

Evolution of multisubunit RNA polymerases in the three domains of life

Finn Werner and Dina Grohmann

Abstract | RNA polymerases (RNAPs) carry out transcription in all living organisms. All multisubunit RNAPs are derived from a common ancestor, a fact that becomes apparent from their amino acid sequence, subunit composition, structure, function and molecular mechanisms. Despite the similarity of these complexes, the organisms that depend on them are extremely diverse, ranging from microorganisms to humans. Recent findings about the molecular and functional architecture of RNAPs has given us intriguing insights into their evolution and how their activities are harnessed by homologous and analogous basal factors during the transcription cycle. We provide an overview of the evolutionary conservation of and differences between the multisubunit polymerases in the three domains of life, and introduce the ‘elongation first’ hypothesis for the evolution of transcriptional regulation.

Last universal common ancestor

The last common ancestor of all three contemporary domains of life.

Transcription factor

A protein that transiently interacts with RNA polymerase, modulates its protein-, DNA- and RNA-binding or catalytic properties, and thereby regulates transcription.

Convergent evolution

The acquisition of the same trait in unrelated lineages.

Homologous

Pertaining to genes: derived from the same ancestor.

*RNA Polymerase Laboratory,
Institute for Structural and
Molecular Biology,
Division of Biosciences,
University College London,
Darwin Building, Gower Street,
London WC1E 6BT, UK.
Correspondence to F.W.
e-mail:
werner@biochem.ucl.ac.uk*

doi:10.1038/nrmicro2507

RNA plays key roles in the expression of genes in all living organisms. As RNA is omnipresent, it is not surprising that the enzymes that synthesize it — RNA polymerases (RNAPs) — are highly conserved in evolution¹. Even though the polymerization of RNA has probably been ‘invented’ several times during evolution, as judged by the five structurally discrete and evolutionarily unrelated folds of RNAP active sites, all multisubunit RNAPs responsible for the transcription of cellular genomes have a common structural framework and operate by closely related molecular mechanisms², suggesting that the last universal common ancestor (LUCA) of the Archaea, Bacteria and Eukarya had an RNAP very similar to the simplest form of contemporary RNAPs found in the Bacteria. Whereas the Archaea and the Bacteria use a single type of RNAP to transcribe their entire gene repertoire, the Eukarya use several classes of RNAPs that are specialized to transcribe distinct and non-overlapping subsets of genes³. A plethora of transcription factors interact with their cognate RNAP to modulate its activities during the three distinct phases of the transcription cycle: initiation, elongation and termination (FIG. 1). The structure and function of some of these factors are conserved across the three domains, whereas some non-homologous factors show an intriguing level of structural and functional similarity, suggesting that convergent evolution has led to alternative means of facilitating the same process. Some homologous proteins that are permanently incorporated into RNAPs in one system are reversibly incorporated in another RNAP. Thus, the boundary between core RNAP subunits and

associated transcription factors is diffuse. Therefore, an understanding of the evolution of transcription machineries requires an appreciation of both RNAP subunits and transcription factors, and for hypotheses to be rooted on our structural, functional and mechanistic knowledge of transcription. Our deep understanding of multisubunit RNAPs in all three domains of life allows us to explain the molecular mechanisms of RNAP and transcription in unprecedented detail. In this Review, we focus on the current understanding of the structure and function of RNAP in an evolutionary context, and propose the ‘elongation-first’ hypothesis, according to which the RNAP of the LUCA was regulated during the elongation phase of transcription rather than the initiation phase.

Molecular anatomy of RNAP

Multisubunit RNAPs differ fundamentally from the single-subunit ‘right-handed’ RNAPs encoded by bacteriophages, such as T7 or SP6, which directly recognize promoter sequences without any requirements for accessory and regulatory factors^{4,5}. All multisubunit RNAPs resemble a crab claw, the ‘jaws’ of which interact with the downstream duplex DNA template^{6–10} (FIG. 2). In the transcribing RNAP, the downstream DNA projects along the floor of the major DNA-binding channel until it encounters the active centre at the RNAP ‘wall’. The DNA–RNA hybrid rises up from the active site perpendicular to the downstream duplex DNA, and the strands are separated by the RNAP ‘lid’ (REF. 7). The DNA–RNA hybrid and the DNA immediately downstream of the RNAP active centre are both secured by the ‘RNAP clamp’.

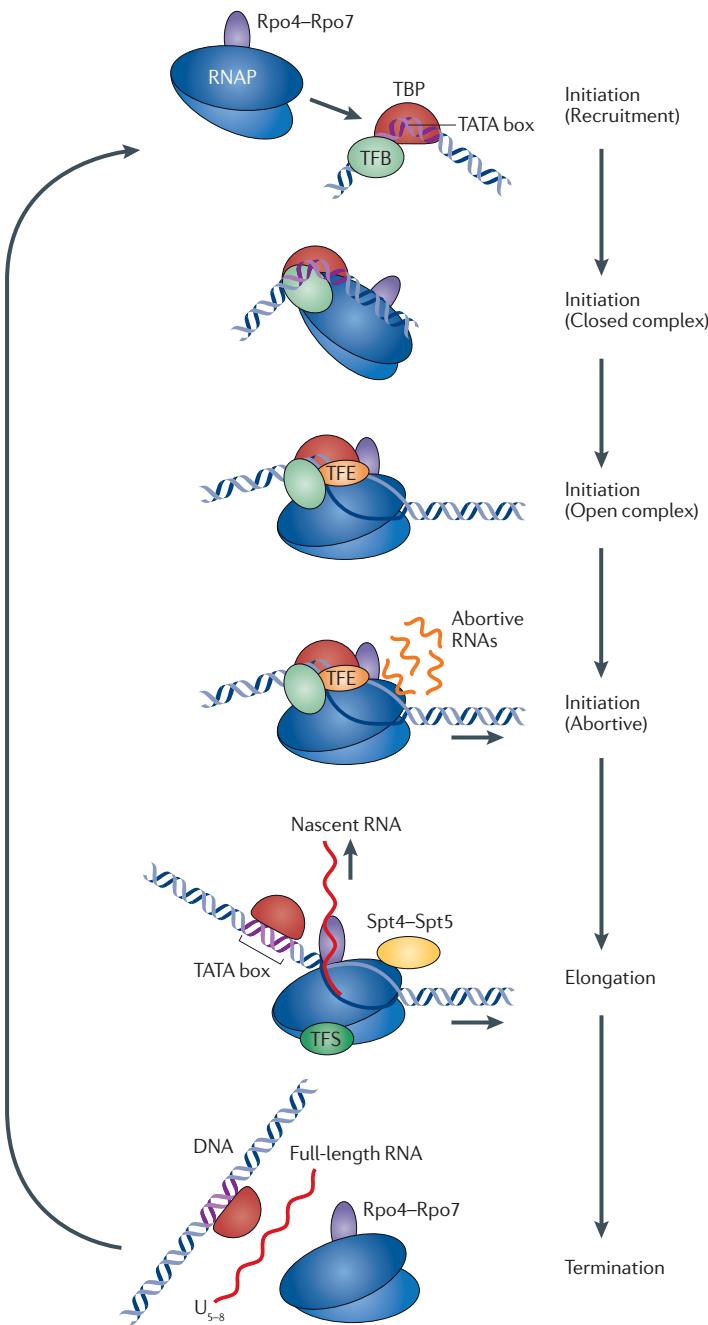


Figure 1 | The transcription cycle of the archaeal RNA polymerase. Multisubunit RNA polymerases (RNAPs) carry out transcription by repeatedly cycling through initiation, elongation and termination phases. RNAP activity is dependent on, and modulated by, exogenous transcription factors. In the Archaea, TATA box-binding protein (TBP), transcription factor B (TFB) and TFE facilitate transcription initiation, whereas TFS and Spt4–Spt5 regulate transcription elongation and Rpo4 and Rpo7 (also known as RpoF and RpoE, respectively) increase the processivity of the RNAP. Some initiation factors (for example, TBP) remain associated with the promoter ready for the recruitment of the next RNAP in the subsequent cycle, whereas other factors (for example, TFE) remain associated with elongating RNAPs¹²⁰. Transcription termination in the Archaea is facilitated by a poly(T) signal at the 3' end of the template gene.

