingly, the 'missing' NusA N-terminal domain and S1 domains bear a limited sequence and structural similarity to Rpo7/RPB7, a conserved RNAP subunit that is stably associated with all archaeal and eukaryotic RNAPs [88]. Rpo7 and RPB7 interact with the nascent RNA during elongation, and increase transcription processivity [89, 90]. Similar to its bacterial homologue, archaeal NusA binds tightly to RNA *in vitro* [91].

Bioinformatic analysis of the first available archaeal genomes discovered one more candidate for a basal transcription factor in archaea. Mbf1 is the only eukaryotic transcription regulator with a homologue in archaea where it is highly conserved [81, 92, 93]. However, yeast Mbf1 is also associated with the small ribosomal subunit and directly involved in reading-frame maintenance during translation [94-96]. Congruently, the archaeal Mbf1 homologue from *S. solfataricus* binds to the small ribosomal subunit and translating ribosomes [97]. Thus, Mbf1 might not function as a transcription factor in archaea after all.

Besides the bioinformatics approach, relatively little work has been undertaken to isolate transcription factors directly from archaeal biomass and unbiased by sequence similarity to eukaryotic or bacterial factors. The sequence homology-based identification of TBP and TFB occurred in parallel with their partial purification from cell lysates, experiments in which they originally were referred to as aTFA and aTFB, respectively [98, 99]. Additional factors affecting transcription fidelity, termination, pausing or release require activity assays that might be far more challenging to establish. A shining example of this approach is Eta (Euryarchaeal

termination activity), which was purified from *Thermococcus kodakarensis* biomass by incubating fractionated cell lysates with stalled TECs and screening for fractions that re-activated the TEC to resume transcription. Eta was identified using mass-spectrometry, and in contrast to all the proteins discussed above, Eta is not homologous to any eukaryotic or bacterial factors. Eta is a unidirectional (5'->3') ATP-dependent DEAD-box helicase with the unique capability of releasing stalled TECs from the template DNA [100]. As such, Eta is a bona fide transcription termination factor that likely plays a specialised role in DNA damage response. In conclusion, our knowledge of regulatory factors in archaea with auxiliary roles in transcription - and without apparent homology to other well-characterized factors - leaves great room for improvement.

Transcriptome levels: RNAP output versus RNA degradation

With the cost of high throughput sequencing becoming more affordable, an increasing number of RNA-seq studies have been published recently. These have provided genome-wide TSS maps and promoter consensus motifs for a range of archaeal species, and insights into cellular adaptations to changes in growth conditions and environmental stresses. While RNA-seq (and previously microarray) data report on the regulation of gene expression, they do not only reflect the RNAP output. The steady-state level of a given RNA transcript is the sum of its synthesis and decay. RNA decay in archaea is mediated by the archaeal exosome and a complement of RNases including beta-CASP enzymes [101, 102] and affected by features such as the phosphorylation state of the RNA 5'-end, polyadenylation, and RNA binding proteins [103, 104]. Rigorous bioinformatics approaches have to be applied for the stringent characterization of global transcriptome responses in particular concerning data normalization and statistical

verification of significance. The common assumption that whole RNA levels remain largely constant during differential gene expression analysis is not met when a substantial fraction of the RNA population is perturbed. For example, during stresses like oxidative- or heat shock, normalization to total RNA can lead to gross over- or under-estimations of changes in gene expression of specific targets [105]. In the absence of any specific inhibitors of the archaeal RNAP, the more pleiotropic reagent actinomycin D has been used to assess archaeal RNA stability. Actinomycin D is a potent DNA intercalator that preferentially binds to GpC sequences in the minor groove of the DNA template trapping elongating RNA polymerase and thereby blocking RNA synthesis [106]. A combination of actinomycin D treatment and microarray transcriptome analysis has been applied to calculate mRNA decay rates in *Sulfolobus* and *Halobacterium* species [107-109]. Archaeal mRNA half-lives varied substantially but are on average within the order of 5-18 minutes, similar to bacterial mRNA turnover. The differences of mRNA half-lives clearly emphasizes that steady-state level transcriptome analysis only in part reflects RNAP output, this is in addition to general technical limitations inherent in the RNA-sequencing techniques including sequence bias during amplification and linker ligation. Methods such as transient transcriptome sequencing (TT-seq) provide a direct approach to study the output of the transcription machinery independent of RNA degradation. TT-seq is a pulse labelling technique that makes use of the incorporation of uracil analogues into RNA and subsequent isolation of derivatised transcripts, followed by sequencing [110]. The recent publication of *in vivo* thio-uracil labelling of RNA in *S. acidocaldarius* and *H. volcanii* is a big step towards establishing this type of transient transcriptome labelling in archaea [111].

Transcription in context of translation, DNA repair and replication

Transcription is inherently connected to the processing of biological information via both its DNA template and RNA product. The dynamics of transcription elongation - pausing and processivity - directly influence the folding of RNAs [112]. E. g. the vectorial nature of transcription is essential for the co-transcriptional assembly of ribosomal proteins and factors in a specific and predetermined order, and this so-called 'assembly-gradient' is thought to be a universal feature of ribosome assembly in all domains of life [113, 114]. Not only the transcription of structural RNAs, but also of mRNAs is coordinated with downstream processes.

Translation. As in bacteria, RNAP and the first (initiating) ribosome are physically coupled in archaea (Fig. 2) [12] and the identification of polarity in gene expression underlines the functional relevance of transcription-translation coupling in archaea [115]. Ribosomes that are trailing TECs are thought to prevent RNAP backtracking and thereby ensure high levels of processivity. In *E. coli*, the coupling of transcription and translation is reflected in the correlation of RNAP and ribosome velocities. Interestingly, this phenomenon applies in both directions in as much as lowered RNAP elongation rates slow down trailing ribosomes, while antibiotics against or mutations that lower ribosome translation elongation rates also lower transcription elongation rates [116]. *E. coli* NusG has been shown to physically connect RNAPs and ribosomes by binding simultaneously to the RNAP clamp module with its N-terminal NGN domain and to ribosomal protein S10 (aka NusE) via its C-terminal KOW domain [117, 118].

However, several structural studies of the RNAP-ribosome combination, the so-called expressome, show the ribosome, in an orientation relative to RNAP that is not compatible with an adaptor role of NusG [10, 119]. An alternative structure of *E. coli* RNAP with the 30S

ribosomal subunit shows the 30S ribosomal subunit in an orientation that would facilitate recognition of the Shine-Dalgarno sequence during translation initiation, but incompatible with the adaptor role of NusG [120]. Coupling could be multistep process where NusG initially assists the recruitment of the ribosome to the RNAP and the RNAP-ribosome complex subsequently undergoes a conformational transition to the processive form. Such model would partially explain the apparent discrepancies between these studies [120]. Archaeal Spt4/5 could play a similar role as NusG in the coupling of transcription and translation in archaea. The recruitment of *M. jannaschii* Spt4/5 to RNAP does not vary with or correlate with the distance between the TSS and the start codon, i.e. the length of the 5'-UTR, as one may predict from a mechanism where Spt4/5 would recruit the ribosome to the translation start site or *vice versa* [40]. Detailed insights into the mechanisms of coupled transcription and translation in archaea are important to address questions about the fundamental principles of gene expression, but this remains uncharted territory. Other processes including DNA repair and replication may experience more rare and stochastic encounters with the transcription machinery.

DNA repair. RNAPs are efficient sensors of DNA damage and trigger recruitment of the DNA repair machinery in bacteria and eukaryotes during transcription-coupled repair (TCR) [121]. Due to their extremophilic lifestyle, many archaea inevitably experience high rates of DNA damage [122]. It appears therefore likely that archaea might utilise TCR, but decisive experimental evidence is still sparse. While *Thermococcus* RNAP appears to be an efficient DNA-damage sensor [123], the components mediating TCR in bacteria and eukaryotes, Mfd/UvrD and CSA/B, respectively, do not have obvious homologues in archaea. A hallmark feature of transcription coupled-repair is the efficient discrimination between the transcribed and the non-transcribed strand that is reflected in different rates of repair. A recent study into

UV-damage repair in haloarchaea indicated that type of strand-specificity [124], while two studies in *S. solfataricus* did not find such differences [125, 126].

Replication. Elongating RNAPs will encounter replication forks during the S-phase of cell division, and while direct evidence is lacking, collisions are likely to have detrimental effects on genome stability in archaea, similar to bacteria and eukaryotes [127]. Interestingly, highly transcribed genes tend to cluster around the three origins of replication in *Sulfolobus* species [128]. In contrast to *E. coli*, the biased chromosomal localisation of highly-transcribed genes in *Sulfolobus* is unlikely to be related to an increased gene dosage, since the S-phase in the *Sulfolobus* cell cycle is very short and not characterised by multiple replication fork pairs and genome equivalents larger than two, as typical for rapidly dividing *E. coli* cells.

