PORTLAND PRESS

Review Article

The cutting edge of archaeal transcription

Thomas Fouqueau, Fabian Blombach, Gwenny Cackett, Alice E. Carty, Dorota M. Matelska, Sapir Ofer, Simona Pilotto, Duy Khanh Phung and Finn Werner

RNAP laboratory, Institute of Structural and Molecular Biology, Division of Biosciences, University College London, Gower Street, London WC1E 6BT, U.K.

Correspondence: Finn Werner (f.werner@ucl.ac.uk)

OPEN ACCESS

The archaeal RNA polymerase (RNAP) is a double-psi β-barrel enzyme closely related to eukaryotic RNAPII in terms of subunit composition and architecture, promoter elements and basal transcription factors required for the initiation and elongation phase of transcription. Understanding archaeal transcription is, therefore, key to delineate the universally conserved fundamental mechanisms of transcription as well as the evolution of the archaeo-eukaryotic transcription machineries. The dynamic interplay between RNAP subunits, transcription factors and nucleic acids dictates the activity of RNAP and ultimately gene expression. This review focusses on recent progress in our understanding of (i) the structure, function and molecular mechanisms of known and less characterized factors including Elf1 (Elongation factor 1), NusA (N-utilization substance A), TFS4, RIP and Eta, and (ii) their evolution and phylogenetic distribution across the expanding tree of Archaea.

Introduction

Transcription — the DNA template-dependent synthesis of RNA — is essential to life. The overall structure of the molecular machine that drives transcription, RNA polymerase (RNAP), is universally conserved in all domains of life, including Bacteria, Archaea and Eukarya. But whereas bacteria and archaea use a single RNAP to transcribe all genes, eukaryotes have compartmentalized the transcription space into distinct subsets of genes that are transcribed by three and five different enzymes in animals and plants, respectively. Most features of archaeal transcription — including the RNAP, general transcription factors that govern its activities and the DNA sequence elements with which they interact — are closely related to the eukaryotic RNAPII system. The archaeal transcription apparatus is likely to resemble the ancestral version of eukaryotic RNAPII and thus worthy of our attention not only because it is interesting in its own right, but also because it serves as highly tractable and thus extremely valuable model system.

Archaea are prokaryotic organisms that occupy a key position in the tree of life. The development of culture-independent sequencing techniques highlighted the abundance of archaea in diverse environments such as soils, deep-sea sediments and hydrothermal systems. Archaea are also well-recognized components of the human microbiome and provide a broader view on biodiversity. To date, archaea comprise at least four major superphyla, each of which comprises various phyla: Euryarchaeota (subdivided into group I and II), DPANN (Diapherotrites, Parv-, Aenigm-, Nano-, Nanohaloarchaeota, and others), TACK (Thaum-, Aig-, Cren-, Kor- and Bathyarchaeota) and ASGARD (Loki-, Odin-, Thor- and Heimdallarchaeota) [1–5]. Genetically very diverse, archaea use a single type of RNAP to transcribe all genes. However, lineage-specific RNAP subunits, such as Rpo8 and Rpo13, shed light on the acquisition of transcription function during evolution.

Architecture and function of the archaeal RNAP subunits

All cellular RNAPs share a subunit core whose ancestry predates the last universal common ancestor and thus the diversification into the lineages that have evolved into extant bacteria, archaea and eukaryotes [6]. The RNAP core is formed by five universally conserved subunits (Rpo1, 2, 3, 6 and 11 in archaea) and contains, in principle, all critical elements for transcription. In addition, RNAP subunits

Table 1 Evolutionary conservation of RNAP subunits and general transcription factors

Table summarizes the archaeal RNAP subunits, transcription initiation- and elongation factors, and indicates the homologous components in bacteria and eukaryotes. The column on the right indicates the molecular functions discussed in detail in the text. Note that the bacterial sigma-70 factor is functionally analogous to the TBP/TFB duo, while only sharing a very limited sequence similarity with TFB. Bacterial Gre factors are functionally analogous, but not homologous, to TFS/TFIIS transcript cleavage factors.

	Bacteria	Archaea		Eukarya	function
		Eury-/DPANN	TACK/ASGARD	RNAPII	
Core Subunits	β'	Rpo1		RPB1	DPBB-1, catalysis (Asp-triad)
	β	Rpo2		RPB2	DPBB-2, catalysis (lysines)
	α_I	Rpo3		RPB3	Assembly platform
	α_{II}	Rpo11		RPB11	
	ω	Rpo6		RPB6	RNAP assembly and stability
Auxiliary Subunits		Rpo5		RPB5	dsDNA binding
		Rpo10		RPB10	Assembly platform
		Rpo12		RPB12	
		Rpo4		RPB4	RNA-binding (processivity and termination)
		Rpo7		RPB7	
			Rpo8*	RPB8	RNA binding (backtracking?)
				RPB9	
			Rpo13*		dsDNA binding
Initiation factors	σ^{70}	TBP		TBP	TATA-box binding
		TFB		TFIIB	BRE binding, RNAP recruitment
		TFEα		TFIIEα	DNA melting and OC formation
			TFEβ*	TFIIEβ	
Elongation factors	Gre	TFS**		TFIIS	Transcript cleavage
		Spt4		Spt4	Antipausing/processivity
	NusG	Spt5		Spt5	
			Elf1*	Elf1	Elongation (?)
	NusA	NusA			Elongation/pausing (?)

*Only found in some species.

**Archaeal TFS is evolutionarily related to RNAPII subunit RPB9 and to the transcript cleavage factor TFIIS in eukaryotes.

not conserved in bacteria play important roles for the assembly and stability of RNAP (Rpo10 and 12), its interactions with downstream DNA (Rpo5 and 13), the RNA transcript as well as the initiation factor TFE (Rpo4 and 7). RNAP subunits and their functions are summarized in Table 1 and Figure 1. The catalytic centre enabling phosphodiester bond formation and cleavage in all multisubunit RNAPs is formed between two structural motifs called double-psi β-barrels (DPBBs) residing in the large subunits (Rpo1 and Rpo2) [7,8]. In many archaea, the genes encoding the large RNAP subunits Rpo1 and Rpo2 are split into two open reading frames [8,9]. The two DPBBs acquired different functions crucial for the activity of extant RNAPs: One DPBB provides three carboxylate residues (aspartic acid) for the active site that chelate one of the two catalytic magnesium ions (Mg-A), while the second DPBB contributes two universally conserved lysine residues that facilitate interactions with nucleic acid and NTP substrates [10,11]. The overall RNAP core resembles a crab claw with a DNA-binding channel (aka main channel) between its pincers that leads the DNA template strand towards the active site (Figure 1). The NTP entry channel (aka secondary channel) connects the external milieu with the RNAP active centre allowing NTP substrates to enter the RNAP active site [12] and the RNA 3′-terminus to be extruded through it in backtracked transcription elongation complexes (ECs) (see below).

Like all molecular machines, the RNAP comprises a combination of rigid and flexible parts; the most prominent conformationally flexible domain of RNAP is the clamp. Movements of the clamp are conserved in all
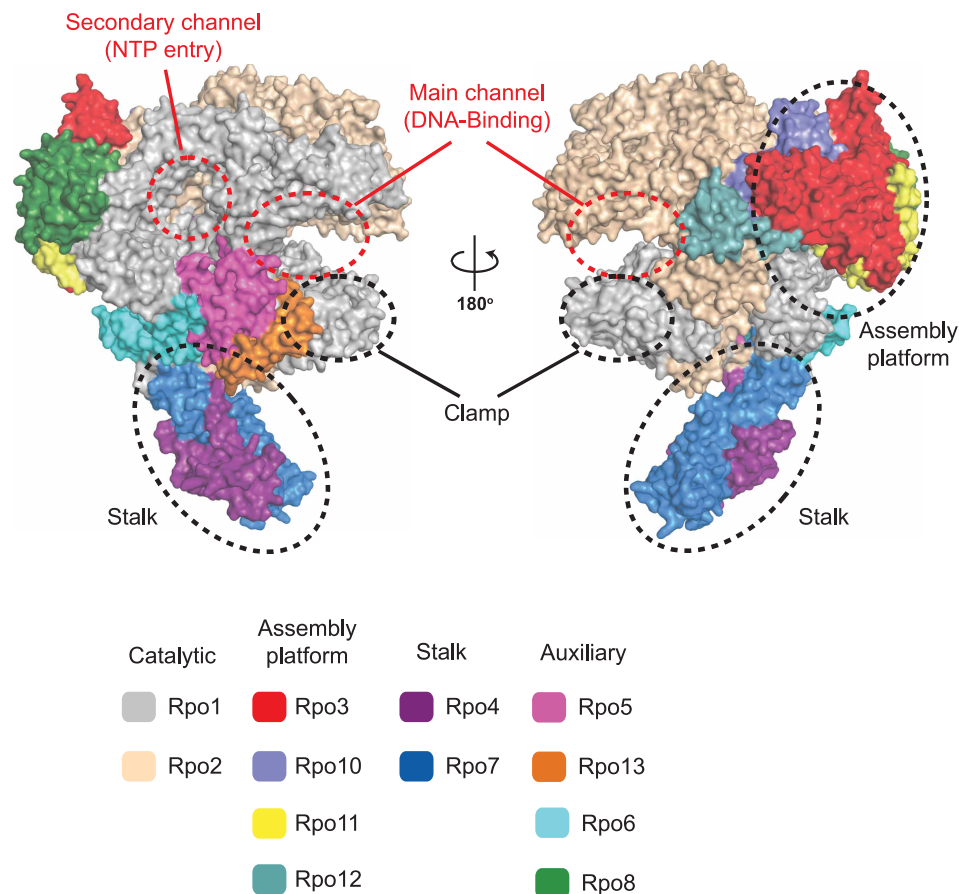
Figure 1. Structure of the archaeal RNAP.