RNAP clamp
A flexible RNA polymerase (RNAP) domain that is predicted to move over the DNA-binding channel during open-complex formation.

Bridge and trigger helices
Flexible RNA polymerase motifs that unfold and refold repeatedly during nucleotide addition and translocation.

In the archaeal and eukaryotic enzymes, the transcript is guided away from the elongating RNAP by interactions with the RNAP 'stalk' (REF. 11), whereas in the bacterial system, the flap tip helix might fulfil a similar function. The pore, or secondary channel, located under the active site, allows substrates and cleavage factors to access the active site, and allows extrusion of the transcript during backtracking¹².

All RNAP subunits can be divided into three overlapping functional classes^{1,2}: assembly platform subunits, which nucleate RNAP assembly; catalytic subunits, which form the catalytic core that harbours the active site, including the Mg²⁺-chelating residues, the bridge and trigger helices, the downstream DNA-binding and DNA–RNA hybrid-binding sites, the secondary NTP entry channel, and the loop and switch regions that are instrumental for the handling of the nucleic acid scaffold (including DNA strand separation); and auxiliary subunits, which include the RNAP stalk. The combination of assembly platform and catalytic subunits is the minimal configuration of active RNAPs. The auxiliary RNAP subunits are not strictly required for basic RNAP operations, including promoter-directed transcription, but add interaction sites with basal transcription factors and/or nucleic acids.

Structural and functional complexity of RNAP

All multisubunit RNAPs in the three domains of life are evolutionary related and contain homologues of the bacterial RNAP β -, β' -, α - and ω -subunits (FIG. 3a; TABLE 1). The eukaryotic RNAPII subunits are called RPB1–RPB12 (numbered from largest to smallest polypeptide, based on the subunits in *Saccharomyces cerevisiae*), whereas the archaeal RNAP subunits are named Rpo1–Rpo13, although an alternative letter designation is often used in the literature (TABLE 1). The two largest RNAP subunits — the β -subunit and the β' -subunit in the Bacteria, RPB1 and RPB2 in the Eukarya, and Rpo1 (also known as RpoA) and Rpo2 (also known as RpoB) in the Archaea — contain two double-psi β -barrel motifs that form the active site and are derived from a common ancestor, although it is difficult to recognize the structural similarity beyond the double-psi β -barrels owing to large insertions and deletions.

RNAP in the RNA world. According to the 'RNA world' hypothesis, the primordial RNAP was a ribozyme. It has been suggested that the ancestor of the contemporary β -subunit and β' -subunit was a homodimeric RNA-binding protein without any catalytic activity¹³. After the emergence of protein synthesis, this homodimer could have acted as a chaperone by binding to and improving the fitness of a ribozyme RNAP. According to this hypothesis, the homodimer subsequently evolved into a heterodimer with useful chemical properties at its interface, and the active site was transferred from the RNA to the protein cofactor. Then, the functionally obsolete RNA component was lost and the protein subunit complexity increased, resulting in the contemporary multisubunit RNAPs. However, this hypothesis remains speculative without the existence

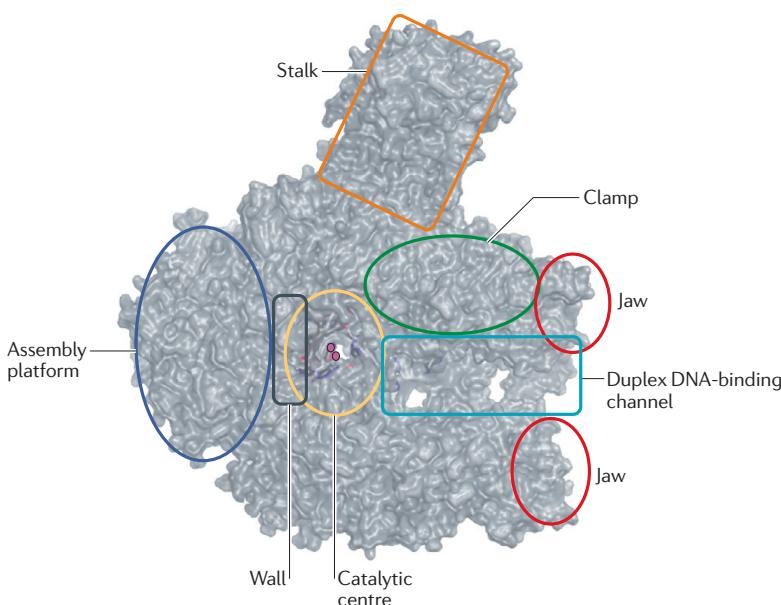


Figure 2 | Overall architecture of RNA polymerase. This simplified diagram of RNA polymerase shows important structural and functional features discussed in the main text, including the assembly platform, the active site, the DNA-binding channel, the jaws and the wall, clamp and stalk domains. The two active-site Mg²⁺ ions are indicated as magenta spheres. The structural information was obtained from eukaryotic RNAPII Protein Data Bank entry 1Y1W.

of experimental data or a naturally occurring ribozyme RNAP¹⁴. During evolution, the template specificity of RNAP gradually changed from RNA to DNA; bacterial RNAP and eukaryotic RNAPII still can use RNA templates in special circumstances, such as the regulation of transcription by 6S RNA¹⁵ and the replication of the hepatitis delta virus genome¹⁶, respectively.

The minimal core of RNAP. The two largest RNAP subunits are each encoded by one gene in most bacteria and eukaryotes, whereas many of the homologous archaeal subunits are split into two genes that are transcribed as one large polycistronic operon². The split sites do not affect the structure or function of the subunits, as has been demonstrated by introducing the archaeal split sites into the bacterial RNAP¹⁷ and by fusing the archaeal subunits (F.W., unpublished observations). In some bacterial lineages, including the genera *Helicobacter* and *Campylobacter*, the two large RNAP subunits are fused, presumably as a result of a frameshift in the intercistronic region¹⁸; furthermore, a synthetic fusion protein of the two large *Escherichia coli* RNAP subunits can be assembled to yield a catalytically active enzyme¹⁹. The evolutionary conservation of the large subunits and the presence of highly conserved blocks of sequence can be easily recognized in all multisubunit RNAPs²⁰, a fact that was further refined and elaborated in an analysis of multiple sequence alignments using a large sequence data set across all domains of life^{21,22}. The availability of high-resolution structural information about RNAPs allows the conserved sequence blocks to be mapped onto the three-dimensional structure of RNAP; these conserved

Double-psi β-barrel motif
A structural module in the two largest RNA polymerase subunits that reveals the common origin of these subunits and constitutes the active site.

Ribozyme
An enzyme made exclusively of RNA.

RNAi RNAP
An RNA polymerase (RNAP) involved in RNA interference (RNAi); it produces double-stranded RNA from aberrant single-stranded RNA templates to trigger RNAi.

regions correspond to mechanistically important motifs of RNAP, thereby linking evolution of sequence with RNAP function^{21,22}. Such analyses make it possible to predict a hypothetical minimal RNAP core, which in essence is composed of the two double-psi β-barrels. Interestingly, two hypothetical proteins encoded by *Corynebacterium glutamicum* (ORF Cgl1702) and the yeast 'killer' plasmid pGKL2 correspond to this design²³, although no experimental data on the expression, structure or catalytic activity of these putative ultra-minimal RNAPs are available. Other examples of minimal RNAPs are RNAi RNAPs such as QDE1, which consist of two identical subunits with one active site in each subunit²⁴. The active sites of QDE1 in *Neurospora crassa* have the same architecture as multisubunit DNA-dependent RNAPs, encompassing two double-psi β-barrels²⁵.

Despite a low sequence identity, the two main classes of multisubunit RNAP — from the Bacteria and from the Archaea and the Eukarya, respectively — share an extensive structural homology⁸ (FIG. 3a–e). The majority of the strictly conserved residues cluster around the RNAP active site, form the pore and are involved in the handling of the template and non-template DNA strands and RNA strands, or form flexible motifs, including the RNAP clamp and the switch regions²³. In addition to the well-characterized RNAP subunits that are conserved in the three domains of life, the archaeal and eukaryotic RNAPs share a complement of additional subunits¹ (TABLE 1). The archaea–eukaryote-specific subunits interact with many of the universally conserved RNAP subunits (FIG. 3c–e) and are not clustered at one particular site of the enzyme. The following section discusses how these subunits contribute to RNAP function.