Paucity of archaeal promoter elements

To date, three archaeal promoter elements have been identified with defined roles in transcription initiation, the TATA-, BRE- and the (Inr) initiator elements (Fig. 3). This compares to a range of sigma factor-specific -10, extended -10, and -35 motifs, UP-, discriminator- and core recognition- (CRE) elements in bacteria, and a complement of eukaryotic RNAPII promoter elements including two discrete TFIIB recognition elements, called BREu and BREd, TATA, Inr, motif ten element (MTE), and distal promoter elements (DPE) as well as several others [129]. Both eukaryotic and bacterial promoters use a 'mix and match' solution [129, 130]. Generally, not one promoter contains all elements, and indeed a large fraction of metazoan promoters do not include a readily identifiable TATA box [131]. This is in sharp contrast to archaeal promoters. One reason for the diversity in eukaryotes are the 13-14 TBP-associated factors (TAFs), not

conserved in archaea, which make direct physical contact with the additional promoter elements [132-134].

Notably, the promoter elements BREd (eukaryotes) [135] and CRE (bacteria) [136] evolved around protein-nucleic-acid interactions that appear to be conserved in the archaeal transcription PIC. The existence of additional promoter elements has been discussed in the literature, including a study about an AT-rich region termed the promoter-proximal element (PPE) [137]. The PPE is located within the initially melted region (IMR) of DNA that is formed within the OC. Unlike the BRE, TATA and Inr, the IMR does not show any specific DNA motifs beyond an above genome average AT-content that is thought to enhance DNA melting during OC formation [40, 48] similar to RNAPII promoters [138]. However, a direct correlation between AT-content and promoter strength could not be observed [48].

The combination of RNA-seq, TSS mapping data and genome-wide occupancy of the basal transcription machinery provides the means to identify promoters that challenge our understanding of what constitutes a strong archaeal promoter. E.g. the *M. jannaschii* *rrnA* promoter is arguable one of the strongest promoters, but does not show any detectable TATA-box motif and no apparent TBP/TFB occupancy according to ChIP-seq analyses [40].

Furthermore, the promoter is inactive in *in vitro* transcription reactions containing recombinant RNAP, TBP, TFB and TFE [40]. It is likely that an unknown transcription factor enables the high levels of transcription by the *rrnA* promoter possibly in an unorthodox manner and independently of strong consensus promoter elements including the TATA box.

Taken together, it currently appears that archaeal promoters are less complex and more uniform compared to their bacterial and eukaryotic counterparts. Future studies, in particular

genome-wide mapping techniques, will be required to reveal the full diversity of archaeal promoters.

Mechanism(s) of transcription termination

Transcription termination can proceed via factor-dependent or factor-independent - also referred to as intrinsic - mechanisms (Fig. 4A). Intrinsic termination relies on terminator sequences consisting of a poly-U stretch, often preceded by a short RNA hairpin in the case of bacteria. Typically, intrinsic termination requires transcription pausing caused by a weakened RNA-DNA hybrid. The relative instability of rUU-dAA pairs compared to rAA-dTT as well as their DNA-DNA and RNA-RNA counterparts [139] might have contributed to the fact that poly-U stretches evolved independently as termination signals for different cellular multisubunit RNAPs as well as for phage-like single subunit RNAPs like T7 [140]. Following pausing of RNAP on the poly-U stretch of the terminator, the terminator hairpin invades the DNA cleft of RNAP which leads to a partial opening of the RNAP clamp and subsequent dissociation of the TEC [141].

Factor-dependent transcription termination relies on ATP-dependent helicases including Rho in bacteria and Sen1 in eukaryotes, and the riboexonuclease Rat1/Xrn2 in eukaryotes [141-146].

ChIP-seq analyses in *E. coli* have shown that Rho is recruited to RNAP during early elongation and proximal to the promoter but only interacts with and translocates along the RNA later on before terminating transcription by inducing conformational changes in the RNAP [147].

Interestingly, Rho-dependent termination likely involves allosteric mechanisms that are closely related to intrinsic termination [141-147]. It is unclear to which extent factor-dependent termination is DNA-sequence specific, but Rho has a bias for rut (rho utilization) sites enriched in poly-C stretches that are located 10-100 bp upstream of rho termination sites [148]. In contrast,

transcription termination has not been extensively studied in archaea. In early *in vitro* studies, intrinsic termination of transcription was observed to occur at uridine-rich sequences downstream of TUs in several archaea [149, 150]. Additional studies confirmed that motifs containing 5-8 U residues were sufficient to efficiently mediate transcription termination *in vitro* and *in vivo* [23, 151, 152]. While bacterial intrinsic terminators rely on RNA secondary structures, archaeal RNAPs are able to terminate *in vitro* in their absence [152, 153]. However, it is unlikely that RNA poly-U motifs are the sole explanation for genome-wide transcription termination in archaea, since only a fraction of TUs in the *T. kodakarensis* genome contains six or more T residues at their 3'-ends while these motifs do also occur sometimes within the coding region [151]. This would suggest that (i) additional factors enable termination, including less obvious sequence signatures, undiscovered termination factors or even the DNA topology, keeping in mind that genomes of hyperthermophilic archaea are positively supercoiled, and that coupled transcription-translation might suppress transcription termination. Lastly, it is not inconceivable that PICs bound to promoters located downstream of TUs contribute to the transcription termination of TU upstream of the promoter by providing a roadblock.

In order to obtain a genome-wide and unbiased overview of transcription termination sites (TTSs) in archaea, a high-throughput sequencing method called term-seq has been used to map the RNA 3'-ends in *M. mazei* and *S. acidocaldarius* at single-base pair resolution. TTSs sites were identified for 25% and 53% of all TUs, respectively [154]. Three important questions arise from the term-seq data: (i) Are the 3'-termini *bona fide* TTSs, or were they generated by RNA processing? (ii) Are the identified terminator motifs compatible with the current model of intrinsic termination on poly-U stretches? (iii) Is the lack of identifiable terminators for a large

fraction of TUs due to technical limitations or because of an alternative, possibly factor-dependent termination mechanisms?

Term-seq itself maps all RNA 3' ends irrespective of how they were generated, but the nucleotide-enrichment - i.e. sequence signature - shows stronger biases upstream of the transcript 3' end compared to the downstream positions in the genome (Fig. 4B), compatible with the idea that these sites reflect transcription termination. The technical bias inherent in the term-seq methodology is currently unknown, but one obvious factor that is expected to limit terminator identification is the mRNA abundance. *S. acidocaldarius* genes that are associated with terminators identified by term-seq have significantly higher mRNA levels than TUs for which no terminator could be identified (Fig. 4C). The alignment of the sequences upstream of *M. mazei* and *S. acidocaldarius* TTSs reveal enrichment in U-residues in a window that stretches 16 or 22 nt, respectively (Fig. 4B). All sequences in the small set of archaeal terminators that were characterised *in vivo* or *in vitro* included a stretch of at least four consecutive U residues within the terminal 7-8 nt [151-153]. However, only 30% for *S. acidocaldarius* and 56% of *M. mazei* terminators identified by term-seq have at least four consecutive U residues (Fig. 4D and data not shown). Furthermore, the size of the termination motifs (16 or 22 bp) identified by term-seq is considerably longer than the RNA-DNA hybrid (~8-9 bp) in the RNAP hybrid-binding compartment [154, 155]. The underlying molecular mechanism explaining this discrepancy is unknown, but could involve RNAP slowing down gradually on the upstream segment of the motif, before coming to a halt, pausing and eventually terminating transcription on the last U-residue. Another striking feature of both *M. mazei* and *S. acidocaldarius* term-seq data is the use of multiple terminators, which points to a potential regulatory function of 3'-UTR length. In cases of TUs with multiple terminators, the last terminators show generally a higher U-content with

longer poly-U stretches as compared to the upstream terminators suggesting that they might act as a backstop for sloppy transcription termination (Fig. 4D).

In *E. coli*, efficient transcription termination requires a minimum distance of 30 bp between the stop codon and the intrinsic TTSs [156], a distance that is likely defined by the terminator sequence and the architecture of the RNAP-ribosome complex [10, 157]. Term-seq data for *E. coli* show a similar minimum distance (Fig. 4E) [158]. *S. acidocaldarius* has a 3'UTR length distribution roughly comparable to *E. coli*, but interestingly, the apparent minimal distance between stop codon and TTS is significantly shorter (~ 20bp, Fig. 4E), possibly due to the lack of a requirement for hairpin structures in archaeal terminators [154]. In conclusion, the two available archaeal term-seq data sets challenge our view on transcription termination in archaea, in particular questioning the nature of the terminator motifs and the conservation of termination across different archaeal species. Ultimately, the integration with high-resolution occupancy profiling reflecting active transcription will provide additional clues to corroborate the term-seq data to further dissect and distinguish distinct mechanisms of transcription.