Overall architecture of the archaeal RNAP (subunits are colour-coded according to the key). The DPBB-1 and -2 comprising the catalytic centre reside in the two largest subunits Rpo1 and -2, respectively. Important structural features and motifs, including the RNAP assembly platform, stalk, clamp and the (main) DNA-binding channel and (secondary) NTP entry channel, are highlighted with dashed circles.

DPBB RNAPs and not only alter the width of the DNA-binding channel but also translate into the microenvironment of the active centre. During transcription, RNAP progresses through three distinct phases of the transcription cycle starting with initiation of transcription, elongation and termination with concomitant release of the transcript (Figure 2A). Interactions with the DNA template and general initiation- and elongation factors modulate the position of the clamp, resulting in distinct clamp closure states that reflect functional states of RNAP at different phases of the transcription cycle. In brief, FRET measurements on *Methanocaldococcus jannaschii* RNAP revealed that clamp opening is important (i) during transcription initiation for DNA melting and template strand loading into the active site enhanced by the initiation factor TFE, (ii) keeping the clamp closed in conjunction with the factor Spt4/5 during elongation enables high processivity [13,14]. In addition, an opening of the clamp accompanies pausing and is a prerequisite for bacterial transcription termination [15].

Rpo13 is the only archaea-specific RNAP subunit, and it is only conserved in species belonging to the Sulfolobales family of Crenarchaeota. Rpo13 is located at the downstream end of the DNA-binding channel and has been speculated to interact with the DNA template. The most prominent difference between bacterial and archaeo-eukaryotic RNAPs is the stalk domain comprising Rpo4 and Rpo7. The stalk interacts with the initiation factor TFE during initiation, and with the nascent RNA transcript during elongation via an oligonucleotide/oligosaccharide binding (OB) S1 domain residing in Rpo7. The interactions between the RNA and Rpo7 have been reported *in vitro*, they increase the processivity during elongation, and enable efficient termination at weak intrinsic terminator signals [16]. Rpo8, like Rpo7, contains an OB-fold [17]. The functional
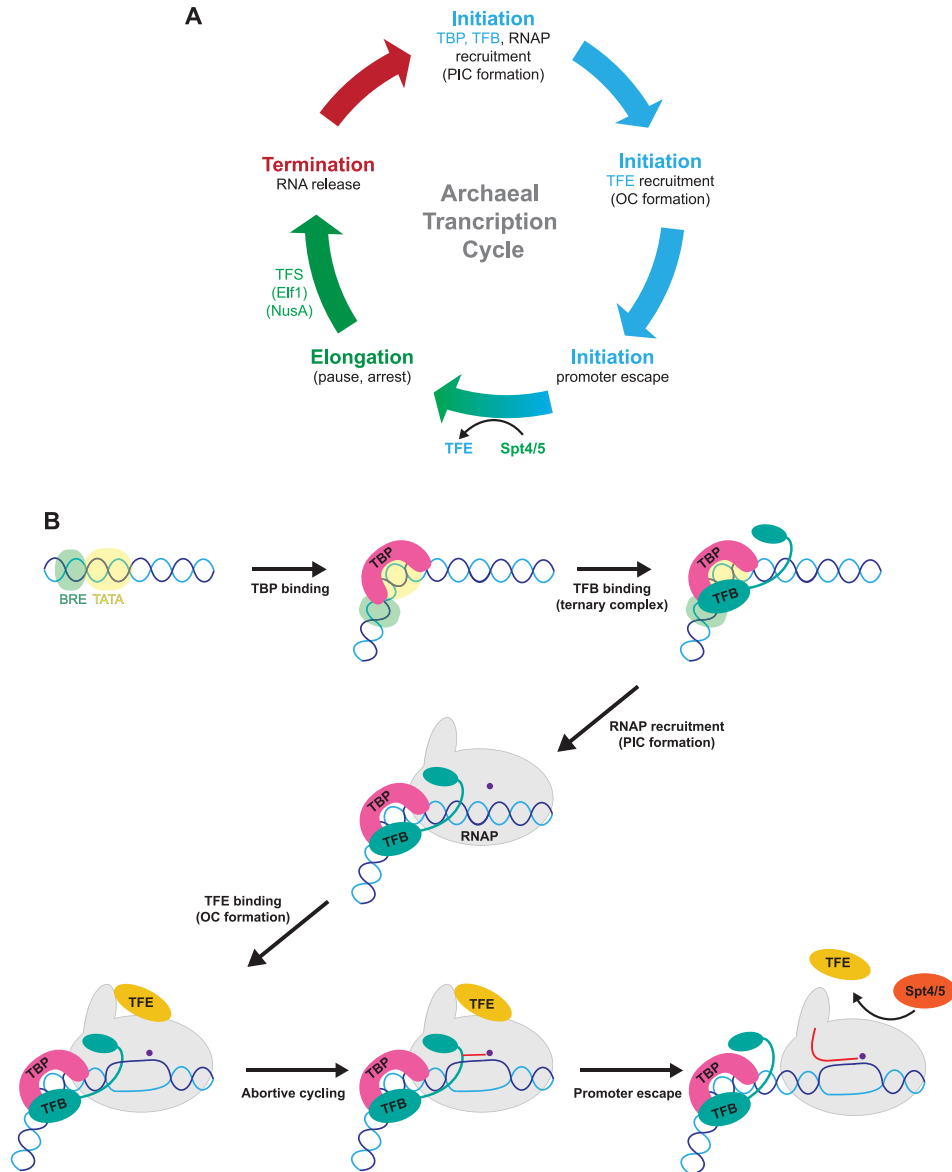
Figure 2. The archaeal transcription cycle.

(**A**) The archaeal transcription cycle consists of initiation, elongation and termination phases during which RNAP is assisted by general transcription factors. (**B**) Transcription cascade. TBP (pink) and TFB (green) bind to the TATA-box and BRE promoter elements, respectively, forming a ternary complex. RNAP (grey) is subsequently recruited to form the minimal PIC. TFE (yellow) is recruited to the PIC, and enhances the transition between the CC and OC which occur concomitantly with DNA strand separation and formation of the transcription bubble. In the presence of NTP substrates, RNAP undergoes abortive initiation that produces 3–9 nt RNA species — also called abortive transcripts or nano-RNAs. The elongation factor Spt4/5 (orange) displaces TFE in a process coined factor swapping during promoter escape or early elongation. Template and non-template DNA strands are shown in dark blue and light blue, respectively. Catalytic Mg-A is shown in purple.

implications of this OB-fold are unclear; however, the location adjacent to the secondary channel suggests that it could interact with the 3′ segments of the RNA that are extruded through the pore in backtracked ECs [11,18]. Rpo8 is the only archeo-eukaryotic acquisition in the RNAP subunit repertoire that is not conserved in all archaea. While Rpo8 is present in species belonging to the TACK (Cren- and Korarchaeota) and ASGARD (Odin- and Heimdallarchaeota) superphyla, it is not conserved in euryarchaeal and DPANN species (Table 1) [19,20]. The eukaryotic homologue of the archaeal transcript cleavage factor TFS corresponds to the RNAPII

subunit RPB9 as well as to the transcription factor TFIIS, a special case that is discussed in depth in section 'RNAP backtracking, arrest and reactivation'. Overall, archaeal RNAPs and in particular the RNAPs of the TACK and ASGARD superphyla are closely related to eukaryotic RNAPII in terms of subunit composition.

Factors and mechanisms enabling transcription initiation in archaea

While bacterial RNAPs require sigma factors to initiate transcription, archaeal RNAP utilizes the three general transcription initiation factors TBP (TATA-binding protein), TFB (Transcription Factor B) and TFE (Transcription Factor E) that are homologous to eukaryotic TBP, TFIIB and TFIIE [21–25], respectively. On the nucleic acid level, three consensus promoter elements direct the assembly of transcription initiation complexes on the promoter: the TATA-box (7–8 bp in length), B-recognition- (BRE, 3–5 bp in length) and the initiator (Inr) elements. The consensus DNA sequences of these elements is differentially conserved in the archaeal lineages; the TATA element is highly conserved, the BRE consensus is juxtaposed to TATA elements and enriched in A-residues, and the Inr (T(A/G)) is conserved in some archaea (e.g. *M. jannaschii*, *Methanolobus psychrophilus*, *Sulfolobus solfataricus* and *Haloferax volcanii*) but not in others (e.g. *Methanosarcina mazei*, *Thermococcus kodakarensis*) [26]. TATA and BRE serve to recruit transcription initiation factors. The sequence and role of the Inr is difficult to decipher because this motif overlaps extensively with start codon positions in many archaeal species such as *S. solfataricus* where most mRNAs are leaderless. A sequence preference for purines at the TSS preceded by a pyridine is a universal feature not only of DPBB RNAPs, but also of other RNAPs as it helps positioning the initiating NTP substrate [27]. A genome-wide comparison of transcript 5′-ends and the Inr motifs of the corresponding promoters revealed that the Inr is important for the exact positioning of the transcription start site TSS in *M. jannaschii* [28].