Domain-specific RNAP subunits

A subset of RNAP subunits emerged following the split of the bacterial and archaeal–eukaryotic branches of the universal tree of life. These subunits are present in archaeal and/or eukaryotic RNAPs but have no homologues in bacteria. Although most of them are essential for cell viability in yeast, they are not strictly required for RNA polymerization. They make important interactions with basal transcription factors and the DNA–RNA scaffold of transcription complexes. Some of these subunits facilitate interactions between distinct RNAP subunits and between RNAP and basal transcription factors. Others make important contacts with the DNA template or the transcript RNA, resulting in a modulation of RNAP function during the transcription cycle.

The stalk. The most pronounced difference between the RNAPs of archaeal and eukaryotic cells and that of bacteria is the stalk (FIG. 2), which is located at the periphery of the archaeal and eukaryotic enzymes and comprises Rpo4–Rpo7 (also known as RpoF–RpoE) in archaea and RPB4–RPB7 in eukaryotes (FIG. 3b,d,e). The stalk is the most thoroughly characterised subunit of the archaea–eukaryote-specific RNAP subunits. The two subunits form a stable complex that reversibly associates with RNAPII in *S. cerevisiae*²⁶, but is stably (non-reversibly) incorporated into the archaeal RNAP²⁷. The

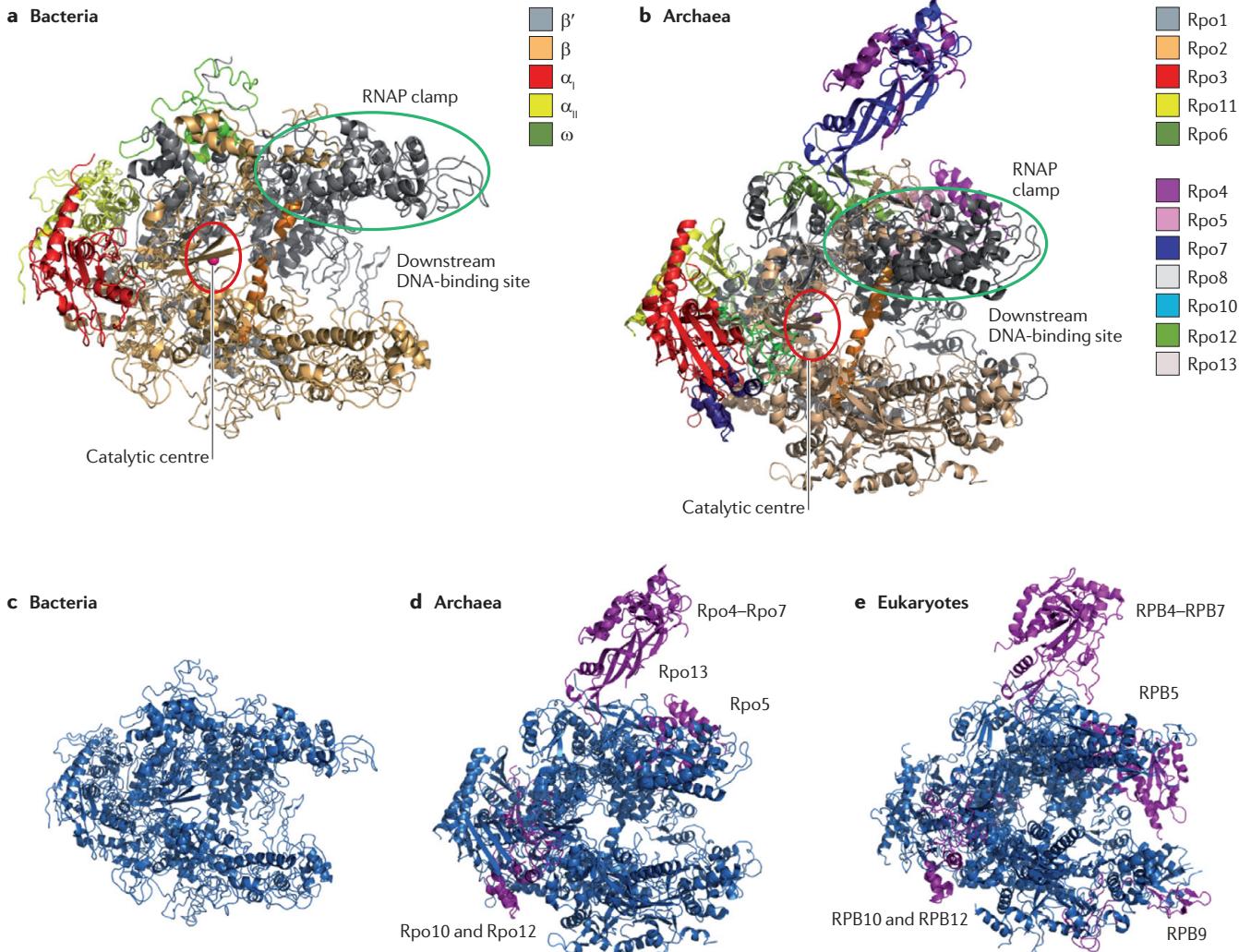


Figure 3 | Important structural and functional features of multisubunit RNA polymerases. **a,b** | The structures of the bacterial RNA polymerase (RNAP) (part **a**; based on the *Thermus aquaticus* Protein Data Bank entry [1I6V](#)) and the archaeal RNAP (part **b**; based on the *Sulfolobus shibatae* Protein Data Bank entry [2WAO](#)) reveal an evolutionarily conserved architecture. Homologous RNAP subunits are colour coded according to the key. The Mg^{2+} ion is highlighted in magenta, and the bridge helix is in orange. **c–e** | The locations of RNAP subunits that are absent in bacteria (part **c**) but present in archaea (part **d**) and eukaryotes (part **e**) are shown. Universally conserved RNAP subunits are blue, and those that are unique to the archaeal–eukaryotic lineage are magenta.

stalk has multiple roles during the transcription cycle^{14,28}: it promotes open-complex formation during transcription initiation²⁹ and facilitates the action of the basal transcription factor TFE^{30,31}. During elongation, Rpo4–Rpo7 and RPB4–RPB7 increase the processivity of RNAP, and Rpo4–Rpo7 also augment transcription termination¹¹.

Assembly platform. The assembly of RNAPs requires an assembly platform³². In eukaryotes, RPB10 and RPB12 (homologous to archaeal Rpo10 (also known as RpoN) and Rpo12 (also known as RpoP)) form a stable complex with RPB3–RPB11 (homologous to archaeal Rpo3–Rpo11 (also known as RpoD–RpoL)), and this complex is homologous to the bacterial α -subunit homodimer³³ (FIG. 3a,b; TABLE 1). In contrast to the bacterial RNAP, which only requires the α -subunit homodimer (FIG. 3a),

all four archaeal subunits are necessary for the efficient assembly and stability of archaeal RNAP³² (FIG. 3b). Rpo10 and Rpo12, and RPB10 and RPB12, fill concave depressions in the second-largest RNAP subunit (Rpo2 and RPB2, respectively) and thereby act as structural adaptors between Rpo2 and Rpo3 or RPB2 and RPB3, respectively (FIG. 3d,e); this explains, at least in part, their role during RNAP assembly and stability. RPB10 and RPB12 are found in the three main classes of eukaryotic RNAPs, but their interaction partners differ in each class; the assembly platform subunits of RNAPI and RNAPIII (AC19–AC40 in both) are distinct from RPB3–RPB11, and the second-largest catalytic subunits of RNAPI and RNAPIII (A135 and C128, respectively) are distinct from RPB2, suggesting that RPB10 and RPB12 have additional functions beyond RNAP assembly. It is

Open-complex formation
The structural transition of RNA polymerase concomitant with the melting of DNA strands during initiation.