Is anti-sense transcription shaping the archaeal transcriptome?

A consistent feature of all archaeal transcriptomes is the widespread antisense transcription, while its role in the regulation of transcription and beyond is only poorly understood. RNA-seq data from *M. jannaschii* cells did not reveal any significant correlation between sense and antisense transcription, and moreover the vast majority of antisense transcripts were not associated with a defined RNA-5' in the global TSS map [40]. The fact that most short noncoding RNA in archaea is antisense RNA produced in cis makes it unlikely that its primary role is on the level of mRNA turnover. Rather, antisense transcription is likely to negatively interfere with sense

transcription by creating head-to-head collisions of TECs. Antisense transcription could be involved in the chromatin organization of the genome (Fig. 2), or as part of pervasive transcription contributing to genome surveillance by TCR, as proposed for bacteria [159]. Lastly, antisense transcription might be initiated from cryptic promoters without any apparent physiological role due to the high inherent AT-content of the many archaeal genomes, including *M. jannaschii* and *Sulfolobus* species. Cryptic transcription occurs abundantly on AT-rich foreign DNA in *E. coli* and is generally considered more harmful than beneficial for the host [160].

Mechanisms of global regulation

The functional boundaries between components of the basal transcription machinery and gene-specific regulators are not sharp. Bacteria and eukaryotes partition their transcription space by using either alternative sigma factor (sigma-70, sigma-54, ECF etc) holo-RNAP enzymes, or distinct types of RNAPs with their cognate basal transcription factors (e.g. RNAPI, II, III), respectively. Archaea use a single type of RNAP to transcribe all genes, noncoding as well as protein-coding. Some archaeal species, but not all, encode a plethora of TBP and TFB paralogues, similar to the bacterial sigma paradigm for the transcription regulation of subsets of genes [161]. A detailed study of TFB paralogs in archaea has demonstrated that their gene duplication occurred independently multiple times and distinct TFB clades differ in terms of evolutionary rates in different archaeal species [162]. Typically, TFB orthologues belonging to one clade, exhibiting a slower rate of evolution, support 'house-keeping' gene transcription, while other TFB variants, belonging to faster evolving clades, have speciated to acquire specialized functions. This is similar to sigma factor evolution in bacteria. In haloarchaea, however, all TFB paralogs show generally higher rates of evolution. Genome-wide mapping and

genetic knockout analyses of TFB paralogues in *Halobacterium* have demonstrated widely overlapping occupancy profiles indicative of promiscuous promoter utilisation, with only a weak sequence BRE preference [67, 68]. Certain TFB paralogues have a truncated domain architecture which suggests that they depend on additional accessory factors to enable DNA binding. *Sulfolobus* TFB3, which is induced by UV irradiation and is predicted to be part of the DNA damage response, lacks the C-terminal cyclin fold. TFB3 cannot interact with the promoter DNA and hence cannot replace the canonical TFB homologue in a reconstituted transcription system. However, it stimulates transcription *in trans* strictly dependent on the presence of the canonical TFB1 variant [163]. The TFB3 regulon has been mapped using *tfb3* deletion strains combined with RNA-seq, and TUs that are regulated by TFB3 include gene products involved in Ups pili formation and the Ced DNA importer [164, 165]. However, the promoter sequence determinants for TFB3 activity as well as molecular mechanism by which TFB3 stimulates transcription remains opaque.

Following RNAP recruitment and PIC formation the closed initiation complex (CC) is converted into the open complex (OC), which involves conformational rearrangements of RNAP and basal factors as well as DNA strand separation of the IMR promoter region [13]. Several independent mechanisms have evolved that catalyze and regulate the CC-to-OC transition in bacteria including stringent response in *E. coli*, which involves the OC destabilization dependent on the discriminator promoter motif, the nucleotide analogue alarmone ppGpp and the secondary channel binding factor DksA [166]. Notably, such mechanisms do not simply reduce or enhance transcriptional output in a uniform manner. Promoter-specific features such as the stability of intermediates in OC formation [167] can turn a regulator of OC formation into a *bona fide* global transcription regulator. In archaea, OC formation occurs in the minimal PIC consisting of

TBP, TFB and RNAP, but the action of TFE greatly enhances the CC to OC transition and thereby stimulates transcription. TFE might be involved in global transcription regulation in *S.*

solfatarius (Fig. 5) [13]. Key to this hypothesis is the labile [4Fe4S] cluster in the C-terminal domain of the TFEbeta subunit that readily undergoes oxidative damage. As a result, the TFEbeta levels are severely depleted during stress conditions including mild hydrogen peroxide treatment [48]. The magnitude of stimulation by *S. solfataricus* TFE *in vitro* is dependent on the promoter context and critically depends on TFEbeta [48].

While regulatory factors can directly influence the rate of initiation by modulating RNAP recruitment, OC formation, and the escape of RNAP from the promoter, transcription output can also be regulated during the initial elongation phase. Slow rates of early elongation of RNAP proximal to the promoter may indirectly influence the rate of initiation, as well as make the RNAP more prone to pausing and premature termination (Fig. 5). All latter mechanisms alter the dynamics of RNAP transcription and result phenotypically in increased RNAP occupancy at the promoter as compared to intragenic regions in ChIP-seq analyses. But these data alone cannot distinguish between the different molecular mechanisms [57]. Promoter-proximal transcription might be more sensitive to pausing and termination caused by a number of factors that counteract backtracking, which includes of transcription-translation coupling. Different transcription factors interact with the RNAP secondary channel including transcript cleavage factors like Gre in bacteria and TFS in archaea. Cleavage factors act through the RNAP secondary channel and modulate RNAP processivity and dynamics by reactivating backtracked and arrested complexes, which is considered to resolve RNAP traffic jams [168-170]. In archaea, the cleavage factor TFS plays a similar role during elongation, and the *S. solfataricus* TFS factor can be incorporated into the PIC [49] and therefore may have a possible role in regulating

promoter-proximal RNAP dynamics. GreA and DksA, a structural homologue of Gre factors, both influence promoter-proximal RNAP occupancies in *B. subtilis* and *E. coli*, respectively [170, 171].

Notably, many archaea encode multiple TFS paralogues [49] and at least one of them, TFS4, has undergone a dramatic functional specialisation into a potent global inhibitor of RNAP [49]. The domain architecture of TFS4 and the binding to RNAP through the secondary channel is identical to the canonical TFS cleavage factor. In contrast, the catalytic residues have undergone a charge reversal (DE in TFS to KKK in TFS4). As a result, TFS4 has gained the property to both lower the affinity of RNAP for substrate NTPs and to strip RNAPs from their nucleic acids interaction partners in PICs and TECs and thereby efficiently inhibiting transcription in a promoter-independent and thus global fashion (Fig. 5) [49]. TFS4 expression is induced by viral infection and in all likelihood the global transcriptome suppression is part of the antiviral response of the infected cell, however, the precise impact on the host transcriptome has not been elucidated yet. Like bacterial and eukaryotic cells, archaea are likely to respond to their environment by adapting their global gene expression profiles, and that applies a broad range of processes from nutrient shifts to viral infection.

Concluding Remarks

While we have gathered a detailed view on the molecular mechanisms of individual parts of the basal archaeal transcription machinery, many key principles of archaeal transcription remain poorly understood. These big questions include the role of pervasive and antisense transcription, the mechanisms underlying the global responses to environmental changes and stresses, and the functional- and physical coupling of RNA polymerase and transcription to other processes including DNA replication, -repair and translation. Integrating structural and

molecular biology with multiomics approaches will be crucial to unravel these mysteries in a truly multiscale fashion - from the molecular to the systems level.

Acknowledgements

The authors thank all colleagues in the RNAP lab and Dr Tine Arnvig at UCL ISMB. Research in the RNAP laboratory is funded by Wellcome Investigator Award in Science WT 207446/Z/17/Z to Finn Werner.

References

- [1] Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. *Nat Rev Microbiol.* 2017;15:711-23.
- [2] Koonin EV. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos Trans R Soc Lond B Biol Sci.* 2015;370:20140333.
- [3] Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature.* 2013;504:231-6.
- [4] Werner F. Molecular mechanisms of transcription elongation in archaea. *Chem Rev.* 2013;113:8331-49.
- [5] Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol.* 2011;9:85-98.
- [6] Fouqueau T, Blombach F, Cackett G, Carty AE, Matelska DM, Ofer S, et al. The cutting edge of archaeal transcription. *Emerging Topics in Life Sciences.* 2018;2:517.
- [7] Fouqueau T, Blombach F, Werner F. Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annu Rev Microbiol.* 2017;71:331-48.
- [8] Peeters E, Driessen RP, Werner F, Dame RT. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol.* 2015;13:333-41.
- [9] Mattioli F, Bhattacharyya S, Dyer PN, White AE, Sandman K, Burkhart BW, et al. Structure of histone-based chromatin in Archaea. *Science.* 2017;357:609-12.
- [10] Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P. Architecture of a transcribing-translating expressome. *Science.* 2017;356:194-7.
- [11] McGary K, Nudler E. RNA polymerase and the ribosome: the close relationship. *Curr Opin Microbiol.* 2013;16:112-7.
- [12] French SL, Santangelo TJ, Beyer AL, Reeve JN. Transcription and translation are coupled in Archaea. *Mol Biol Evol.* 2007;24:893-5.
- [13] Blombach F, Smollett KL, Grohmann D, Werner F. Molecular Mechanisms of Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol.* 2016;428:2592-606.
- [14] Littlefield O, Korkhin Y, Sigler PB. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *P Natl Acad Sci USA.* 1999;96:13668-73.
- [15] Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* 2011;30:1302-10.