Archaeal TBP corresponds to the eukaryotic TBP core domain that binds to and distorts the TATA-containing promoter DNA by ∼90° (Figure 2B) [29,30]. TBP has an internal symmetry consisting of two repeats that are derived from an ancestral DNA-binding domain present in RNaseHIII [31,32]. The kinetics and stability of the TBP–DNA interaction differs significantly between archaeal species suggesting lineage-specific adaptation. For many archaea, the formation of a stable TBP:DNA complex requires TFB recruitment concomitant with TBP binding [29,33]. TFB, like its eukaryotic counterparts, consists of an N-terminal ZR (Zn-ribbon) domain connected by a flexible linker region to two cyclin fold domains at the C-terminus. The linker region itself comprises the B-reader and the B-linker motifs [34]. The orientation of the ternary TBP–TFB–DNA complex determines the directionality of transcription and relies on interactions between the second cyclin fold of TFB and the BRE upstream of the TATA-box [35]. The ZR domain of TFB interacts with the RNAP dock domain and recruits RNAP to the promoter forming a minimal DNA–TBP–TFB–RNAP pre-initiation complex (PIC). The B-linker penetrates deep into the RNAP and stabilizes the template DNA strand (TS) in the active site [34,36,37]. Many archaea, most prominently haloarchaea, utilize combinations of multiple TBP and TFB homologues, allowing different combinations of TBP–TFB which enable a certain degree of promoter specificity [38–40]. Additional TFB paralogues do not necessarily function the same way as canonical TFBs. TFB3, a TFB paralogue in *Sulfolobus* that is induced by UV-radiation and DNA damage, cannot replace the canonical TFB homologue, but rather appears to activate transcription in conjunction with canonical TFB in trans via a mechanism that is still poorly understood [41]. Recent insights into the genes under direct control of TFB3 provide now a basis for functional studies into the molecular mechanism of transcription activation by TFB3 [42,43].

To load the DNA TS into the RNAP active centre, the DNA strands are locally melted in a region 12 bp upstream of the TSS. This process is accompanied by large-scale conformational changes of the PIC that are referred as closed (CC) to open complex (OC) transition [21,44–46]. The initially melted region (IMR) shows an increased AT-content that might aid DNA melting in some, but not all archaea. DNA melting and OC formation are facilitated by the third archaeal initiation factor termed TFE. Canonical archaeal TFE and its eukaryotic counterpart TFIIE are composed of two subunits (α and β) [21]. TFEα and TFIIEα share the bipartite WH (winged helix) and ZR domain organization that interact with RNAP in a bidentate fashion: the TFEα ZR domain interacts with the RNAP clamp and stalk, whereas the TFEα WH domain interacts with the RNAP clamp coiled-coil (clamp-CC) domain [47]. The interactions of TFEα with both the stalk and clamp domains of RNAP together with interactions of TFE with the non-template strand (NTS) of the promoter DNA retain the clamp in the open conformation, and stabilize the transcription bubble, respectively, facilitating OC formation [45,47,48] (Figure 2B). In the presence of NTP substrates, the RNAP enters into abortive cycles of

synthesis which release short RNA transcripts (2–15 nt) prior to the full extension of the RNA–DNA hybrid and escape from the promoter [46,49,50].

Is TFEβ a global regulator of transcription?

The TFEβ subunit has a patchy phylogenetic distribution and is present in most group I euryarchaeota (with the exception of Thermoplasmata that lack both α- and β-subunits), TACK (missing in Korarchaeota) and ASGARD superphyla, but is absent from species from group II Euryarchaeota and DPANN [26,51]. TFEβ consists of an N-terminal WH- and a C-terminal cubane [4Fe–4S] cluster domain. The [4Fe–4S] cluster easily undergoes oxidative damage rendering TFE sensitive to oxidative stress. TFEβ expression levels vary dramatically with growth conditions and environmental stresses in *S. solfataricus*, unlike the remaining general transcription factors. Since TFE is a general factor and its activation of transcription varies considerably between promoters, the depletion of TFEβ has the potential to alter the RNA synthesis globally in *S. solfataricus*. In essence, modulation of OC formation provides an opportunity for the regulation of transcription, a mechanism which has previously been shown to operate in bacterial and eukaryotic transcription systems [26]. Interestingly, *H. volcanii* TFEβ and indeed all haloarchaeal homologues lack the [4Fe–4S] cluster that is essential for *S. solfataricus* TFEβ function. Nevertheless, in line with TFEβ being a bona fide general transcription factor, the deletion of the *H. volcanii* TFEβ results in the misregulation of approximately one-third of all transcription units [52]. The group II Euryarchaeota lacks TFEβ altogether and monomeric TFEα can fully support OC formation [23,36,48]. This broad, though patchy phylogenetic, distribution suggests that both TFEα and β subunits were present in the last archaea common ancestor (LACA) [26,52].

Promoter escape: early transcription elongation

All DPBB RNAPs face similar mechanical engineering problems when entering the early elongation phase of the transcription cycle. A network of high affinity interactions between DNA-bound initiation factors (TBP, TFB and TFE) and RNAP are important to enable efficient recruitment to the promoter. However, these interactions need to be disrupted for RNAP to escape the promoter and enter processive transcription elongation. Structures of the initially transcribing complex of yeast RNAPII as well as recent cross-linking studies in *Pyrococcus* have shown that once the nascent RNA exceeds 5 nt in length, it collides with the TFB B-reader and B-linker domains, disrupting the interaction with and displacing TFB from the active site of RNAP [34,53,54]. Promoter escape of archaeal RNAP has not been well studied thus far and probably differs from its eukaryotic counterpart RNAPII with its drastically increased repertoire of initiation factors. Exonuclease and permanganate foot-printing studies revealed that promoter escape is initiated once the nascent RNA reaches 10 nt in length [46]. Once the elongating RNAP has reached register +15, the interactions between TFB and the DNA downstream of the TATA-box are disrupted [54]. An additional feature of the promoter escape is the swapping of initiation (TFE) and elongation factors (Spt4/5), both of which bind to overlapping binding sites on the RNAP clamp-CC motif in a mutually exclusive manner. This mechanism was initially discovered using biochemical and biophysical interaction analysis and transcription assays *in vitro* [47] and it is supported with the early recruitment of Spt4/5 to the vast majority of transcription units *in vivo* determined using chromatin immunoprecipitation (ChIP-seq) [28]. The association of Spt4/5 possibly induces allosteric changes in RNAP from an initiation- to elongation competent conformation. In line with this idea, single molecule FRET experiments showed that TFE and Spt4/5 exert opposing effects on the position of the RNAP clamp [14]. The global occupancy analysis revealed that a subset of non-coding RNA transcription units, including the ribosomal RNA operons and CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) loci, displayed a delayed Spt4/5 recruitment to the promoter, suggestive of an alternative promoter escape mechanism possibly reliant on additional uncharacterized transcription factors [28].

Factors and mechanisms that enable efficient transcription processivity

A subset of evolutionarily conserved regulatory factors assist RNAP during transcription elongation by modulating the elongation rate and/or by improving the processivity (defined as polymerized nucleotides per initiation). Elongation factors belonging to the Spt4/5 family (the bacterial homologue of Spt5 is called NusG) stimulate transcription by binding to the RNAP clamp-CC on one side of the DNA-binding channel and to the RNAP gate loop on the other [13,55–58]. This locks the clamp into the closed state and seals the

DNA-binding channel of RNAP, which counteracts the dissociation of the EC [13,59,60]. In addition, the interaction between the Spt5/NusG NGN (NusG N-terminal) domain and the template DNA enhances the annealing of the TS and NTS at the upstream edge of the transcription bubble and thereby suppress backtracking and pausing, which overall improves processivity and increases the elongation rate [61]. Bacterial NusG is monomeric, while in archaea and eukaryotes, NusG homologue Spt5 forms a heterodimer with a small ZR-containing protein Spt4. Spt4 not only exerts a stabilizing effect on the Spt5 NGN domain but also may functionally interact with the upstream DNA of the EC [13,59]. In addition to the NGN domain, Spt5 contains one (in archaea and bacteria) or several (in eukaryotes) KOW (Kyrpides–Ouzounis–Woese) domains. In bacteria and probably in archaea, the KOW domain interacts with the ribosomal protein S10 thereby physically coupling RNAP and the first co-translating ribosome, coordinating transcription and translation [55,62].

Structural insight into complete yeast transcription ECs encompassing RNAP, DNA, TFIIS, Spt4/5 and Elf1 reveals a striking reoccurring theme, by which the latter two elongation factors form entry and exit tunnels for the DNA and RNA strands [58,59]. Elf1 (Elongation factor 1) is a transcription elongation factor conserved in eukaryotes and several archaeal species [63]. Homologues of Elf1 have been identified in of the TACK (Cren-, Kor-, Aig- and Bathyarchaeota) as well as in the ASGARD superphylum (Table 1) [3,63,64]. Elf1 comprises a positively charged N-terminal α-helical tail, a structurally discrete ZR domain and a negatively charged unstructured C-terminal tail [59,65]. ChIP-Seq analyses in yeast demonstrate that Elf1 accompanies elongating RNAPII in similar manner to Spt4/5 [66], and *in vitro* transcription assays showed that Elf1 inhibits transcription elongation, possibly by interacting with downstream DNA via its N-terminal tail [59]. Elf1 is likely also part of the archaeal EC (Figure 3); however, the mechanism and function of Elf1 during transcription elongation in archaea remains enigmatic.