Table 1 | Conserved RNA polymerase (RNAP) subunits and transcription factors*

Bacteria	Archaea	Eukaryotes				
		RNAPII	RNAPIII	RNAPI	Plant RNAPIV‡	Plant RNAPV‡
RNAP subunits						
β-subunit	Rpo1 (RpoA)	RPB1	C160	A190	NRPD1	NRPE1
β-subunit	Rpo2 (RpoB)	RPB2	C128	A135	NRPD/E2	NRPD/E2
α-subunit	Rpo3 (RpoD)	RPB3	AC40	AC40	RPB3 [1]	RPB3 [1]
α-subunit	Rpo11 (RpoL)	RPB11	AC19	AC19	RPB11	RPB11
ω-subunit	Rpo6 (RpoK)	RPB6	RPB6	RPB6	RPB6 [1]	RPB6 [1]
	Rpo5 (RpoH)	RPB5	RPB5	RPB5	RPB5 [3]	NRPE5
	Rpo8§ (RpoG)	RPB8	RPB8	RPB8	RPB8 [1]	RPB8 [1]
	Rpo10 (RpoN)	RPB10	RPB10	RPB10	RPB10	RPB10
	Rpo12 (RpoP)	RPB12	RPB12	RPB12	RPB12	RPB12
	Rpo4 (RpoF)	RPB4	C17	A14	NRPD/E4	NRPD/E4
	Rpo7 (RpoE)	RPB7	C25	A43	NRPD7 [1]	NRPE7
		RPB9	C11	A12	NRPD9b	RPB9
	Rpo13§					
Transcription factors						
		TFIIFα (RAP74)	C53 (C4)	A49		
		TFIIFβ (RAP30)	C37(C5)	A34.5		
	TFEα	TFIIEα	C82			
	TFEβ/C34§	TFIIEβ	C34			
			C31			
	TBP	TBP	TBP	TBP		
	TFB	TFIIB	BRF1			
		TFIIA				
		TFIIC				
	TFS	TFIIS	TFIIS			
	Spt4	SPT4	SPT4			
NusG	Spt5	SPT5	SPT5			
NusA	NusA					
Rho						
σ-factors						

TBP, TATA box-binding protein. *Alternative names that are common in the literature are shown in brackets. ‡The numbers in square brackets indicate the number of orthologues of RNAPIV and RNAPV subunits. §Found in some but not all archaeal species.

TATA box-binding protein
One of the two minimal transcription factors required for initiation by archaeal RNA polymerase (RNAP) and eukaryotic RNAPII. It binds to a sequence recognition motif in promoters that is called the TATA element or TATA box.

TFIIB
The second of the two minimal transcription factors required for initiation by archaeal RNA polymerase (RNAP) and eukaryotic RNAPII.

possible that RPB10 and RPB12 are incorporated into multiple classes of RNAPs owing to functional and/or physical interactions with basal transcription factors that facilitate transcription of all eukaryotic and archaeal RNAPs, such as TATA box-binding protein (TBP). Indeed, RPB10 and RPB12 are localized in proximity to TBP in a structural model of the DNA–TBP–TFIIB–RNAPII transcription initiation complex³⁴. In addition, the archaeal homologue Rpo12 plays a part during transcription initiation by promoting DNA melting and stabilizing the open complex³⁵.

DNA melting. Like RPB10 and RPB12, RPB5 is present in all eukaryotic RNAPs, and the archaeal homologue Rpo5 (also known as RpoH) plays a part in DNA melting and early transcription³⁶. RPB5 consists of two discrete domains; the carboxy-terminal domain, which corresponds to the full-length Rpo5, makes intricate contacts with the C terminus of the largest RNAP subunit (RPB1 and Rpo1; FIG. 3b). Similarly to RPB5 and RPB1 in eukaryotes, Rpo5 and a fragment of Rpo1 form the lower jaw domain of RNAP in archaea, which is more extended than its bacterial counterpart^{8,9} (FIG. 3). The jaw interacts with the downstream duplex DNA and undergoes substantial conformational changes between the initiation and elongation phases of transcription^{36,37}.

RPB8 and Rpo8 are located at the underside of the RNAP between the assembly platform and the pore. These homologues contain a nucleic acid-binding OB-fold³⁸ and could potentially interact with the 3' end

of the nascent transcript in backtracked elongation complexes, as this end is extruded through the pore³⁹. Yeast Rpb8 is essential, but its precise function during transcription is not clear⁴⁰. In archaea, Rpo8 is present only in the Crenarchaeota, one of two main archaeal phyla, reflecting the closer kinship of the Crenarchaeota to the Eukarya compared with the relationship between the Eukarya and the other main archaeal phylum, the Euryarchaeota⁴¹ (TABLE 1).

Domain-specific subunits. Some subunits are specific for a single domain. RPB9 is probably the only subunit found exclusively in eukaryotic RNAPs. RPB9 influences the interaction of RNAP with the basal factor TFIIF, and therefore transcription start site selection and transcription fidelity^{42,43}. Rpo13 is the only archaea-specific subunit; its function is unknown, and it is only present in a subset of archaeal genomes⁹.

Mechanisms of transcription initiation

Promoter-directed transcription requires sequence-specific recruitment of RNAP to the promoter, initiation of RNA polymerization in a primer-independent manner and efficient escape from the promoter (FIG. 1), all of which are stimulated by evolutionarily unrelated basal initiation factors in the Bacteria and in the Eukarya and the Archaea. However, as the molecular mechanisms of initiation are the same in all three domains, the mechanisms of action of the non-homologous factors are closely related. They have similar structures, utilize the same RNAP-binding sites and carry out nearly identical functions. This is an important insight into the evolution of transcription machineries and provides an excellent example of convergent evolution.

Bacteria. The bacterial core RNAP associates with a σ-factor to form ‘holo’-RNAP, which is directly recruited to the promoter in a sequence-specific manner (FIG. 4a,b). The canonical σ-factors consist of four conserved regions that contribute in distinct ways to transcription initiation⁴⁴. The σ-factors determine promoter specificity by binding directly to the –10 element in the promoter through region 2 and to the –35 element through region 4; they also interact with RNAP in a complex way to recruit the polymerase (predominantly through region 2 and region 4, but also through other regions), facilitate DNA melting and template strand loading during the closed-to-open complex transition (through region 2), stabilize the binding of the initiating nucleotide substrate and affect abortive initiation (through region 3.2), and modulate promoter escape (through multiple regions)⁴⁵.

In addition to the ‘housekeeping’ factor σ⁷⁰ (also known as RpoD), numerous alternative σ-factors (up to 60 in *Streptomyces coelicolor*), including σ^S (also known as RpoS), σ^{S4} and σ^{S2}, direct RNAP to subsets of genes that are induced, for example, in stationary phase, under phage shock conditions and under heat shock conditions, respectively⁴⁶. In *Bacillus subtilis*, one alternative σ-factor consists of two subunits, YvrHa (also known as RsoA) and YvrI (also known as SigO)^{47,48}. YvrI binds to the RNAP flap motif and mediates promoter recognition

of the –35 element, similarly to region 4 in σ⁷⁰, whereas YvrHa interacts with the –10 element of the promoter and the coiled coil of the RNAP clamp, thereby stabilizing the open complex⁴⁷, similarly to region 2 in σ⁷⁰ (FIG. 4a,b).

Whereas transcription initiation facilitated by the σ⁷⁰ holo-RNAP is a spontaneous, energy-independent process, transcription initiation by the σ^{S4} holo-RNAP requires an activator from the enhancer-binding protein family (which is part of the AAA+ family) and hydrolysis of ATP⁴⁹. Transcription initiation complexes formed by σ^{S4} holo-RNAP are trapped in a conformation that is not conducive to DNA strand separation, and ATP hydrolysis mediated by the RNAP-bound enhancer-binding proteins induces a conformational change in the transcription initiation complex that overcomes this restriction and results in efficient induction of genes under the control of σ^{S4} promoters⁴⁹. Conceptually, this mechanism is reminiscent of the RNAPII system, in which transcription initiation from many promoters is strongly enhanced by, if not dependent on, the hydrolysis of ATP by the basal transcription factor TFIID, which facilitates DNA strand melting⁵⁰. However, the underlying mechanisms are distinct and not conserved in evolution^{49,50}.

Archaea and eukaryotes. The archaeal RNAP and eukaryotic RNAPII have identical minimal requirements for homologous basal transcription initiation factors, and these factors operate via the same mechanisms (TABLE 1). Neither RNAP can be recruited to the promoter without the aid of additional factors; instead, the RNAP recognizes a complex of basal factors, including TBP and TFIIB, pre-assembled on the promoter². Unlike bacterial σ-factors⁵¹, TBP and TFIIB can be recruited to the promoter independently of RNAP. TBP consists of a highly conserved DNA-binding core domain and a less conserved eukaryote-specific amino-terminal domain⁵². The TBP core domain has a bipartite symmetrical structure encoded by a sequence repeat, which suggests that it evolved by a gene duplication event. However, no eukaryotic or archaeal protein with a single TBP repeat has been identified. Only one other protein, a bacterial RNase H subtype (RNase HIII), contains a domain that is homologous to a single TBP repeat, and this domain probably facilitates interactions with its DNA–RNA substrate⁵².