- [16] Todone F, Brick P, Werner F, Weinzierl RO, Onesti S. Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell*. 2001;8:1137-43.
- [17] Meinhart A, Blobel J, Cramer P. An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem*. 2003;278:48267-74.
- [18] Werner F, Weinzierl RO. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell*. 2002;10:635-46.
- [19] Naji S, Grünberg 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. *The Journal of biological chemistry*. 2007;282:11047-57.
- [20] Fouqueau T, Zeller ME, Cheung AC, Cramer P, Thomm M. The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res*. 2013;41:7048-59.
- [21] Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell*. 2011;43:263-74.
- [22] Hirtreiter A, Damsma GE, Cheung AC, Klose D, Grohmann D, Vojnic E, et al. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res*. 2010;38:4040-51.
- [23] 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 research*. 2010;38:585-96.
- [24] Nagy J, Grohmann D, Cheung AC, Schulz S, Smollett K, Werner F, et al. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nature communications*. 2015;6:6161.
- [25] Naji S, Bertero MG, Spitalny P, Cramer P, Thomm M. Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res*. 2008;36:676-87.
- [26] Hilpert R, Winter J, Hammes W, Kandler O. The sensitivity of archaebacteria to antibiotics. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I Abt Originale C: Allgemeine, angewandte und ökologische Mikrobiologie*. 1981;2:11-20.
- [27] Zillig W, Stetter KO, Schnabel R, Madon J, Gierl A. Transcription in Archaeobacteria. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I Abt Originale C: Allgemeine, angewandte und ökologische Mikrobiologie*. 1982;3:218-27.
- [28] Kwak H, Lis JT. Control of transcriptional elongation. *Annu Rev Genet*. 2013;47:483-508.
- [29] Forbes AJ, Patrie SM, Taylor GK, Kim YB, Jiang L, Kelleher NL. Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proc Natl Acad Sci U S A*. 2004;101:2678-83.
- [30] Botting CH, Talbot P, Paytubi S, White MF. Extensive lysine methylation in hyperthermophilic crenarchaea: potential implications for protein stability and recombinant enzymes. *Archaea*. 2010;2010.
- [31] Esser D, Hoffmann L, Pham TK, Brasen C, Qiu W, Wright PC, et al. Protein phosphorylation and its role in archaeal signal transduction. *FEMS Microbiol Rev*. 2016;40:625-47.
- [32] Kort JC, Esser D, Pham TK, Noirel J, Wright PC, Siebers B. A cool tool for hot and sour Archaea: proteomics of *Sulfolobus solfataricus*. *Proteomics*. 2013;13:2831-50.
- [33] Harlen KM, Churchman LS. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol*. 2017;18:263-73.
- [34] Gehring AM, Walker JE, Santangelo TJ. Transcription Regulation in Archaea. *J Bacteriol*. 2016;198:1906-17.
- [35] Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*. 1989;56:549-61.
- [36] Gietl A, Holzmeister P, Blombach F, Schulz S, von Voithenberg LV, Lamb DC, et al. Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res*. 2014;42:6219-31.

- [37] Bell SD, Kosa PL, Sigler PB, Jackson SP. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A*. 1999;96:13662-7.
- [38] Dexl S, Reichelt R, Kraatz K, Schulz S, Grohmann D, Bartlett M, et al. Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res*. 2018;46:10066-81.
- [39] Werner F, Weinzierl RO. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol*. 2005;25:8344-55.
- [40] Smollett K, Blombach F, Reichelt R, Thomm M, Werner F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat Microbiol*. 2017;2:17021.
- [41] Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*. 2006;314:1139-43.
- [42] Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*. 2013;493:437-U191.
- [43] Qureshi SA, Bell SD, Jackson SP. Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J*. 1997;16:2927-36.
- [44] Blombach F, Ausiannikava D, Figueiredo AM, Soloviev Z, Prentice T, Zhang M, et al. Structural and functional adaptation of *Haloferax volcanii* TFE α /beta. *Nucleic Acids Res*. 2018;46:2308-20.
- [45] Sarmiento F, Mrazek J, Whitman WB. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*. 2013;110:4726-31.
- [46] Zhang CY, Phillips APR, Wipfler RL, Olsen GJ, Whitaker RJ. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nature communications*. 2018;9.
- [47] Schulz S, Gietl A, Smollett K, Tinnefeld P, Werner F, Grohmann D. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci U S A*. 2016;113:E1816-25.
- [48] Blombach F, Salvadori E, Fouqueau T, Yan J, Reimann J, Sheppard C, et al. Archaeal TFE α /beta is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife*. 2015;4:e08378.
- [49] Fouqueau T, Blombach F, Hartman R, Cheung ACM, Young MJ, Werner F. The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nature communications*. 2017;8:1914.
- [50] Fouqueau T, Werner F. The architecture of transcription elongation. *Science*. 2017;357:871-2.
- [51] 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.
- [52] 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.
- [53] Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci U S A*. 2011;108:546-50.
- [54] Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M, Sekine SI. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*. 2017;357:921-4.
- [55] Sanders TJ, Lammers M, Marshall CJ, Walker JE, Lynch ER, Santangelo TJ. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol Microbiol*. 2018.
- [56] Laroche S, Amat R, Glover-Cutter K, Sanso M, Zhang C, Allen JJ, et al. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature structural & molecular biology*. 2012;19:1108-15.
- [57] Ehrensberger AH, Kelly GP, Svejstrup JQ. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell*. 2013;154:713-5.
- [58] Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*. 2011;469:368-73.

- [59] Core LJ, Waterfall JJ, Lis JT. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*. 2008;322:1845-8.
- [60] Jordan-Pla A, Perez-Martinez ME, Perez-Ortin JE. Measuring RNA polymerase activity genome-wide with high-resolution run-on-based methods. *Methods*. 2019.
- [61] Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*. 2012;483:295-301.
- [62] Li J, Liu YY, Rhee HS, Ghosh SKB, Bai L, Pugh BF, et al. Kinetic Competition between Elongation Rate and Binding of NELF Controls Promoter-Proximal Pausing. *Molecular Cell*. 2013;50:711-22.
- [63] Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R. Regulator trafficking on bacterial transcription units in vivo. *Mol Cell*. 2009;33:97-108.
- [64] Reppas NB, Wade JT, Church GM, Struhl K. The Transition between Transcriptional Initiation and Elongation in *E. coli* Is Highly Variable and Often Rate Limiting. *Molecular Cell*. 2006;24:747-57.
- [65] Kusuya Y, Kurokawa K, Ishikawa S, Ogasawara N, Oshima T. Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in *Bacillus subtilis* cells. *Journal of bacteriology*. 2011;193:3090-9.
- [66] Wojtas MN, Mogni M, Millet O, Bell SD, Abrescia NGA. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Research*. 2012;40:9941-52.
- [67] Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, Bonneau R, et al. General transcription factor specified global gene regulation in archaea. *Proc Natl Acad Sci U S A*. 2007;104:4630-5.
- [68] Seitzer P, Wilbanks EG, Larsen DJ, Facciotti MT. A Monte Carlo-based framework enhances the discovery and interpretation of regulatory sequence motifs. *BMC Bioinformatics*. 2012;13:317.
- [69] Wilbanks EG, Larsen DJ, Neches RY, Yao AI, Wu CY, Kjolby RA, et al. A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. *Nucleic Acids Res*. 2012;40:e74.
- [70] Ouhammouch M, Dewhurst RE, Hausner W, Thomm M, Geiduschek EP. Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc Natl Acad Sci U S A*. 2003;100:5097-102.
- [71] Ouhammouch M, Werner F, Weinzierl RO, Geiduschek EP. A fully recombinant system for activator-dependent archaeal transcription. *The Journal of biological chemistry*. 2004;279:51719-21.
- [72] Wilkinson SP, Ouhammouch M, Geiduschek EP. Transcriptional activation in the context of repression mediated by archaeal histones. *Proc Natl Acad Sci U S A*. 2010;107:6777-81.
- [73] Artsimovitch I, Landick R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell*. 2002;109:193-203.
- [74] Hausner W, Wettach J, Hethke C, Thomm M. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *The Journal of biological chemistry*. 1996;271:30144-8.
- [75] Qureshi SA, Khoo B, Baumann P, Jackson SP. Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc Natl Acad Sci U S A*. 1995;92:6077-81.
- [76] Rowlands T, Baumann P, Jackson SP. The TATA-binding protein: a general transcription factor in eukaryotes and archaeobacteria. *Science*. 1994;264:1326-9.
- [77] Creti R, Londei P, Cammarano P. Complete nucleotide sequence of an archaeal (*Pyrococcus woesei*) gene encoding a homolog of eukaryotic transcription factor IIB (TFIIB). *Nucleic Acids Res*. 1993;21:2942.
- [78] Ouzounis C, Sander C. TFIIB, an evolutionary link between the transcription machineries of archaeobacteria and eukaryotes. *Cell*. 1992;71:189-90.
- [79] Bell SD, Brinkman AB, van der Oost J, Jackson SP. The archaeal TFIIE α homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep*. 2001;2:133-8.