A likeness between the Rpo4/7 RNAP stalk and NusA

The origin of the RNAP stalk — a hallmark feature of archaeal and eukaryotic RNAPs — is opaque, but a combination of recent structural and functional studies has revealed a striking resemblance to a bacterial elongation factor. The bacterial NusA (N-utilization substance A) interacts with the RNAP via the NusA N-terminal
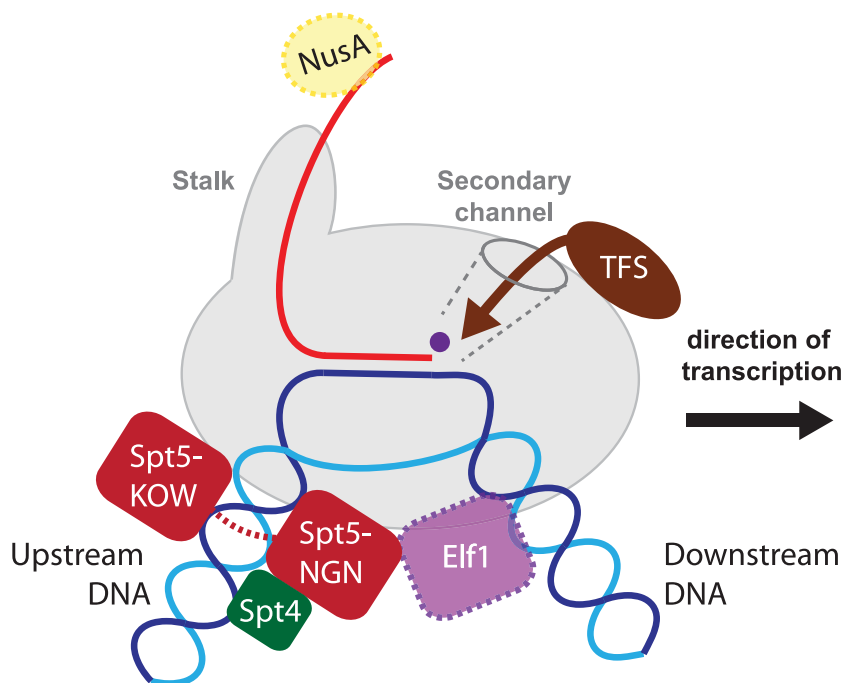

Figure 3. The archaeal transcription elongation complex.
Schematic representation of the complete archaeal transcription elongation complex encompassing RNAP-DNA/RNA, TFS, Spt4/5 and possibly Elf1. The function of the cleavage and processivity factors are discussed in detail in the main text. The RNA-bound NusA is indicated beyond the tip of the RNAP stalk. Factors with unknown function is archaea are highlighted in dashed lines.
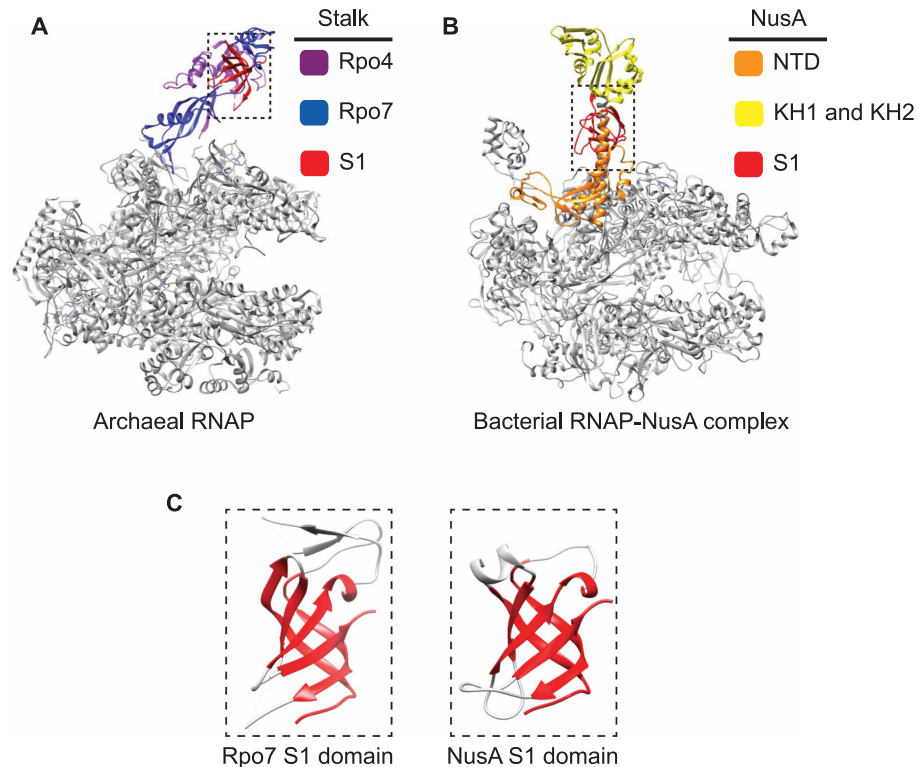
Figure 4. Structure comparisons of archaeal RNAP and the bacterial RNAP–NusA complex.
(**A**) The archaeo-eukaryotic Rpo4/7 subunits form a stalk-like protrusion highly reminiscent of (**B**) the RNAP-bound bacterial elongation and antitermination factor NusA. The S1 domains of Rpo7 and NusA are highlighted in dashed lines. The insertion domains SI1, SI2 and SI3 of *E. coli* RNAP and regulatory C-terminal domain of NusA were omitted for clarity. (**C**) RNA-binding S1 domains of archaeal Rpo7 (PDB: 1GO3) and bacterial NusA (PDB: 1K0R).

domain (NTD), and with RNA in at least two distinct ways that have different effects on transcription elongation. Interactions of the NusA S1 domain with an RNA hairpin enhances transcription pausing. In contrast, interactions between the NusA KH1 and -2 domains and the RNA nut (N-utilization target) sequence promote antitermination on ribosomal RNA operons [67]. This increases the elongation rate and renders the RNAP inert to the action of the termination factor rho [68]. All archaea encode one or several genes homologous to NusA, but archaeal NusA variants only encompass the two RNA-binding KH domains and not the N-terminal RNAP-interaction- and S1 domains altogether [69,70]. The archaeal RNAP stalk subunit Rpo7 includes an RNAP interaction domain, as well as an S1 domain that interacts with the nascent RNA transcript, which in turn modulates both transcription elongation and termination properties of the elongation complex [16] (Figure 4A). In combination, the RNAP interaction- and S1 domains of Rpo7 in conjunction with the two KH domains of archaeal NusA provide the complete domain complement of bacterial NusA [71]. Moreover, a recent structure of the bacterial RNAP–NusA complex shows that NusA forms an elongated stalk protruding from the RNAP proximal to the RNA exit channel, somewhat reminiscent of the archaeal and eukaryotic RNAP structures (Figure 4B) [72]. The possibility of a relationship between Rpo7 and NusA is enticing, and the S1 domains of archaeal Rpo7 and eukaryotic RPB7 and bacterial NusA are homologous [71] (Figure 4C). If indeed Rpo7 is homologous to NusA, an important question remains how the division of one polypeptide (NusA) in bacteria into two distinct (Rpo7 and NusA) polypeptides in archaea-altered NusA function.

Transcription bubble maintenance by flexible RNAP motifs

During transcription elongation, RNAP translocates along the template DNA via a thermal Brownian ratchet mechanism [73–75]. The active centre of RNAP contains several polypeptide loops that were shown to be critical for the proper arrangement of the RNA–DNA hybrid and its stability during RNAP elongation. The
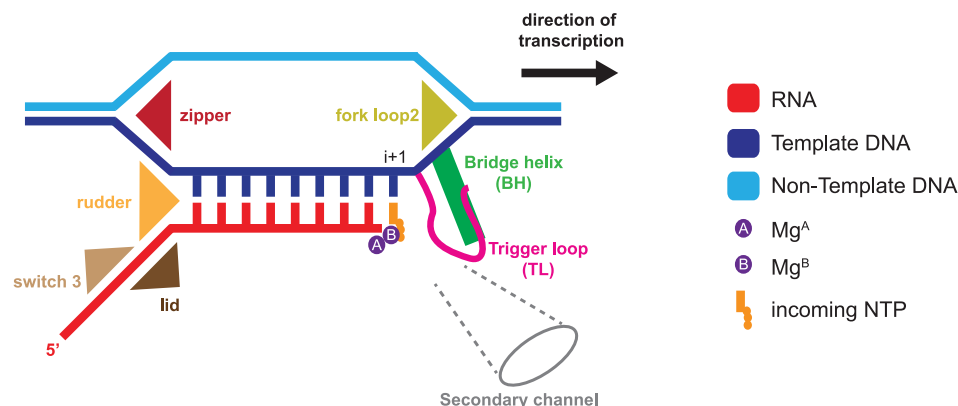
Figure 5. Flexible motifs enable the nucleotide translocation cycle of DPBB–RNAP.
Schematic representation of RNAP active centre in the transcribing RNAP elongation complex (EC), the NTP insertion site corresponds to register i + 1. The motifs critical for RNA–DNA hybrid maintenance are shown as coloured triangles, and the trigger loop (TL) and bridge helix (BH) are shown in magenta and green, respectively, RNA, DNA strands, catalytic Magnesium ions and NTP substrate are coded according to the key.

downstream DNA stacks on the fork loop 2, which plays a critical role in double-strand DNA separation [76] (Figure 5). The lid serves as a wedge to facilitate RNA displacement by sterically blocking the formation of an overextended hybrid [76–78], while the rudder interacts with the RNA and overall stabilizes the EC [79]. Switch 3 binds to each RNA base in a nascent transcript as it dissociates from the RNA–DNA hybrid, stabilizing the EC [80,81]. Finally, the double-stranded DNA is reformed at the upstream edge of the transcript bubble by the zipper motif [11,77].