TFIIB consists of two discrete domains connected by a flexible linker. The C-terminal core domain of TFIIB binds to the TATA box–TBP complex and makes contacts with the DNA both upstream and downstream of the TATA box. From a structural perspective, the TBP–TFIIB–TATA box complex is reminiscent of the complex formed by region 4 of σ⁷⁰ and the –35 element of bacterial promoters, which forms independently of RNAP⁵³. Although the structure of the DNA–TBP–TFIIB–RNAPII complex has not been solved, partial crystal structures combined with cross-linking and cleavage experiments have allowed a model of the initiation complex to be generated^{34,54,55}. In this model, eukaryotic and archaeal RNAPs make very few direct contacts with the DNA in the closed complex, similar to bacterial RNAP. Instead, archaeal and eukaryotic RNAPs

σ-factor

A bacterial transcription initiation factor that binds specific sequences in the promoter. Each σ-factor regulates the transcription of a specific set of genes.

–10 and –35 elements

Key sequence motifs of bacterial promoters that are recognized by the transcription factor σ⁷⁰. Promoter strength is in part regulated by the strength of the interaction between σ⁷⁰ and these elements.

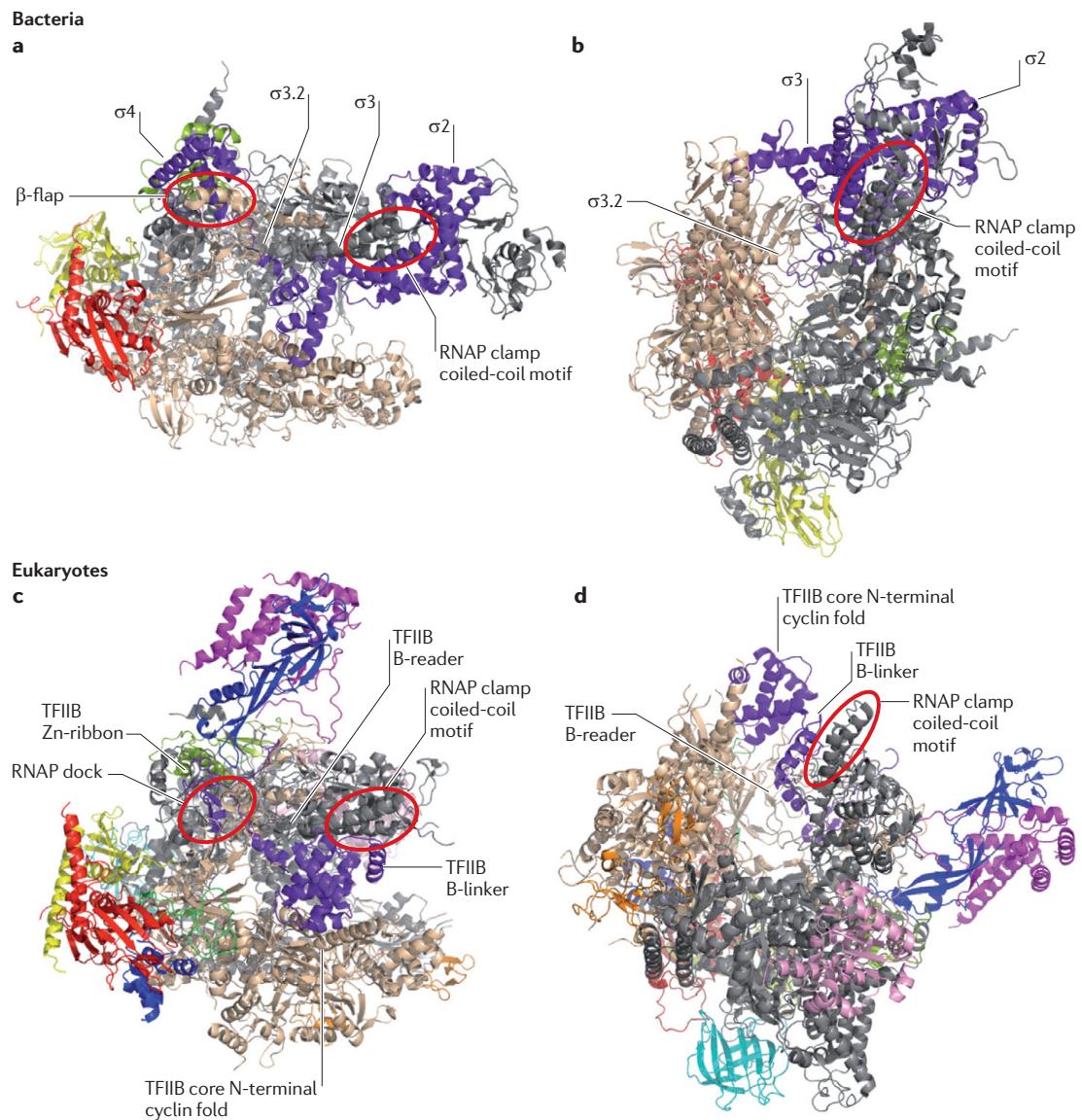


Figure 4 | Transcription initiation — holo-RNA polymerase structures from bacteria and eukaryotes. **a,b** | The bacterial core RNA polymerase (RNAP) forms a stable holoenzyme complex with the initiation factor σ^{70} (also known as RpoD) (based on the *Thermus thermophilus* Protein Data Bank entry 2A6E). The holoenzyme structure is shown in top (part **a**) and front (part **b**) views. The density of the bacterial β' -subunit downstream of region 2 is lineage specific and not representative of all bacterial RNAPs. **c,d** | Eukaryotic RNAPII forms a stable complex with basal transcription factor IIB (TFIIB) (based on the *Saccharomyces cerevisiae* Protein Data Bank entry 3K1F). The RNAPII-TFIIB structure is shown in top (part **c**) and front (part **d**) views. Important features in RNAP (the β -flap, RNAP dock and clamp coiled-coil motifs), TFIIB (the Zn-ribbon, the B-reader, the B-linker and the core amino-terminal cyclin repeats) and σ^{70} (regions 2, 3, 3.2 and 4) are indicated. All RNAP subunits are colour coded as in FIG. 1.

Zn-ribbon domain

A domain found in many transcription factors, including TFIIB, TFIIS and TFIIE in eukaryotes (and TFB, TFS and TFE, respectively, in archaea). The amino-terminal Zn-ribbon domain of TFIIB and TFB interacts with the RNA polymerase (RNAP) dock domain and is important for efficient RNAP recruitment.

are anchored to the promoter via multiple interactions with TFIIB. The N-terminal Zn-ribbon domain of TFIIB interacts with the dock domain of RNAP (FIG. 4c), and the C-terminal core domain is positioned across the DNA-binding channel (FIG. 4c,d). The RNAP clamp coiled-coil motif, which is conserved across the three domains of life, is an important binding site for region 2 of σ^{70} and for TFIIB^{34,56} (FIG. 4). The highly flexible linker region that connects the TFIIB domains (consisting of the B-reader helix and the B-linker) penetrates deep into the active centre of RNAP (FIG. 4c,d). The linker can be

cross-linked to the template DNA strand⁵⁷. The B-reader is displaced by the growing RNA transcript (longer than 5 nucleotides), whereas the B-linker is displaced by the rewinding of upstream DNA during TFIIB release and promoter escape⁵⁸. The position of the TFIIB Zn-ribbon and its interaction with RNAP, as well as the TFIIB core domain and the B-reader, are reminiscent of region 4, region 3 and region 3.2 of σ^{70} , respectively^{34,59} (FIG. 4a,c).

Bacterial, archaeal and eukaryotic RNAPs form a ‘composite’ active site that is complemented by the cognate initiation factor, resulting in a higher affinity

for the initiating nucleotide substrate⁶⁰ and stimulating catalysis³¹. Thus, despite a lack of similarity between these transcription factors at the sequence level, the interaction networks between RNAP and σ -factors in bacteria and between RNAP and TFIIB in archaea and eukaryotes are strikingly similar.