- [80] Hanzelka BL, Darcy TJ, Reeve JN. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFII α . *J Bacteriol.* 2001;183:1813-8.
- [81] Aravind L, Koonin EV. DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.* 1999;27:4658-70.
- [82] Blombach F, Makarova KS, Marrero J, Siebers B, Koonin EV, van der Oost J. Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol Direct.* 2009;4:39.
- [83] Ponting CP. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* 2002;30:3643-52.
- [84] Langer D, Zillig W. Putative tfls 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 Res.* 1993;21:2251.
- [85] Daniels JP, Kelly S, Wickstead B, Gull K. Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol Direct.* 2009;4:24.
- [86] Ehara H, Kujirai T, Fujino Y, Shirouzu M, Kurumizaka H, Sekine SI. Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science.* 2019;363:744-7.
- [87] Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol.* 2010;17:1272-+.
- [88] Belogurov GA, Artsimovitch I. Regulation of Transcript Elongation. *Annu Rev Microbiol.* 2015;69:49-69.
- [89] 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;38:585-96.
- [90] Ujvari A, Luse DS. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol.* 2006;13:49-54.
- [91] Shibata R, Bessho Y, Shinkai A, Nishimoto M, Fusatomi E, Terada T, et al. Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem Biophys Res Commun.* 2007;355:122-8.
- [92] Takemaru K, Harashima S, Ueda H, Hirose S. Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation. *Mol Cell Biol.* 1998;18:4971-6.
- [93] Takemaru K, Li FQ, Ueda H, Hirose S. Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. *Proc Natl Acad Sci U S A.* 1997;94:7251-6.
- [94] Opitz N, Schmitt K, Hofer-Pretz V, Neumann B, Krebber H, Braus GH, et al. Capturing the Asc1p/Receptor for Activated C Kinase 1 (RACK1) Microenvironment at the Head Region of the 40S Ribosome with Quantitative BioID in Yeast. *Mol Cell Proteomics.* 2017;16:2199-218.
- [95] Wang J, Zhou J, Yang Q, Grayhack EJ. Multi-protein bridging factor 1(Mbf1), Rps3 and Asc1 prevent stalled ribosomes from frameshifting. *Elife.* 2018;7.
- [96] Hendrick JL, Wilson PG, Edelman, II, Sandbaken MG, Ursic D, Culbertson MR. Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis. *Genetics.* 2001;157:1141-58.
- [97] Blombach F, Launay H, Snijders AP, Zorraquino V, Wu H, de Koning B, et al. Archaeal MBF1 binds to 30S and 70S ribosomes via its helix-turn-helix domain. *Biochem J.* 2014;462:373-84.
- [98] Frey G, Thomm M, Brudigam B, Gohl HP, Hausner W. An archaeobacterial cell-free transcription system. The expression of tRNA genes from *Methanococcus vanniellii* is mediated by a transcription factor. *Nucleic Acids Res.* 1990;18:1361-7.
- [99] Hausner W, Thomm M. Purification and characterization of a general transcription factor, aTFB, from the archaeon *Methanococcus thermolithotrophicus*. *The Journal of biological chemistry.* 1993;268:24047-52.
- [100] Walker JE, Luyties O, Santangelo TJ. Factor-dependent archaeal transcription termination. *Proc Natl Acad Sci U S A.* 2017;114:E6767-E73.

- [101] Clouet-d'Orval B, Batista M, Bouvier M, Quentin Y, Fichant G, Marchfelder A, et al. Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol Rev.* 2018;42:579-613.
- [102] Phung DK, Rinaldi D, Langendijk-Genevaux PS, Quentin Y, Carpousis AJ, Clouet-d'Orval B. Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res.* 2013;41:1091-103.
- [103] Hasenohrl D, Lombo T, Kaberdin V, Londei P, Blasi U. Translation initiation factor α /eIF2(- γ) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proc Natl Acad Sci U S A.* 2008;105:2146-50.
- [104] Mura C, Randolph PS, Patterson J, Cozen AE. Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biol.* 2013;10:636-51.
- [105] Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol.* 2015;36:662-7.
- [106] Goldberg IH, Rabinowitz M, Reich E. Basis of actinomycin action. I. DNA binding and inhibition of RNA-polymerase synthetic reactions by actinomycin. *Proc Natl Acad Sci U S A.* 1962;48:2094-101.
- [107] Evgenieva-Hackenberg E, Wagner S, Klug G. In vivo and in vitro studies of RNA degrading activities in Archaea. *Methods Enzymol.* 2008;447:381-416.
- [108] Hundt S, Zaigler A, Lange C, Soppa J, Klug G. Global analysis of mRNA decay in *Halobacterium salinarum* NRC-1 at single-gene resolution using DNA microarrays. *J Bacteriol.* 2007;189:6936-44.
- [109] Andersson AF, Lundgren M, Eriksson S, Rosenlund M, Bernander R, Nilsson P. Global analysis of mRNA stability in the archaeon *Sulfolobus*. *Genome Biol.* 2006;7:R99.
- [110] Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, et al. TT-seq maps the human transient transcriptome. *Science.* 2016;352:1225-8.
- [111] Knuppel R, Kuttenger C, Ferreira-Cerca S. Toward Time-Resolved Analysis of RNA Metabolism in Archaea Using 4-Thiouracil. *Front Microbiol.* 2017;8:286.
- [112] Zhang J, Landick R. A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci.* 2016;41:293-310.
- [113] Ferreira-Cerca S. Life and Death of Ribosomes in Archaea. In: Clouet-d'Orval B, editor. *RNA Metabolism and Gene Expression in Archaea*. Cham: Springer International Publishing; 2017. p. 129-58.
- [114] Nierhaus KH. Assembly Of The Prokaryotic Ribosome. In: Nierhaus KH, Wilson DN, editors. *Protein synthesis and ribosome structure : translating the genome*. Weinheim: Wiley-VCH; 2004. p. 85-105.
- [115] Santangelo TJ, Cubonova L, Matsumi R, Atomi H, Imanaka T, Reeve JN. Polarity in archaeal operon transcription in *Thermococcus kodakaraensis*. *J Bacteriol.* 2008;190:2244-8.
- [116] Proshkin S, Rahmouni AR, Mironov A, Nudler E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science.* 2010;328:504-8.
- [117] Saxena S, Myka KK, Washburn R, Costantino N, Court DL, Gottesman ME. *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol Microbiol.* 2018;108:495-504.
- [118] Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, et al. A NusE:NusG complex links transcription and translation. *Science.* 2010;328:501-4.
- [119] Fan H, Conn AB, Williams PB, Diggs S, Hahm J, Gamper HB, Jr., et al. Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res.* 2017;45:11043-55.
- [120] Demo G, Rasouly A, Vasilyev N, Svetlov V, Loveland AB, Diaz-Avalos R, et al. Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife.* 2017;6.
- [121] Portman JR, Strick TR. Transcription-Coupled Repair and Complex Biology. *J Mol Biol.* 2018;430:4496-512.
- [122] Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* 1993;362:709-15.