RNAP backtracking, arrest and reactivation

Transcription elongation is a discontinuous process during which the EC pauses relatively frequently. This pausing can be induced in DNA sequence-dependent fashion (e.g. poly-A stretches in the TS) or by roadblocks such as DNA-bound proteins (e.g. chromatin proteins) or DNA lesions [82–85]. Upon pausing, RNAP can move in a retrograde direction along the DNA, which is referred to as backtracking. During this process, the RNA–DNA hybrid maintains its 8–9 bp length, while one or more nucleotides of the RNA 3′-end are displaced from the downstream edge of the RNA–DNA hybrid out of the active site rendering the backtracked EC catalytically inert. If backtracking proceeds further, longer RNA 3′ segments are extruded from the RNAP through the secondary channel. Backtracked ECs pose a severe problem for the cell since they act as roadblocks for upstream transcription ECs and replication forks, which can lead to double-stranded DNA breaks compromising genome integrity [86]. Transcript cleavage factors resolve this conflict by inducing an endonucleolytic cleavage activity inherent in DPBB RNAPs. This generates a new RNA 3′-end conducive to RNA polymerization and transcription elongation can commence.

While archaea and eukaryotes utilize evolutionary related factors, TFS and TFIIS, respectively, the non-homologous bacterial Gre factors carry out the same function while providing a stunning case of convergent evolution [87] (Table 1). With the exception of the euryarchaeon *Methanopyrus kandleri*, TFS is conserved in all archaeal species [88]. Both TFS and TFIIS are evolutionarily related to RPB9-like subunits of eukaryotic RNAPs, but while RPB9 subunits are stably incorporated into RNAP, TFS/TFIIS associate with their cognate RNAP in a reversible fashion (Table 1) [89]. All transcript cleavage factors position two acidic residues at the tip of an elongated insertion domain through the secondary into the RNAP active site. The carboxylate moieties stabilize a magnesium ion required for the stable coordination of a water molecule that carries out a nucleophilic attack on the RNA phosphodiester bond triggering RNA cleavage [89–91].

Functional diversification of archaeal transcript cleavage factors

Several archaeal species encode more than one TFS paralogue, e.g. the genome of the crenarchaeon *S. solfataricus* includes four apparent TFS paralogues (TFS1 to 4). While TFS1 carries out the canonical transcript

cleavage function, TFS4 has evolved into a highly potent RNAP inhibitor [91]. TFS4 shares a high degree of sequence similarity with TFS1 but lacks the catalytic acidic (DE) residues required for transcript cleavage activity. Rather, three lysine residues replace the acidic residues at the tip of the insertion domain. TFS4 binding to RNAP destabilizes transcription initiation and ECs which suggests that it exerts an allosteric effect that compromises the interactions between RNAP and the nucleic acid scaffold. These conformational changes are likely characteristics for all DPBB RNAP and related to the mechanism by which the bacterial regulator Gfh1 inhibits RNAP activity [92,93]. In addition to the allosteric mechanism, TFS4 acts as a competitive inhibitor for NTP binding to RNAP, possibly by sterical blockage as suggested by its binding site within the secondary channel. Expression of the *tfs4* gene is not detectable under standard growth conditions. However, infection with STIV (*Sulfolobus* turreted icosahedral virus) leads both to a dramatic increase in TFS4 expression, and induces a dormant state in the infected cell [94] (Figure 6). TFS4 is likely to play a key role in this process, since the ectopic overexpression of a TFS4 variant is sufficient to induce a severe growth retardation in line with its potent inhibitory effect on global transcription. Our understanding of the biological function of TFS4 during infection leaves much room for improvement, but it seems likely that the inhibition of transcription is an antiviral response that enables host survival by persistence. This is a survival strategy employed by bacteria in response to bacteriophage infection; the infected cells enter a hiatus to inhibit virus proliferation often in conjunction with additional, more active, defence mechanisms [95].

Global inhibition of transcription in the host–virus arms race

TFS4 is a host encoded archaeal transcription factor that inhibits transcription on a global level in response to viral infection [91]. Surprisingly, archaeal viruses themselves use a very similar strategy to their own advantage. Archaeal cells are exposed to a plethora of viruses in their natural environment, and an ongoing arms race between the two has shaped the relationship between them [96]. One of the primary antiviral defence
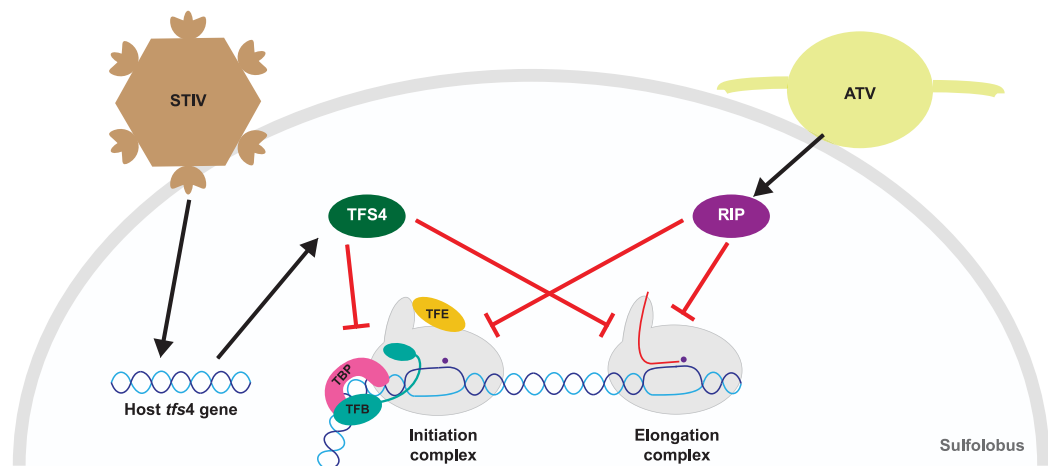

Figure 6. Global transcription inhibition in the virus–host relationship.
Both host- (TFS4) and virus-encoded factors (RIP) can directly associate with the archaeal RNAP and efficiently shut down transcription on a genome-wide scale. The *S. solfataricus* transcript cleavage factor homologue TFS4 interacts with RNAP like other cleavage factors such as TFS1 through the NTP entry channel. Rather than promoting transcription elongation, TFS4 dramatically lowers the affinity of RNAP for NTP substrates thereby inhibiting catalysis, and induces allosteric changes that destabilize RNAP-nucleic acid interactions. TFS4 expression is repressed during normal cell growth but highly induced by infection with the STIV. In comparison, the ATV that infects Sulfolobales encodes the small protein RIP, which is derived from a viroid coat protein but has evolved into a potent inhibitor of the archaeal RNAP. RIP binds to the RNAP clamp in the DNA-binding channel, locks the clamp into a fixed position and inhibits RNAP activity in a global fashion. Both TFS4 and RIP inhibit transcription in a DNA sequence-independent fashion, i.e. they repress host as well as virus promoters. While the former has been speculated to provide a survival mechanism for the infected host akin to persistence, the latter probably serves the virus by preventing or attenuating the activation of cellular antiviral type III-B CRISPR–Cas system.

mechanisms of the archaeal hosts is provided by CRISPR–Cas systems [97]. Sophisticated viral counter measures involve the subjugation of the host transcription machineries, as well as strategies to stay 'under the radar' of surveillance mechanisms including the type IIIb CRISPR–Cas system that is triggered by active transcription [98–100]. Encounters between the *Acidianus* two-tailed virus (ATV) and *S. solfataricus* are witnessed by the presence of several ATV genome-derived CRISPR spacers in the hosts' CRISPR arrays, including sequences mapping to a small gene called ORF145 [101]. ORF145, also called RIP (RNAP inhibitory protein), binds tightly to the inside of the DNA-binding channel of the host RNAP, thereby locking the RNAP clamp into a fixed position [102]. This counteracts the formation of transcription initiation complexes and inhibits abortive and productive transcription (Figure 6).

The interaction of RIP with RNAP differs in a fundamental way from TFS4, but the outcome is surprisingly similar. RNAP-nucleic acid complexes are destabilized and transcription initiation and elongation are inhibited. Because RIP, like TFS4, binds directly to RNAP, both host and virus promoters are inhibited in a global fashion. While the regulatory rationale behind this mechanism is still unclear, it is likely that the inhibition of transcription attenuates or even prevents the activation of the type IIIb CRISPR–Cas system and expression of anti-ATV CRISPR RNAs, while still enabling transcription on viral genes required for virus proliferation [102].

Mechanisms and factors that enable transcription termination

Transcription termination not only defines the nascent 3′ terminus of the RNA transcript but is important to prevent transcription read through of RNAP from upstream into the adjacent transcription units in the densely crowded environment of small archaeal genomes. Despite its biological significance, transcription termination remains one of the least understood processes of gene expression in archaea. *In vitro* and *in vivo* studies have shown that euryarchaeal RNAPs are capable of terminating transcription directed by short poly-U stretches and unaided by exogenous factors, a property reminiscent of the eukaryotic RNAPIII system (Figure 7) [103–109].