All RNAPs enter a non-productive phase of transcription following recruitment to the promoter called abortive cycling, during which the downstream DNA template is repeatedly reeled in⁶¹, and small transcripts of 3 to 9 nucleotides are synthesized and released without the RNAP disengaging from the promoter⁶². The exact mechanical nature of abortive initiation is still unclear, but it is likely to be caused by several factors. First, the linker regions of σ^{70} (region 3.2) and TFIIB clash sterically with the growing RNA chain^{34,63}. Second, the interactions within the initiation complex between RNAP and σ -factors or TFIIB need to be disrupted during promoter escape, and this presents a substantial energy barrier. In all initiation complexes, multiple low-affinity interactions between initiation factors and RNAP combine to form a stable complex^{34,45,59}. These interactions guarantee efficient recruitment of RNAP to the promoter and simultaneously enable dissociation by small conformational changes of the complex, during which the individual contacts are broken in a stepwise manner⁴⁵.

In archaea and eukaryotes, additional basal factors contribute to transcription initiation, including TFE, TFIIE, TFIIF and TFIIH (TABLE 1). The interaction of TFIIE with the RNAP clamp is not strictly required for initiation but stimulates DNA melting and stabilizes the open complex^{29,31,64}. The α -subunit of TFIIE shows very weak sequence homology to region 2 and region 4 of σ^{70} , and the β -subunit shows a weak homology to region 3 (REFS 65,66). The eukaryote-specific factor TFIIF also displays very weak sequence similarities to σ^{70} region 4 (REF. 67). However, it is unlikely that these proteins are bona fide homologues, as there is no apparent structural homology or shared RNAP-binding sites between TFIIE, TFIIF and the σ -factors. TFIIF interacts with RNAPII during transcription initiation and elongation^{68–70}. Paralogous TFIIF-like factors are involved in transcription in all three major eukaryotic RNAP systems: the general transcription factor TFIIF (the α -subunit and the β -subunit; also known as RAP74 and RAP30, respectively) in RNAPII, A49–A34.5 in RNAPI and C53–C37 (also known as C4–C5) in RNAPIII⁷¹ (TABLE 1).

Modulation of elongation and termination

Following promoter escape, RNAP enters the elongation phase of transcription, which is modulated by DNA and RNA sequences (including secondary structures) and RNAP subunits, and regulated by general and gene-specific transcription factors^{72,73}. Transcription elongation is intrinsically discontinuous and is interrupted by frequent pausing, stalling and arrest. The two important parameters that vary during elongation are the translocation rate (the average number of nucleotides polymerized per second) and the processivity (the number of nucleotides polymerized per initiation event). Transcription pausing can be instrumental for the regulation and

timing of gene expression but can also decrease RNA synthesis, as pausing reduces processivity⁷⁴. The catalytic cycle and the resulting mechanism of RNAP translocation along the DNA template involves alternative structures of the bridge and trigger helices in the active site⁷⁵. In the ternary elongation complex (TEC), composed of RNAP–DNA–RNA (FIG. 5), RNAP interacts with the downstream duplex DNA template, the DNA–RNA hybrid and the nascent RNA transcript⁷⁶. The interaction of the archaea–eukaryote-specific RNAP stalk with the RNA is dynamic, and the path of the transcript could not be resolved by X-ray crystallography. Whereas the regions that interact with the duplex DNA and DNA–RNA hybrid are highly conserved in all RNAPs, the stalk (consisting of RNAP subunits RPB4–RPB7) (FIG. 5c,d) is not present in bacterial RNAP (FIG. 5a,b). In the bacterial TEC, the nascent transcript is secured by the RNAP flap domain (FIG. 5a), which also serves as the primary interaction site with the elongation termination factor NusA⁷⁷ (TABLE 1). Both bacterial and archaeal NusA interact with RNA, but it is unclear whether archaeal NusA modulates transcription or indeed interacts with the RNAP, as it lacks the N-terminal domain that is present in bacterial NusA and that facilitates recruitment to the RNAP^{2,78,79}. Two lines of evidence demonstrate that the stalk is involved in elongation for eukaryotic and archaeal RNAP: recombinant archaeal RNAP lacking Rpo4–Rpo7 has a substantially lowered processivity *in vitro*¹¹, and *S. cerevisiae* RNAPII lacking RPB4 is depleted in the 3' regions of long ORFs *in vivo*⁸⁰. Rpo4–Rpo7 modulates both transcription elongation and termination via two mechanisms: by interacting with the nascent transcript, and by an allosteric mechanism that probably involves a repositioning of the RNAP clamp⁸¹ (FIG. 5c,d). As Rpo4–Rpo7-like complexes increase the processivity of transcription, they may enable RNAPs to transcribe longer genes^{14,28}. This probably applies only to eukaryotes, as eukaryotic genes are up to two orders of magnitude longer than bacterial and archaeal genes. However, it does not argue against the hypothesis that stalk-like complexes in ancestral eukaryotic RNAPs enabled the expansion of ORF size.

Rescue of arrested elongation complexes

Paused RNAPs have a tendency to move in a retrograde direction along the DNA template *in vivo* and *in vitro*. During this 'backtracking', the RNA 3' end is extruded from the RNAP through the pore and RNA polymerization cannot occur. RNAPs can overcome this impediment by cleaving the transcript internally, releasing short (3 to 18 nucleotides) RNA 3'-cleavage products and creating a new 3'-OH on the RNA that is aligned in the active site and conducive to catalysis^{82,83}. This endonucleolytic cleavage activity of RNAP is prominent at elevated pH and stimulated by transcript cleavage factors under physiological conditions. In eukaryotes and archaea, TFIIS and TFS, respectively, stimulate transcript cleavage^{84,85}, and the structure of the yeast RNAPII–TFIIS complex provides insights into its molecular mechanism⁸⁶. TFIIS is recruited to the TEC via interactions between TFIIS domain II and the RNAP

Paralogous

Pertaining to genes: separated by a gene duplication event.