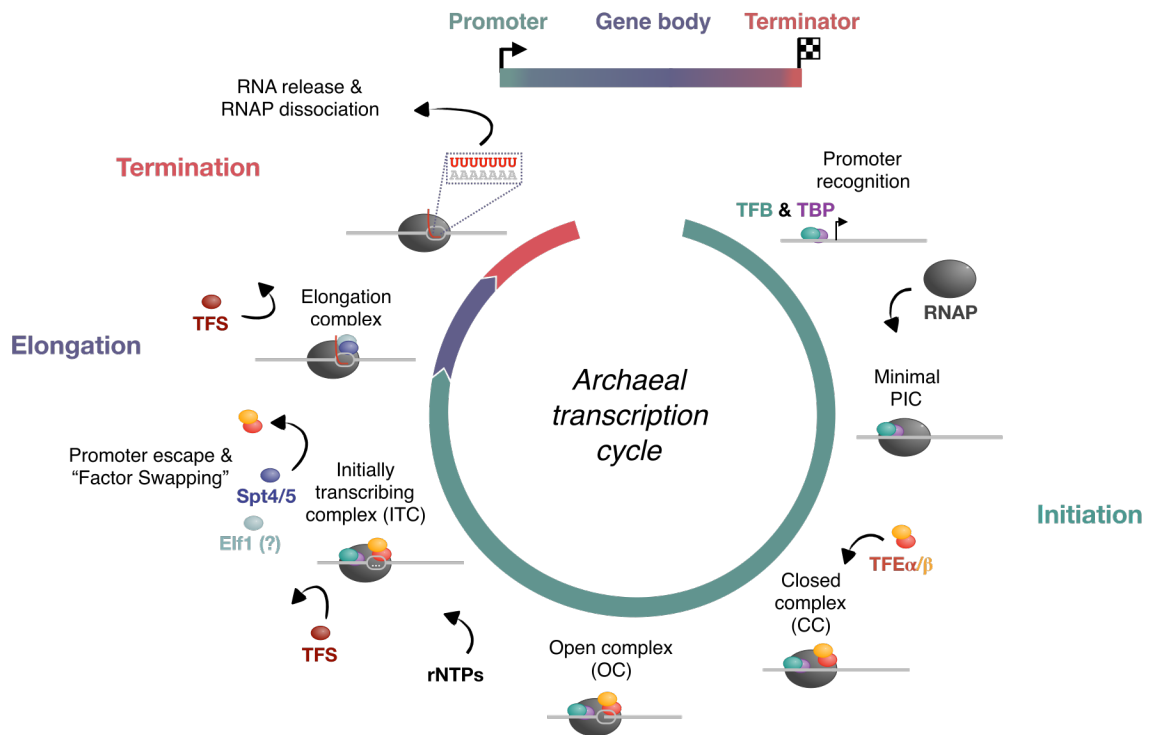
- [123] Gehring AM, Santangelo TJ. Archaeal RNA polymerase arrests transcription at DNA lesions. *Transcription*. 2017;8:288-96.
- [124] Stantial N, Dumpe J, Pietrosimone K, Baltazar F, Crowley DJ. Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst)*. 2016;41:63-8.
- [125] Dorazi R, Gotz D, Munro S, Bernander R, White MF. Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfataricus*. *Mol Microbiol*. 2007;63:521-9.
- [126] 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;365:921-9.
- [127] McGlynn P, Savery NJ, Dillingham MS. The conflict between DNA replication and transcription. *Molecular Microbiology*. 2012;85:12-20.
- [128] Andersson AF, Pelve EA, Lindeberg S, Lundgren M, Nilsson P, Bernander R. Replication-biased genome organisation in the crenarchaeon *Sulfolobus*. *BMC Genomics*. 2010;11:454.
- [129] Vo Ngoc L, Wang YL, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. *Genes Dev*. 2017;31:1289-301.
- [130] Hook-Barnard IG, Hinton DM. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul Syst Bio*. 2007;1:275-93.
- [131] Donczew R, Hahn S. Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters. *Mol Cell Biol*. 2018;38.
- [132] Malkowska M, Kokoszynska K, Rychlewski L, Wyrwicz L. Structural bioinformatics of the general transcription factor TFIID. *Biochimie*. 2013;95:680-91.
- [133] Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, et al. The architecture of human general transcription factor TFIID core complex. *Nature*. 2013;493:699-702.
- [134] Joo YJ, Ficarro SB, Soares LM, Chun Y, Marto JA, Buratowski S. Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes Dev*. 2017;31:2162-74.
- [135] Deng WS, Roberts SGE. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Gene Dev*. 2005;19:2418-23.
- [136] Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural Basis of Transcription Initiation. *Science*. 2012;338:1076-80.
- [137] Reiter WD, Hudepohl U, Zillig W. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. *Proc Natl Acad Sci U S A*. 1990;87:9509-13.
- [138] Dienemann C, Schwalb B, Schilbach S, Cramer P. Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Mol Cell*. 2018.
- [139] Sugimoto N, Nakano S, Katoh M, Matsumura A, Nakamuta H, Ohmichi T, et al. Thermodynamic Parameters to Predict Stability of Rna/DNA Hybrid Duplexes. *Biochemistry-U S*. 1995;34:11211-6.
- [140] Jeng ST, Gardner JF, Gumport RI. Transcription termination in vitro by bacteriophage T7 RNA polymerase. The role of sequence elements within and surrounding a rho-independent transcription terminator. *The Journal of biological chemistry*. 1992;267:19306-12.
- [141] Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, Nudler E. An allosteric path to transcription termination. *Mol Cell*. 2007;28:991-1001.
- [142] Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedeja E, Greenblatt JF, et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*. 2004;432:517-22.
- [143] West S, Gromak N, Proudfoot NJ. Human 5' \rightarrow 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*. 2004;432:522-5.
- [144] Steinmetz EJ, Brow DA. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol*. 1996;16:6993-7003.
- [145] El Hage A, Koper M, Kufel J, Tollervey D. Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev*. 2008;22:1069-81.

- [146] Han Z, Libri D, Porrua O. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res.* 2016.
- [147] Epshtein V, Dutta D, Wade J, Nudler E. An allosteric mechanism of Rho-dependent transcription termination. *Nature.* 2010;463:245-9.
- [148] Ciampi MS. Rho-dependent terminators and transcription termination. *Microbiology.* 2006;152:2515-28.
- [149] Muller B, Allmansberger R, Klein A. Termination of a transcription unit comprising highly expressed genes in the archaeobacterium *Methanococcus voltae*. *Nucleic Acids Res.* 1985;13:6439-45.
- [150] Thomm M, Hausner W, Hethke C. Transcription Factors and Termination of Transcription in *Methanococcus*. *Syst Appl Microbiol.* 1994;16:648-55.
- [151] Santangelo TJ, Cubonova L, Skinner KM, Reeve JN. Archaeal intrinsic transcription termination in vivo. *J Bacteriol.* 2009;191:7102-8.
- [152] Santangelo TJ, Reeve JN. Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol.* 2006;355:196-210.
- [153] Spitalny P, Thomm M. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol Microbiol.* 2008;67:958-70.
- [154] Dar D, Prasse D, Schmitz RA, Sorek R. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nature Microbiology.* 2016;1.
- [155] Washio T, Sasayama J, Tomita M. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res.* 1998;26:5456-63.
- [156] Li R, Zhang Q, Li JB, Shi HL. Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. *Nucleic Acids Research.* 2016;44:2554-63.
- [157] Castro-Roa D, Zenkin N. In vitro experimental system for analysis of transcription-translation coupling. *Nucleic Acids Res.* 2012;40:e45.
- [158] Dar D, Sorek R. High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res.* 2018;46:6797-805.
- [159] Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol.* 2014;12:647-53.
- [160] Lamberte LE, Baniulyte G, Singh SS, Stringer AM, Bonocora RP, Stracy M, et al. Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat Microbiol.* 2017;2:16249.
- [161] Baliga NS, Goo YA, Ng WV, Hood L, Daniels CJ, DasSarma S. Is gene expression in *Halobacterium NRC-1* regulated by multiple TBP and TFB transcription factors? *Mol Microbiol.* 2000;36:1184-5.
- [162] Petitjean C, Makarova KS, Wolf YI, Koonin EV. Extreme Deviations from Expected Evolutionary Rates in Archaeal Protein Families. *Genome Biol Evol.* 2017;9:2791-811.
- [163] Paytubi S, White MF. The crenarchaeal DNA damage-inducible transcription factor B paralogue TFB3 is a general activator of transcription. *Mol Microbiol.* 2009;72:1487-99.
- [164] Feng X, Sun M, Han W, Liang YX, She Q. A transcriptional factor B paralog functions as an activator to DNA damage-responsive expression in archaea. *Nucleic Acids Res.* 2018;46:7465.
- [165] Schult F, Le TN, Albersmeier A, Rauch B, Blumenkamp P, van der Does C, et al. Effect of UV irradiation on *Sulfolobus acidocaldarius* and involvement of the general transcription factor TFB3 in the early UV response. *Nucleic Acids Res.* 2018;46:7179-92.
- [166] Potrykus K, Cashel M. (p)ppGpp: still magical? *Annu Rev Microbiol.* 2008;62:35-51.
- [167] Barker MM, Gaal T, Josaitis CA, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol.* 2001;305:673-88.
- [168] Borukhov S, Sagitov V, Goldfarb A. Transcript cleavage factors from *E. coli*. *Cell.* 1993;72:459-66.