An unbiased mapping of RNA 3′-ends in a euryarchaeal- (*M. mazei*) and crenarchaeal (*S. acidocaldarius*) species using Term-seq has provided an overview of RNA 3′-ends *in vivo* on a genome-wide scale [110]. However, Term-seq alone cannot discriminate between genuine transcription termination sites and RNA 3′ ends resulting from RNA processing; therefore, additional prior information, such as high-resolution RNAP occupancy profiling (ChIP-seq) and -transcriptome mapping (RNA-seq) and the position of the RNA 3′ end relative to operon structures and stop codons (at the end of ORFs), needs to be taken into account for a rigorous and meaningful analysis. Overall, the archaeal Term-seq study supports the notion that transcription terminates immediately downstream of uridine-rich sequences but also highlights additional, lineage-specific sequence features in *M. mazei* and *S. acidocaldarius* that are not accounted for in current models of transcription termination [110]. Termination in about one-third of genes in either system is enabled by multiple
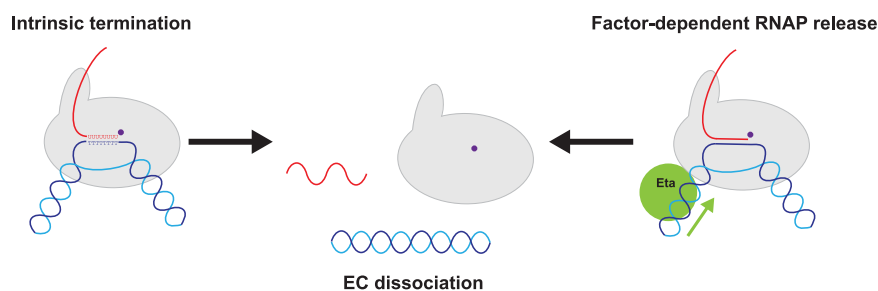

Intrinsic termination   **Factor-dependent RNAP release**

EC dissociation

Figure 7. Transcription termination and RNA release.
Schematic representations of transcription termination in archaea. Termination events that do not rely on exogenous factors are known as intrinsic termination. In archaea, intrinsic termination does not rely on secondary structures in the transcript such as RNA hairpins. Rather, a poly-U stretch seems sufficient to enable termination *in vitro* and *in vivo*. Recently, the first archaeal termination factor, Eta, has been shown to enhance RNA transcript release from stalled ECs. Eta is a Ski2-like DEAD box helicase that in an ATP-hydrolysis-dependent fashion translocates along the DNA; upon impact with the RNAP from the upstream direction the transcript is released, the TEC dissociates, and transcription is terminated.

terminator elements, resulting in variations of non-coding 3′-untranslated regions (3′-UTRs) with differing lengths that could be involved in the regulation of gene expression by small non-coding RNAs. Alternatively, or in addition, transcription termination in archaea could be less precise compared with prototypical bacterial intrinsic terminators, possibly due to the lack of any RNA hairpin secondary structures in archaeal terminators.

Little is known about factor-dependent transcription termination mechanisms in archaea. In *T. kodakarensis*, the Ski2-like RNA helicase Eta (Euryarchaeal Termination Activity) is a DEAD box helicase that is recruited to stalled transcription complexes via interactions with the DNA immediately upstream of the arrested RNAP [111,112]. Eta translocates along the DNA in an ATP-dependent fashion, pushes the EC forward and thereby releases the nascent RNA (Figure 7). It is important to point out that Eta, unlike other transcription termination factors in bacteria and eukaryotes, is not essential for cell viability and does not trigger transcription termination of actively elongating RNAPs [112]. Eta's properties suggest that it is not a general transcription termination factor but rather likely to be a component of the DNA damage response akin to the Mfd factor in bacteria [113,114].

Future perspectives

RNAPs are among the most well-studied molecular machines of life. The initiation phase of transcription has been characterized over the last two decades. These studies have elucidated the structure, function and detailed mechanisms that govern the archaeal PIC. Many studies have identified positive and negative transcription factors that enhance or prevent its recruitment of the PIC. While the structure and mechanisms of elongation factors like TFS and Spt4/5 are reasonably well understood *in vitro*, a thorough understanding of how these factors influence transcription *in vivo* just starts to emerge. An integrated, genome-wide view of transcription in archaea shows promise to bring to light more sophisticated mechanisms of transcription regulation beyond the initial recruitment, probably involving promoter escape and transcription processivity during the elongation phase of the transcription cycle.

Recent discoveries of virus and host encoded global inhibitors of RNAP transcription have shed light on novel molecular mechanisms and regulatory strategies that seemingly play a key role in the host–virus arms race. Finally, our field is coming to terms with the fact that the chromatin structure, histone-based or otherwise, plays an important role in gene regulation in archaea. Novel approaches, including high-throughput sequencing techniques, live cell imaging, as well atomic-resolution cryo-electron microscopy, will lead to key discoveries and a new dawn of archaeal gene expression, with an ever more detailed understanding of transcription from the molecular to the systems level.

Summary

- The catalytic centre or the archaeal RNAP is formed between two DPBB.

- Combined ChIP-seq and RNA-seq analyses reveal the genome-wide organization of transcription and generate new mechanistic hypotheses that can be tested *in vitro*.

- The general transcription initiation factor TFEβ has the potential to regulate transcription globally in response to environmental stresses.

- Transcription elongation is modulated by RNAP subunits (Rpo4/7) and transcription factors (Spt4/5, TFS1 and likely Elf1).

- The RNAP stalk subunit Rpo7 shows an intriguing structural and functional similarity to the bacterial pausing/antitermination factor NusA.

- The expression of the transcript cleavage factor paralogue TFS4 is induced by STIV virus infection and acts as a powerful global inhibitor of RNAP in *S. solfataricus*.

PORTLAND
PRESS

- The ATV virus-encoded regulator RIP binds directly to RNAP and results in the global inhibition of transcription; thus both host- and virus-encoded RNAP-binding transcription factors globally inhibit or attenuate total RNA synthesis.

- Genome-wide mapping of transcript 3′-ends changes our view on the sequence context of archaeal transcription terminators.

- RNAPs stalled by DNA-damage can be efficiently removed by the termination-like factor Eta.

Abbreviations

ATV, *Acidianus* two-tailed virus; CC, closed; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; DPBB, double-psi β-barrel; ECs, elongation complexes; Elf1, Elongation factor 1; Eta, Euryarchaeal Termination Activity; IMR, initially melted region; Inr, initiator; KOW, Kyrpides–Ouzounis–Woese; NTD, N-terminal domain; NTS, non-template strand; NusA, N-utilization substance A; OB, oligonucleotide/oligosaccharide binding; OC, open complex; PIC, pre-initiation complex; RIP, RNAP inhibitory protein; RNAP, RNA polymerase; STIV, *Sulfolobus* turreted icosahedral virus; TBP, TATA-binding protein; TFB, transcription factor B; TFE, transcription factor E; WH, winged helix.

Acknowledgements

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1   Guy, L. and Ettema, T.J. (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 https://doi.org/10.1016/j.tim.2011.09.002

2   Spang, A., Caceres, E.F. and Ettema, T.J.G. (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 https://doi.org/10.1126/science.aaf3883

3   Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Backstrom, D., Juzokaite, L., Vancaester, E. et al. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 https://doi.org/10.1038/nature21031

4   Adam, P.S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. (2017) The growing tree of *Archaea*: new perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407–2425 https://doi.org/10.1038/ismej.2017.122

5   Castelle, C.J. and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 https://doi.org/10.1016/j.cell.2018.02.016

6   Werner, F. and Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 https://doi.org/10.1038/nrmicro2507

7   Iyer, L.M., Koonin, E.V. and Aravind, L. (2003) Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **3**, 1 https://doi.org/10.1186/1472-6807-3-1

8   Fouqueau, T., Blombach, F. and Werner, F. (2017) Evolutionary origins of two-barrel RNA polymerases and site-specific transcription initiation. *Annu. Rev. Microbiol.* **71**, 331–348 https://doi.org/10.1146/annurev-micro-091014-104145

9   Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F. et al. (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl Acad. Sci. U.S.A.* **86**, 4569–4573 https://doi.org/10.1073/pnas.86.12.4569

10 Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 https://doi.org/10.1016/S0092-8674(00)81515-9

11 Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 https://doi.org/10.1126/science.1059493

12 Zhang, L., Silva, D.A., Pardo-Avila, F., Wang, D. and Huang, X. (2015) Structural model of RNA polymerase II elongation complex with complete transcription bubble reveals NTP entry routes. *PLoS Comput. Biol.* **11**, e1004354 https://doi.org/10.1371/journal.pcbi.1004354

13 Hirtreiter, A., Damsma, G.E., Cheung, A.C., Klose, D., Grohmann, D., Vojnic, E. et al. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 https://doi.org/10.1093/nar/gkq135

14   Schulz, S., Gietl, A., Smollett, K., Tinnefeld, P., Werner, F. and Grohmann, D. (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl Acad. Sci. U.S.A.* **113**, E1816–E1825 https://doi.org/10.1073/pnas.1515817113

15   Weixlbaumer, A., Leon, K., Landick, R. and Darst, S.A. (2013) Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 https://doi.org/10.1016/j.cell.2012.12.020

16   Hirtreiter, A., Grohmann, D. and Werner, F. (2010) Molecular mechanisms of RNA polymerase –– the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Res.* **38**, 585–596 https://doi.org/10.1093/nar/gkp928