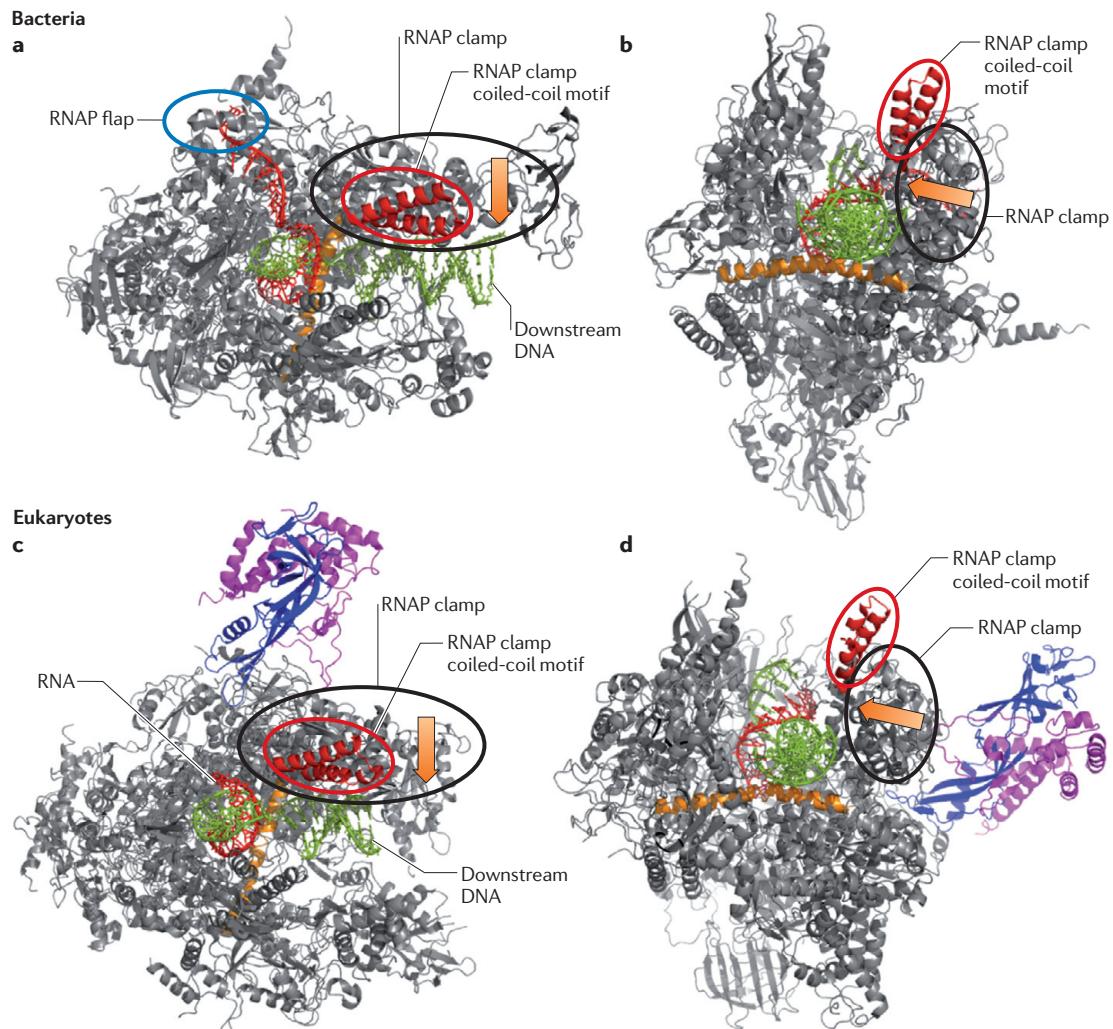


Figure 5 | The architecture of the transcription elongation complex. The ternary elongation complexes (TECs) of *Thermus thermophilus* RNA polymerase (RNAP) (parts **a,b**; Protein Data Bank entry [2O5I](#)) and *Saccharomyces cerevisiae* RNAPII (parts **c,d**; Protein Data Bank entry [1Y1W](#)) are shown in top (parts **a,c**) and front (parts **b,d**) views. RPB4 is the magenta ribbon, and RPB7 is the blue ribbon. The RNAP clamp domain coiled-coil motif is the red ribbon, and the bridge helix is the orange ribbon. Template and non-template DNA strands are green, and the RNA transcript is red. The RNAP flap secures the nascent transcript in the bacterial TEC. The arrows indicate an inwards closing movement of the RNAP clamp over the DNA-binding channel in response to an association of RNAP with RPB4–RPB7 in eukaryotes, and possibly with the elongation factor NusG in bacteria, and with Spt4–Spt5 and SPT4–SPT5 in archaea and eukaryotes, respectively.

jaw, while domain III is inserted into the active site through the pore (FIG. 6). Domain II forms a Zn-ribbon domain with a thin protruding β -hairpin, which complements the active site without either denying access of NTP substrates to the active site or blocking the extrusion of the transcript through the pore. Two invariant acidic residues on the tip of the hairpin are essential for the stimulatory effect of TFIIS on transcript cleavage⁸⁷ (FIG. 6). They are brought into close proximity to the Mg²⁺ ions in the active site and probably alter the binding characteristics of the metal ions and/or modulate the structure and location of the RNA or DNA–RNA hybrid in the active site in a manner that stimulates endonucleolysis⁸⁶ (FIG. 6). The archaeal TFS resembles both TFIIS and RPB9 at the sequence level; it has the domain structure of RPB9, but carries out the function

of TFIIS⁸⁴. Like TFIIS and TFS, the bacterial GreB stimulates transcript cleavage of its cognate RNAP and thereby augments transcription elongation^{85,88}. Although GreB is not evolutionarily related to TFIIS in sequence or structure, their interactions with RNAP and their molecular mechanisms of action are very similar. GreB is recruited to the bacterial TEC by interacting with the RNAP jaw domain, while its N-terminal coiled-coil domain is inserted into the active site through the pore. As in TFIIS, two acidic residues positioned at the tip of the coiled-coil domain are brought into close proximity to the active site Mg²⁺ ions and are crucial for GreB activity⁸⁹. As most if not all genes contain frequent pause sites, TFIIS and GreB regulate RNAP activity by increasing its processivity and thereby increasing the overall transcription elongation rate.

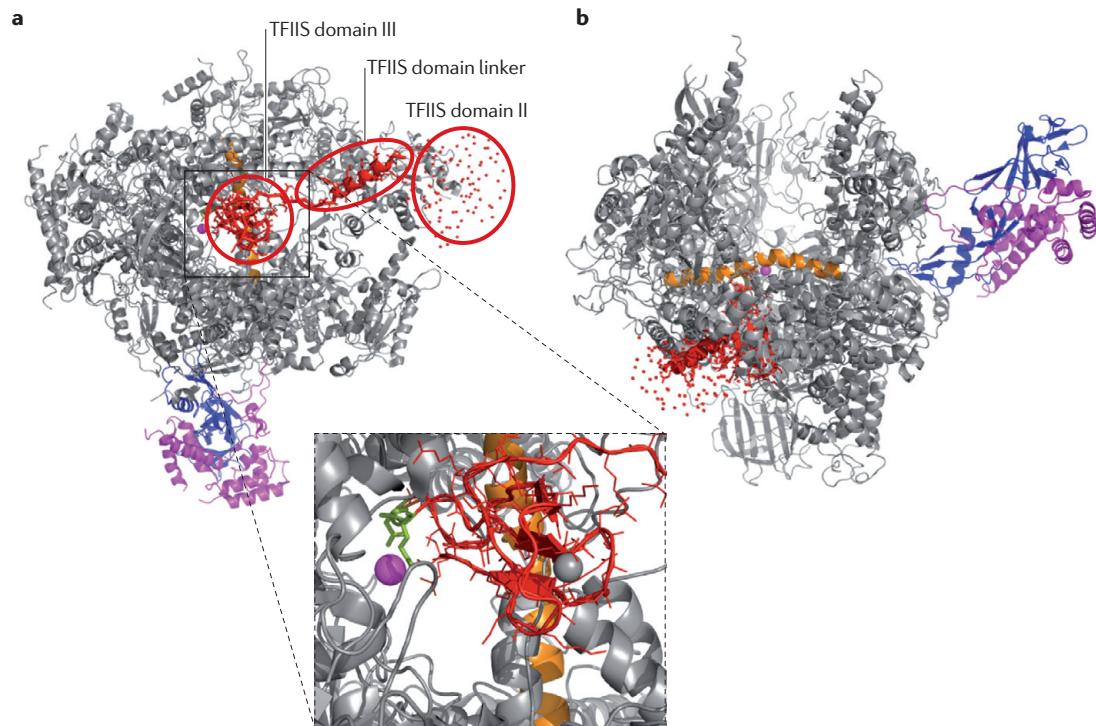


Figure 6 | The transcript cleavage complex. The transcript cleavage factor TFIIS (red) forms a stable complex with RNA polymerase II (RNAPII) (based on the *Saccharomyces cerevisiae* Protein Data Bank entry 1Y1Y), shown in bottom (part **a**) and front (part **b**) views. The TFIIS domain II interacts with the RNAPII jaw (resolved at lower resolution and shown as dots), while domain III penetrates deeply into the RNAPII active site through the pore. Two acidic residues (green) are located in close proximity to the Mg^{2+} ion (magenta sphere). The RNAPII bridge helix and RNAP subunits RPB4 and RPB7 are orange, magenta and blue ribbons, respectively.

The mechanisms of TFIIB, σ -factors, TFIIS and GreB bear some similarity, as all of these factors invade the catalytic centre of RNAP (either via the major DNA-binding channel or by the pore), complement the active site, alter the interactions with nucleic acids, Mg^{2+} ions and NTP substrates, and thereby modulate the catalytic properties of RNAP. Notably, neither initiation factors (TFIIB and σ -factors) nor cleavage factors (TFIIS and GreB) are evolutionarily conserved between the Archaea and Eukarya and the Bacteria; instead, they have adapted a similar structure to interact with RNAP and execute the same function, by convergent evolution.