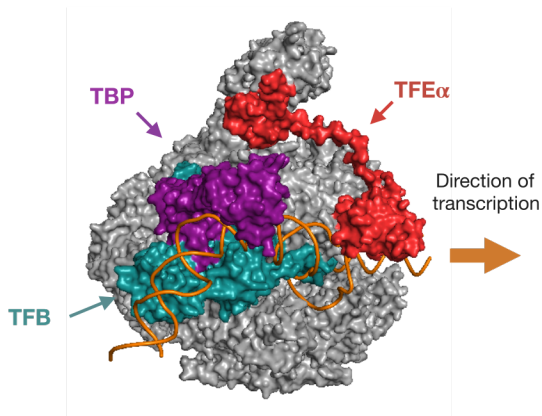
- [169] Cheung AC, Cramer P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. 2011;471:249-53.
- [170] Zhang Y, Mooney RA, Grass JA, Sivaramakrishnan P, Herman C, Landick R, et al. DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest. *Mol Cell*. 2014;53:766-78.
- [171] Kusuya Y, Kurokawa K, Ishikawa S, Ogasawara N, Oshima T. Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in *Bacillus subtilis* cells. *J Bacteriol*. 2011;193:3090-9.
- [172] Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, et al. Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS biology*. 2009;7:e102.
- [173] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-12.
- [174] Dar D, Prasse D, Schmitz RA, Sorek R. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat Microbiol*. 2016;1:16143.
- [175] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188-90.
- [176] Cohen O, Doron S, Wurtzel O, Dar D, Edelheit S, Karunker I, et al. Comparative transcriptomics across the prokaryotic tree of life. *Nucleic Acids Research*. 2016;44:W46-W53.
- [177] Dar D, Shamir M, Mellin JR, Koutero M, Stern-Ginossar N, Cossart P, et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*. 2016;352.
- [178] Sheppard C, Blombach F, Belsom A, Schulz S, Daviter T, Smollett K, et al. Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP. *Nat Commun*. 2016;7:13595.

Figures

A

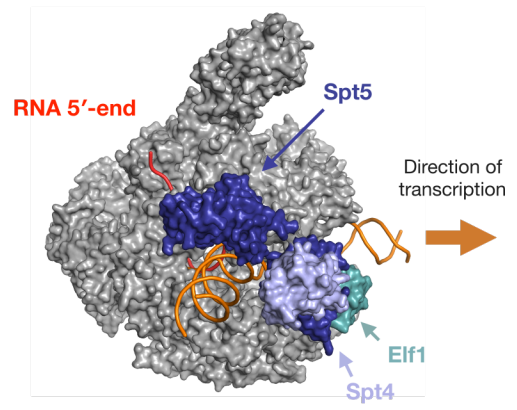


B



PIC

C



TEC

Figure 1: Molecular mechanisms of the archaeal transcription cycle. (A) Transcription of a gene is initiated at the promoter, proceeds through the gene body and comes to completion at the terminator. RNAP makes repeatedly use of the template by progressing repeatedly through the cycle, and specific subset of initiation, elongation and possibly termination factors assist RNAP

during this process. These mechanisms are described in detail in the text. Atomic-resolution structural information on the RNAP transcription pre-initiation complexes (PIC) and RNAP transcription elongation complexes (TEC) has greatly contributed to our understanding of RNAP function. The structural model of the archaeal PIC (DNA-TBP-TFB-RNAP-TFEalpha) is derived from [13]. The structural model of the archaeal TEC (DNA/RNA-RNAP-Spt4/5-Elf1) was generated using Chimera. The *Sulfolobus shibatae* RNAP (PDB: 2WAQ) was aligned to RNAP II of the eukaryotic TEC (PDB: 5XOG, 5XON) [54, 172, 173].

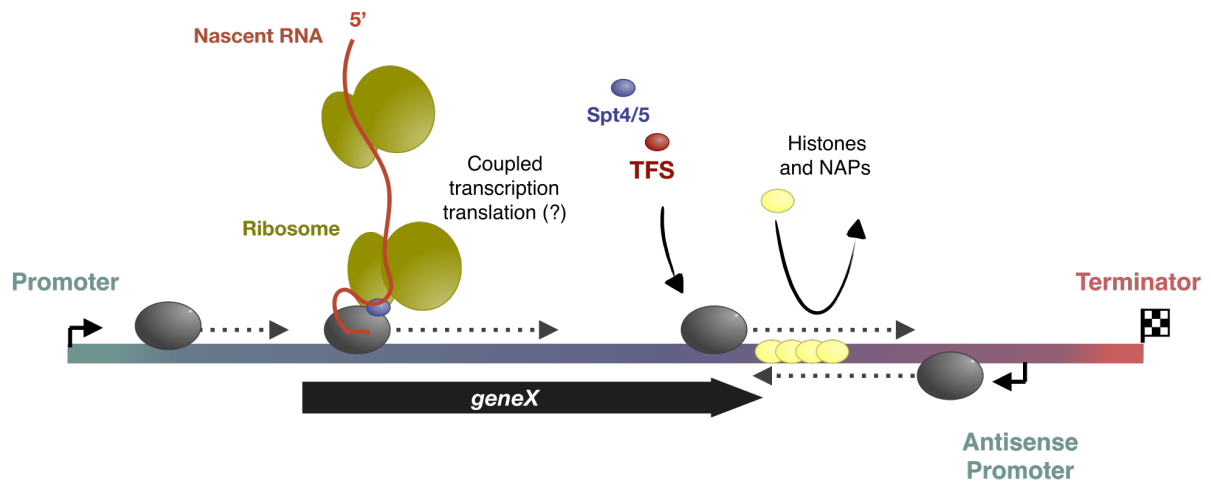


Figure 2: Archaeal transcription in context. The properties of RNAP transcription is influenced by several additional processes that handle biological information in the cell. Transcription and translation are coupled in archaea, and the RNAP and the initiating ribosome are possibly coupled by Spt4/5. The archaeal genome is chromatinised by histones and other nucleoid associated proteins (NAPs); elongation factors like Spt4/5 and TFS enhance transcription of archaeal RNAP through chromatin. Pervasive and antisense transcription have the potential to interfere with transcription by RNAP collisions between convergent TEC.

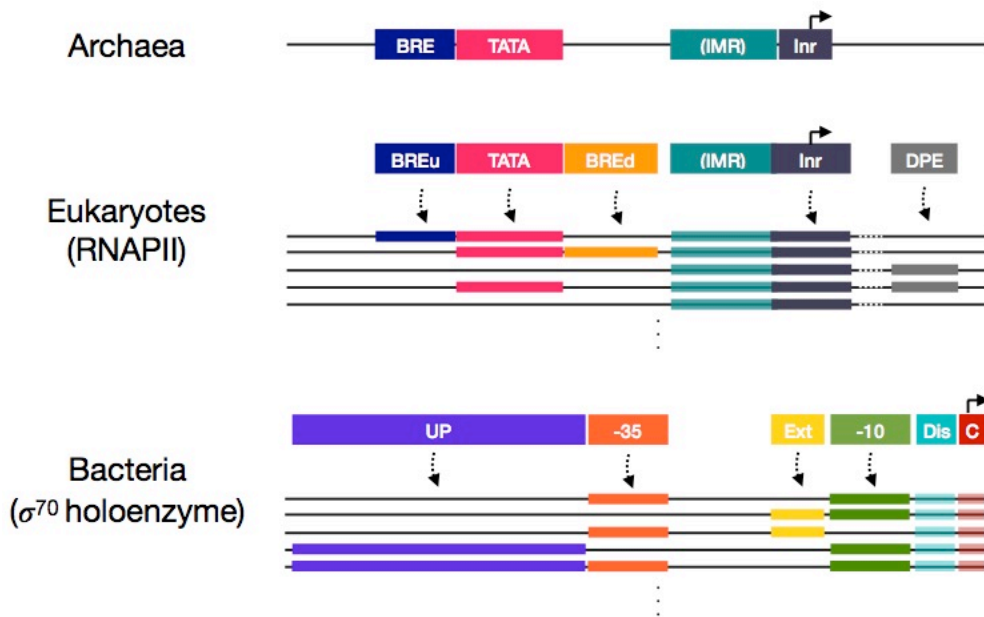


Figure 3: Promoter consensus elements in the three domains. The paucity and uniformity of archaeal DNA elements seems in contrast to both bacterial and eukaryotic RNAPII promoters. Archaeal promoter elements comprise the BRE and TATA box as well as the initiator (Inr). The initially melted region (IMR) is enriched in A and T residues but does not have a bona fide sequence motif. RNAPII core promoter elements comprise in addition to BRE and TATA, the downstream promoter element (DPE), two distinct BRE motifs upstream and downstream of the TATA-box (BREu and BREd) as well as a range of other more varied and gene specific elements (reviewed in [129]). Bacterial promoters include, in addition to the canonical -35/-10 elements, UP elements (UP), extended -35 and -10 regions (-35, Ext, -10), the discriminator -(Dis) and the core recognition element (C). Examples for the 'mix and match' solution of promoter architecture of RNAPII and bacterial RNAP (Sigma70) are depicted below for a subset of promoter element shown in full color (ignoring IMR, Dis and C presence/absence).