17   Krapp, S., Kelly, G., Reischl, J., Weinzierl, R.O. and Matthews, S. (1998) Eukaryotic RNA polymerase subunit RPB8 is a new relative of the OB family. *Nat. Struct. Biol.* **5**, 110–114 https://doi.org/10.1038/nsb0298-110

18   Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B. et al. (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e1000102 https://doi.org/10.1371/journal.pbio.1000102

19   Koonin, E.V., Makarova, K.S. and Elkins, J.G. (2007) Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and 'Korarchaeota'. *Biol. Direct* **2**, 38 https://doi.org/10.1186/1745-6150-2-38

20   Jun, S.H., Hirata, A., Kanai, T., Santangelo, T.J., Imanaka, T. and Murakami, K.S. (2014) The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat. Commun.* **5**, 5132 https://doi.org/10.1038/ncomms6132

21   Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C. et al. (2015) Archaeal TFEalpha/beta is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife* **4**, e08378 https://doi.org/10.7554/eLife.08378

22   Qureshi, S.A., Bell, S.D. and Jackson, S.P. (1997) Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J.* **16**, 2927–2936 https://doi.org/10.1093/emboj/16.10.2927

23   Hanzelka, B.L., Darcy, T.J. and Reeve, J.N. (2001) TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEα. *J. Bacteriol.* **183**, 1813–1818 https://doi.org/10.1128/JB.183.5.1813-1818.2001

24   Hausner, W., Wettach, J., Hethke, C. and Thomm, M. (1996) Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J. Biol. Chem.* **271**, 30144–8 https://doi.org/10.1074/jbc.271.47.30144

25   Bell, S.D., Brinkman, A.B., van der Oost, J. and Jackson, S.P. (2001) The archaeal TFIIEα homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep.* **2**, 133–138 https://doi.org/10.1093/embo-reports/kve021

26   Blombach, F., Smollett, K.L., Grohmann, D. and Werner, F. (2016) Molecular mechanisms of transcription initiation-structure, function, and evolution of TFE/TFIIE-like factors and open complex formation. *J. Mol. Biol.* **428**, 2592–2606 https://doi.org/10.1016/j.jmb.2016.04.016

27   Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C. et al. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289**, 24549–24559 https://doi.org/10.1074/jbc.M114.584037

28   Smollett, K., Blombach, F., Reichelt, R., Thomm, M. and Werner, F. (2017) A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* **2**, 17021 https://doi.org/10.1038/nmicrobiol.2017.21

29   Gietl, A., Holzmeister, P., Blombach, F., Schulz, S., von Voithenberg, L.V., Lamb, D.C. et al. (2014) Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res.* **42**, 6219–6231 https://doi.org/10.1093/nar/gku273

30   Rowlands, T., Baumann, P. and Jackson, S.P. (1994) The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* **264**, 1326–1329 https://doi.org/10.1126/science.8191287

31   Kosa, P.F., Ghosh, G., DeDecker, B.S. and Sigler, P.B. (1997) The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl Acad. Sci. U.S.A.* **94**, 6042–6047 https://doi.org/10.1073/pnas.94.12.6042

32   Brindefalk, B., Dessailly, B.H., Yeats, C., Orengo, C., Werner, F. and Poole, A.M. (2013) Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res.* **41**, 2832–2845 https://doi.org/10.1093/nar/gkt045

33   Qureshi, S.A., Khoo, B., Baumann, P. and Jackson, S.P. (1995) Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc. Natl Acad. Sci. U.S.A.* **92**, 6077–6081 https://doi.org/10.1073/pnas.92.13.6077

34   Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M. et al. (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 https://doi.org/10.1038/nature08548

35   Bell, S.D., Kosa, P.L., Sigler, P.B. and Jackson, S.P. (1999) Orientation of the transcription preinitiation complex in archaea. *Proc. Natl Acad. Sci. U.S.A.* **96**, 13662–7 https://doi.org/10.1073/pnas.96.24.13662

36   Werner, F. and Weinzierl, R.O. (2005) Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell Biol.* **25**, 8344–8355 https://doi.org/10.1128/MCB.25.18.8344-8355.2005

37   Wiesler, S.C. and Weinzierl, R.O. (2011) The linker domain of basal transcription factor TFIIB controls distinct recruitment and transcription stimulation functions. *Nucleic Acids Res.* **39**, 464–474 https://doi.org/10.1093/nar/gkq809

38   Baliga, N.S., Goo, Y.A., Ng, W.V., Hood, L., Daniels, C.J. and DasSarma, S. (2000) Is gene expression in *Halobacterium* NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol. Microbiol.* **36**, 1184–1185 https://doi.org/10.1046/j.1365-2958.2000.01916.x

39   Coker, J.A. and DasSarma, S. (2007) Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet.* **8**, 61 https://doi.org/10.1186/1471-2156-8-61

40   Facciotti, M.T., Reiss, D.J., Pan, M., Kaur, A., Vuthoori, M., Bonneau, R. et al. (2007) General transcription factor specified global gene regulation in archaea. *Proc. Natl Acad. Sci. U.S.A.* **104**, 4630–4635 https://doi.org/10.1073/pnas.0611663104

41   Paytubi, S. and White, M.F. (2009) The crenarchaeal DNA damage-inducible transcription factor B paralogue TFB3 is a general activator of transcription. *Mol. Microbiol.* **72**, 1487–1499 https://doi.org/10.1111/j.1365-2958.2009.06737.x

42   Feng, X., Sun, M., Han, W., Liang, Y.X. and She, Q. (2018) A transcriptional factor B paralog functions as an activator to DNA damage-responsive expression in archaea. *Nucleic Acids Res.* **46**, 7465 https://doi.org/10.1093/nar/gky302

43   Schult, F., Le, T.N., Albersmeier, A., Rauch, B., Blumenkamp, P., van der Does, C. et al. (2018) Effect of UV irradiation on *Sulfolobus acidocaldarius* and involvement of the general transcription factor TFB3 in the early UV response. *Nucleic Acids Res.* **46**, 7179–7192 https://doi.org/10.1093/nar/gky527

44   Bell, S.D., Jaxel, C., Nadal, M., Kosa, P.F. and Jackson, S.P. (1998) Temperature, template topology, and factor requirements of archaeal transcription. *Proc. Natl Acad. Sci. U.S.A.* **95**, 15218–15222 https://doi.org/10.1073/pnas.95.26.15218

45   Nagy, J., Grohmann, D., Cheung, A.C., Schulz, S., Smollett, K., Werner, F. et al. (2015) Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun.* **6**, 6161 https://doi.org/10.1038/ncomms7161

46   Spitalny, P. and Thomm, M. (2003) Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. *J. Biol. Chem.* **278**, 30497–30505 https://doi.org/10.1074/jbc.M303633200

47   Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H. et al. (2011) The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43**, 263–274 https://doi.org/10.1016/j.molcel.2011.05.030

48   Naji, S., Grünberg, S. and Thomm, M. (2007) The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J. Biol. Chem.* **282**, 11047–11057 https://doi.org/10.1074/jbc.M611674200

49   Fouqueau, T., Zeller, M.E., Cheung, A.C., Cramer, P. and Thomm, M. (2013) The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.* **41**, 7048–7059 https://doi.org/10.1093/nar/gkt433

50   Goldman, S.R., Ebright, R.H. and Nickels, B.E. (2009) Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927–928 https://doi.org/10.1126/science.1169237

51   Blombach, F., Makarova, K.S., Marrero, J., Siebers, B., Koonin, E.V. and van der Oost, J. (2009) 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* **4**, 39 https://doi.org/10.1186/1745-6150-4-39

52   Blombach, F., Ausiannikava, D., Figueiredo, A.M., Soloviev, Z., Prentice, T., Zhang, M. et al. (2018) Structural and functional adaptation of *Haloferax volcanii* TFEalpha/beta. *Nucleic Acids Res.* **46**, 2308–2320 https://doi.org/10.1093/nar/gkx1302

53   Sainsbury, S., Niesser, J. and Cramer, P. (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437–440 https://doi.org/10.1038/nature11715

54   Dexl, S., Reichelt, R., Kraatz, K., Schulz, S., Grohmann, D., Bartlett, M. et al. (2018) Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gky699

55   Werner, F. (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27 https://doi.org/10.1016/j.jmb.2012.01.031

56   Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I. et al. (2018) Structural basis for transcript elongation control by NusG family universal regulators. *Cell* **173**, 1650–1662 e14 https://doi.org/10.1016/j.cell.2018.05.017

57   Sevostyanova, A., Belogurov, G.A., Mooney, R.A., Landick, R. and Artsimovitch, I. (2011) The beta subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 https://doi.org/10.1016/j.molcel.2011.05.026

58   Klein, B.J., Bose, D., Baker, K.J., Yusoff, Z.M., Zhang, X. and Murakami, K.S. (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl Acad. Sci. U.S.A.* **108**, 546–550 https://doi.org/10.1073/pnas.1013828108

59   Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M. and Sekine, S.I. (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 https://doi.org/10.1126/science.aan8552

60   Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C. and Cramer, P. (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 https://doi.org/10.1038/emboj.2011.64

61   Turtola, M. and Belogurov, G.A. (2016) Nusg inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* **5**, e18096 https://doi.org/10.7554/eLife.18096

62   Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E. et al. (2010) A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 https://doi.org/10.1126/science.1184953

63   Daniels, J.P., Kelly, S., Wickstead, B. and Gull, K. (2009) Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* **4**, 24 https://doi.org/10.1186/1745-6150-4-24