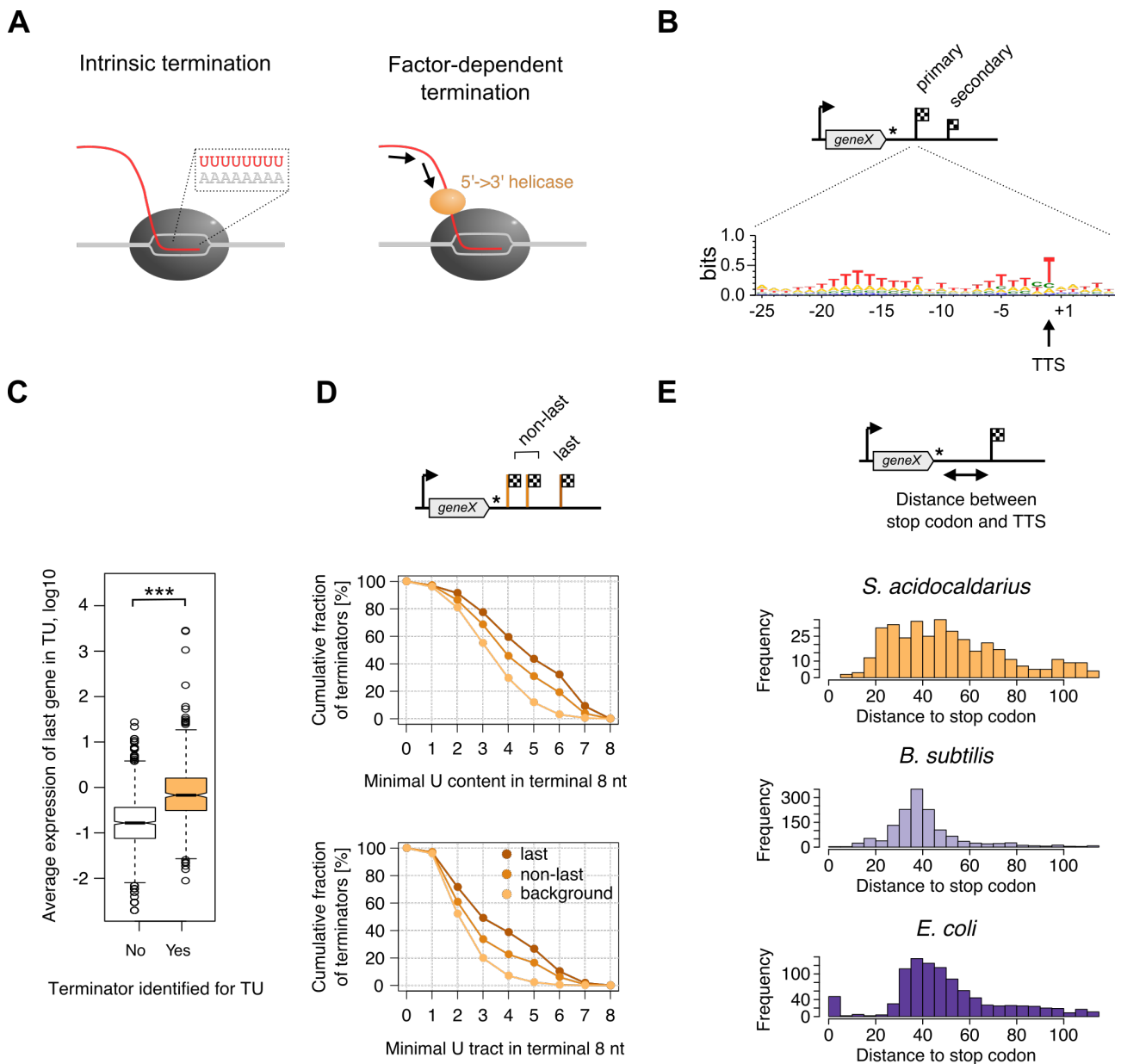


Figure 4: Transcription termination in archaea in the era of term-seq. (A) Schematic illustrations of intrinsic and factor-dependent transcription termination. (B) DNA sequence logo of the primary terminator context in *S. acidocaldarius* based on the terminators with highest signal for each TU [174]. The logo was created with WebLogo 3 [175] and corrected for the genomic background nucleotide frequencies. (C) The term-seq data correlate with RNA expression levels in *S. acidocaldarius* [176]. The RNA-seq coverage of the last genes encoded in multicistronic TUs with identified terminators (n=964) [174] is higher than for TUs without identified

terminators (n=499) (Wilcoxon rank sum test, $p < 0.001$). (D) Term-seq has identified that many archaeal TUs encode multiple terminators. Terminator U-content and U-stretch length is higher for the last terminator compared to the upstream terminators, suggesting that they function as a backstop. Plots depicting the (cumulative) percentage of terminators as a function of a given minimum number of U residues within the last (terminal) eight nucleotides (top panel), or a given minimum length of consecutive U residues (bottom panel). (E) The distance distribution between stop codons and TTSs in *S. acidocaldarius* [174] shows avoidance of distances < 20 nt. This limit is considerably shorter than that for *B. subtilis* [177] and *E. coli* [158], which likely reflects the absence of hairpin RNA secondary structures in archaeal terminators, or possible differences in the topology of the RNAP-ribosome complexes between bacteria and archaea.

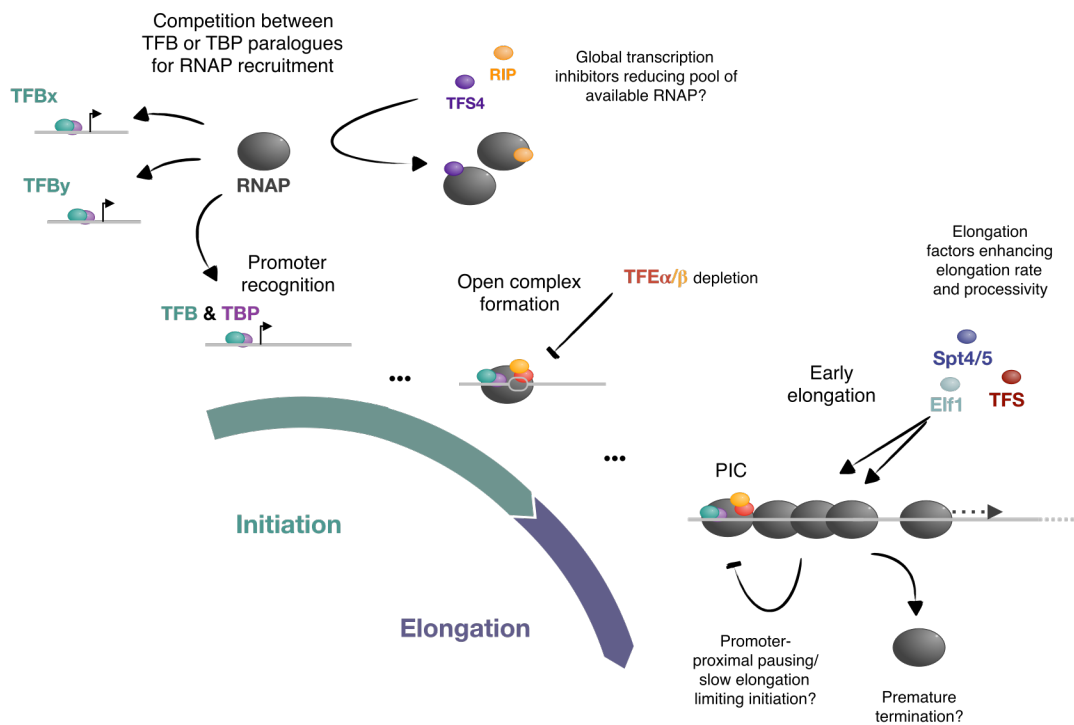


Figure 5: Putative mechanisms of global transcription regulation in archaea. This figure highlights how distinct steps in the progression of RNAP through the transcription cycle are leveraged to regulate transcription on a global and gene-specific level. Promoter-specific binding of combinations of multiple TBP and TFB paralogues may confer gene-specific regulation similar to the sigma paradigm in bacteria. TFE stimulates OC formation, and the depletion of TFE during environmental stresses [48] may globally reduce transcription output while the magnitude of this effect differs between promoters. Pausing or slow elongation during [170]the early elongation phase can indirectly reduce the promoter firing rate by negatively interfering with promoter escape. Premature transcription termination can reduce the amount of RNA synthesised. Both processes could be influenced by elongation factors such as Spt4/5, Elf1 and TFS. Lastly, RNAP engaged in transcription initiation and elongation complexes can be removed entirely from the transcription cycle by global RNAP inhibitors encoded by viruses (e.g. ATV-RIP) or the host (e.g. TFS4) [49, 178].