64   Spang, A., Saw, J.H., Jorgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E. et al. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 https://doi.org/10.1038/nature14447

65   Prather, D., Krogan, N.J., Emili, A., Greenblatt, J.F. and Winston, F. (2005) Identification and characterization of Elf1, a conserved transcription elongation factor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 10122–10135 https://doi.org/10.1128/MCB.25.22.10122-10135.2005

66   Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J. and Cramer, P. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278 https://doi.org/10.1038/nsmb.1903

67   Arnvig, K.B., Pennell, S., Gopal, B. and Colston, M.J. (2004) A high-affinity interaction between NusA and the *rrn nut* site in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. U.S.A.* **101**, 8325–8330 https://doi.org/10.1073/pnas.0401287101

68   Vogel, U. and Jensen, K.F. (1997) Nusa is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *J. Biol. Chem.* **272**, 12265–12271 https://doi.org/10.1074/jbc.272.19.12265

69   Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T. et al. (2007) Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem. Biophys. Res. Commun.* **355**, 122–128 https://doi.org/10.1016/j.bbrc.2007.01.119

70   Werner, F. (2013) Molecular mechanisms of transcription elongation in archaea. *Chem. Rev.* **113**, 8331–8349 https://doi.org/10.1021/cr4002325

71   Belogurov, G.A. and Artsimovitch, I. (2015) Regulation of transcript elongation. *Annu. Rev. Microbiol.* **69**, 49–69 https://doi.org/10.1146/annurev-micro-091014-104047

72   Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M. et al. (2018) Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* **69**, 816–827 e4 https://doi.org/10.1016/j.molcel.2018.02.008

73   Guajardo, R. and Sousa, R. (1997) A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **265**, 8–19 https://doi.org/10.1006/jmbi.1996.0707

74   Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A. and Nudler, E. (2005) A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 https://doi.org/10.1016/j.cell.2004.11.045

75   Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. and Block, S.M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 https://doi.org/10.1038/nature04268

76   Naji, S., Bertero, M.G., Spitalny, P., Cramer, P. and Thomm, M. (2008) Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res.* **36**, 676–687 https://doi.org/10.1093/nar/gkm1086

77   Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. and Kornberg, R.D. (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 https://doi.org/10.1126/science.1059495

78   Naryshkina, T., Kuznedelov, K. and Severinov, K. (2006) The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. *J. Mol. Biol.* **361**, 634–643 https://doi.org/10.1016/j.jmb.2006.05.034

79   Kuznedelov, K., Korzheva, N., Mustaev, A. and Severinov, K. (2002) Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* **21**, 1369–1378 https://doi.org/10.1093/emboj/21.6.1369

80   Kent, T., Kashkina, E., Anikin, M. and Temiakov, D. (2009) Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. *J. Biol. Chem.* **284**, 13497–13504 https://doi.org/10.1074/jbc.M901898200

81   Santangelo, T.J. and Reeve, J.N. (2010) Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J. Biol. Chem.* **285**, 23908–23915 https://doi.org/10.1074/jbc.M109.094565

82   Kotlajich, M.V., Hron, D.R., Boudreau, B.A., Sun, Z., Lyubchenko, Y.L. and Landick, R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *eLife* **4**, e04970 https://doi.org/10.7554/eLife.04970

83   Tornaletti, S., Reines, D. and Hanawalt, P.C. (1999) Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *J. Biol. Chem.* **274**, 24124–24130 https://doi.org/10.1074/jbc.274.34.24124

84   Brueckner, F. and Cramer, P. (2007) DNA photodamage recognition by RNA polymerase II. *FEBS Lett.* **581**, 2757–2760 https://doi.org/10.1016/j.febslet.2007.05.014

85   Charlet-Berguerand, N., Feuerhahn, S., Kong, S.E., Ziserman, H., Conaway, J.W., Conaway, R. et al. (2006) RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J.* **25**, 5481–5491 https://doi.org/10.1038/sj.emboj.7601403

86   Nudler, E. (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438–1445 https://doi.org/10.1016/j.cell.2012.06.003

87   Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W. and Darst, S.A. (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**, 335–345 https://doi.org/10.1016/S0092-8674(03)00600-7

88   Brochier, C., Forterre, P. and Gribaldo, S. (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* **5**, R17 https://doi.org/10.1186/gb-2004-5-3-r17

89   Hausner, W., Lange, U. and Musfeldt, M. (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**, 12393–9 https://doi.org/10.1074/jbc.275.17.12393

90   Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995) Transcription in archaea: similarity to that in eucarya. *Proc. Natl Acad. Sci. U.S.A.* **92**, 5768–5772 https://doi.org/10.1073/pnas.92.13.5768

91   Fouqueau, T., Blombach, F., Hartman, R., Cheung, A.C.M., Young, M.J. and Werner, F. (2017) The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nat. Commun.* **8**, 1914 https://doi.org/10.1038/s41467-017-02081-3

92   Laptenko, O., Kim, S.S., Lee, J., Starodubtseva, M., Cava, F., Berenguer, J. et al. (2006) pH-dependent conformational switch activates the inhibitor of transcription elongation. *EMBO J.* **25**, 2131–2141 https://doi.org/10.1038/sj.emboj.7601094

93   Tagami, S., Sekine, S., Kumarevel, T., Hino, N., Murayama, Y., Kamegamori, S. et al. (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978–982 https://doi.org/10.1038/nature09573

94   Ortmann, A.C., Brumfield, S.K., Walther, J., McInnerney, K., Brouns, S.J., van de Werken, H.J. et al. (2008) Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. *J. Virol.* **82**, 4874–4883 https://doi.org/10.1128/JVI.02583-07

95   Lennon, J.T. and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 https://doi.org/10.1038/nrmicro2504

96   Forterre, P. and Prangishvili, D. (2013) The major role of viruses in cellular evolution: facts and hypotheses. *Curr. Opin. Virol.* **3**, 558–565 https://doi.org/10.1016/j.coviro.2013.06.013

97   Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 https://doi.org/10.1038/nature15386

98   Sheppard, C. and Werner, F. (2017) Structure and mechanisms of viral transcription factors in archaea. *Extremophiles* **21**, 829–838 https://doi.org/10.1007/s00792-017-0951-1

99   Deng, L., Garrett, R.A., Shah, S.A., Peng, X. and She, Q. (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol. Microbiol.* **87**, 1088–1099 https://doi.org/10.1111/mmi.12152

100   Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A. and Marraffini, L.A. (2015) Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell* **161**, 1164–1174 https://doi.org/10.1016/j.cell.2015.04.027

101   Shah, S.A., Hansen, N.R. and Garrett, R.A. (2009) Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans.* **37**, 23–28 https://doi.org/10.1042/BST0370023

102   Sheppard, C., Blombach, F., Belsom, A., Schulz, S., Daviter, T., Smollett, K. et al. (2016) Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP. *Nat. Commun.* **7**, 13595 https://doi.org/10.1038/ncomms13595

103   Bogenhagen, D.F. and Brown, D.D. (1981) Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. *Cell* **24**, 261–270 https://doi.org/10.1016/0092-8674(81)90522-5

104   Muller, B., Allmansberger, R. and Klein, A. (1985) Termination of a transcription unit comprising highly expressed genes in the archaebacterium *Methanococcus voltae*. *Nucleic Acids Res.* **13**, 6439–6445 https://doi.org/10.1093/nar/13.18.6439

105   Reiter, W.D., Palm, P., Yeats, S. and Zillig, W. (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* **209**, 270–275 https://doi.org/10.1007/BF00329653

106   Thomm, M., Hausner, W. and Hethke, C. (1994) Transcription factors and termination of transcription in *Methanococcus*. *Syst. Appl. Microbiol.* **16**, 648–655 https://doi.org/10.1016/S0723-2020(11)80336-X

107   Santangelo, T.J. and Reeve, J.N. (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J. Mol. Biol.* **355**, 196–210 https://doi.org/10.1016/j.jmb.2005.10.062

108   Santangelo, T.J., Cubonova, L., Skinner, K.M. and Reeve, J.N. (2009) Archaeal intrinsic transcription termination in vivo. *J. Bacteriol.* **191**, 7102–7108 https://doi.org/10.1128/JB.00982-09

109 Arimbasseri, A.G., Rijal, K. and Maraia, R.J. (2013) Transcription termination by the eukaryotic RNA polymerase III. *Biochim. Biophys. Acta* **1829**, 318–330 https://doi.org/10.1016/j.bbagrm.2012.10.006

110 Dar, D., Prasse, D., Schmitz, R.A. and Sorek, R. (2016) Widespread formation of alternative 3′ UTR isoforms via transcription termination in archaea. *Nat. Microbiol.* **1**, 16143 https://doi.org/10.1038/nmicrobiol.2016.143

111 Chamieh, H., Ibrahim, H. and Kozah, J. (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* **576**(1 Pt 2), 214–228 https://doi.org/10.1016/j.gene.2015.10.007

112 Walker, J.E., Luyties, O. and Santangelo, T.J. (2017) Factor-dependent archaeal transcription termination. *Proc. Natl Acad. Sci. U.S.A.* **114**, E6767–E6E73 https://doi.org/10.1073/pnas.1704028114

113 Park, J.S., Marr, M.T. and Roberts, J.W. (2002) *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**, 757–767 https://doi.org/10.1016/S0092-8674(02)00769-9

114 Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J. et al. (2006) Structural basis for bacterial transcription-coupled DNA repair. *Cell* **124**, 507–520 https://doi.org/10.1016/j.cell.2005.11.045