Transcription elongation, NusG and Spt5

Decreased processivity and transcription termination are widely used to regulate gene expression in bacteriophages, and in ribosomal (*rrn*) operons in bacteria⁹⁰. The association of bacteriophage-encoded (such as phage λ anti-termination protein Q) and/or bacterial (such as NusA, NusB, NusE and NusG) elongation factors with RNAP converts the TEC into a termination-resistant anti-termination complex, which can transcribe genes beyond termination signals in the phage λ genome or downstream of the 5' leader region of *rrn* operons^{91–93}. Only NusG is universally conserved (FIG. 7; TABLE 1). NusG and its archaeal and eukaryotic homologues, Spt5 and SPT5, respectively, associate with their cognate RNAPs and enhance transcription elongation by

stimulating processivity and possibly the elongation rate of RNAP. Archaeal Spt5 and bacterial NusG consist of an N-terminal NusG (NGN) domain and a C-terminal Kypridis–Ouzounis–Woese (KOW) domain (FIG. 7a). In contrast to the non-homologous basal factors that facilitate transcription initiation, the NGN domain structure, the NusG–RNAP interaction sites and the stimulatory activity on transcription elongation are conserved in archaeal Spt5 (REFS 94,95). Eukaryotic SPT5 is much larger owing to the presence of four to six copies of the KOW domain, and two heptad repeat motifs that regulate SPT5 through phosphorylation⁹⁶.

Archaeal and eukaryotic NGN domains form stable complexes with Spt4 and SPT4, respectively, a protein that is not conserved in bacteria (FIG. 7). The bacterial and archaeal NGN domains are sufficient for RNAP binding and stimulation of transcription elongation^{94,95}. Archaeal Spt5 and bacterial NusG associate with RNAP via interactions between a hydrophobic cavity in the NGN domain (FIG. 7b,c) and the tip of the RNAP clamp coiled-coil domain, which protrudes from the RNAP surface^{94,95} (FIG. 4). Interestingly, the clamp coiled-coil domain is also an important RNAP recruitment site during initiation in all RNAPs, through the interaction with region 4 of σ^{70} and with TFIIB^{34,56}. This overlapping binding site might have an important role during promoter escape or a stalled RNAP proximal to the promoter, where elongation factors (Spt5, SPT5, NusG or

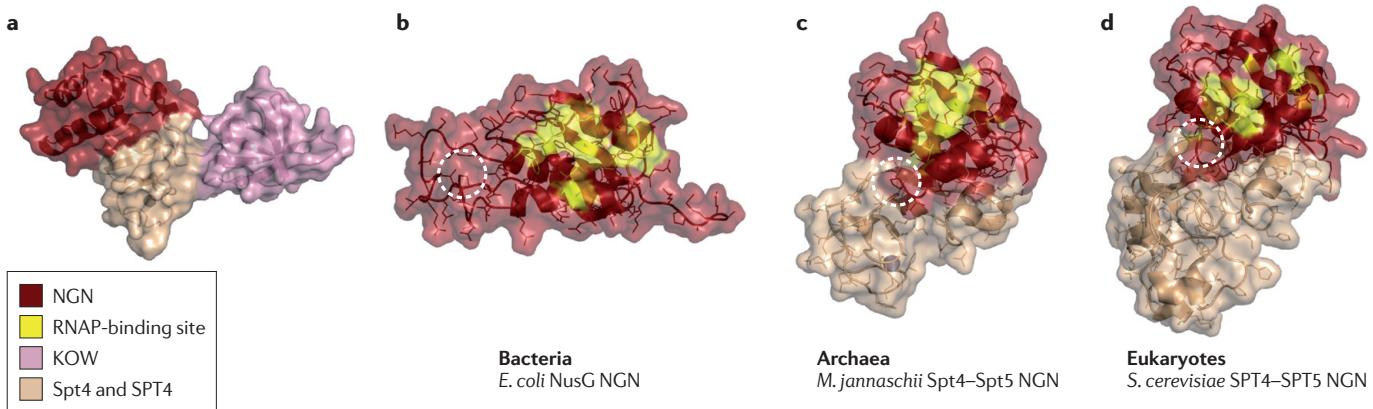


Figure 7 | Universal evolutionary conservation of the elongation factor Spt4–Spt5 and NusG. **a** | The X-ray structure of full-length *Pyrococcus furiosus* Spt4–Spt5, containing the carboxy-terminal Kypridis–Ouzounis–Woese (KOW) domain (K. S. Murakami, personal communication). **b** | The bacterial amino-terminal NusG (NGN) domain (based on the *Escherichia coli* Protein Data Bank entry 2K06). **c** | The archaeal Spt5 NGN domain bound to Spt4 (based on the *Methanocaldococcus jannaschii* Protein Data Bank entry 3LPE). Note that the Spt5 NGN domain is homologous to NusG, whereas Spt4 does not have a bacterial homologue. **d** | The eukaryotic *Saccharomyces cerevisiae* Protein Data Bank entry 2EXU). The C-terminal residue of the NGN domain that connects to the KOW domain is indicated with a dashed circle.

the parologue RfaH) could displace initiation factors (TFIIB and region 4 of σ^{70}) and thereby release RNAP into the elongation phase of transcription^{56,97,98}. Efficient binding of SPT5 to RNAPII requires contacts with the RNA transcript in the eukaryotic system^{99,100}.

Thus, non-conserved RNAP subunits (Rpo4–Rpo7) and universally conserved transcription factors (Spt5, SPT5 and NusG) modulate RNAP elongation, with both the conserved factors and non-conserved factors interacting with the RNAP clamp. These interactions could lead to subtle alterations in the RNAP clamp that have the potential to alter the catalytic properties of RNAP directly via an allosteric signal to the bridge and trigger helices^{94,101–103}. In addition, a closure of the clamp, which forms one side of the DNA-binding channel, is likely to affect the interactions of RNAP with the downstream template DNA or the DNA–RNA hybrid and thereby alter the ‘traction’ of the RNAP in the TEC^{14,28}; the interaction of Rpo4–Rpo7 and eukaryotic Spt5 with the nascent transcript could stabilize the TEC and result in increased processivity⁹⁴.

The ‘elongation-first hypothesis’

Despite the high degree of homology between all RNAPs, the basal transcription factors that are required for transcription initiation in bacteria and in archaea and eukaryotes are not evolutionarily related. By contrast, the only RNAP-associated transcription factor that is universally conserved in evolution, the Spt5–SPT5–NusG family, controls the elongation phase of transcription. What is the significance of this observation, and what does it tell us about the regulation of the ancestral form of RNAPs in the LUCA? Owing to the complete absence of any bona fide σ -factor homologues in archaea and eukaryotes, and of any TBP or TFIIB homologues in bacteria, it is unlikely that the RNAP of the LUCA initiated transcription aided by σ -like, TBP-like or TFIIB-like

transcription factors. There are four possible scenarios for the emergence of ancestral forms of the transcription initiation factors in the three domains of life (FIG. 8). In the first scenario, RNAP in the LUCA did not use any transcription initiation factors, and ancestral variants of σ -factors and of TBP and TFIIB emerged independently in the bacterial and archaeal–eukaryotic lineages, following their split. In the second scenario, RNAP in the LUCA used both σ -factors and TBP and TFIIB factors, followed by loss of TBP and TFIIB in the bacterial lineage and loss of σ -factors in the archaeal–eukaryotic lineage. In the third scenario, RNAP in the LUCA used σ -factors, which were lost and replaced by TBP and TFIIB in the archaeal–eukaryotic lineage. In the fourth scenario, RNAP in the LUCA used TBP and TFIIB proteins, which were lost and replaced by σ -factors in the bacterial lineage. Based on parsimony, we favour the first scenario, as it requires only three independent evolutionary events, versus four events in the third scenario, five events in the fourth scenario and six events in second scenario (FIG. 8). Rather than regulating transcription by recruiting RNAPs to proto-promoters such as the TATA box or the –35 and –10 elements, RNAPs could have initiated transcription non-specifically by directly associating with the template DNA, without basal transcription factors. Suitable candidates for these RNAP ‘entry sites’ are sequences that have a high A and T content, as they have a propensity to distort the DNA topology (that is, to cause DNA bending)¹⁰⁴ and they melt readily and thus allow loading of the template DNA strand into the RNAP active site. It may not be accidental, therefore, that both TATA boxes and –10 elements in contemporary promoters are rich in T and A residues. Auxiliary protein factors could have evolved independently in the bacterial and archaeal–eukaryotic lineages to enhance this process, and eventually these sequences could have co-evolved with their cognate factors, resulting in the TBP–TATA

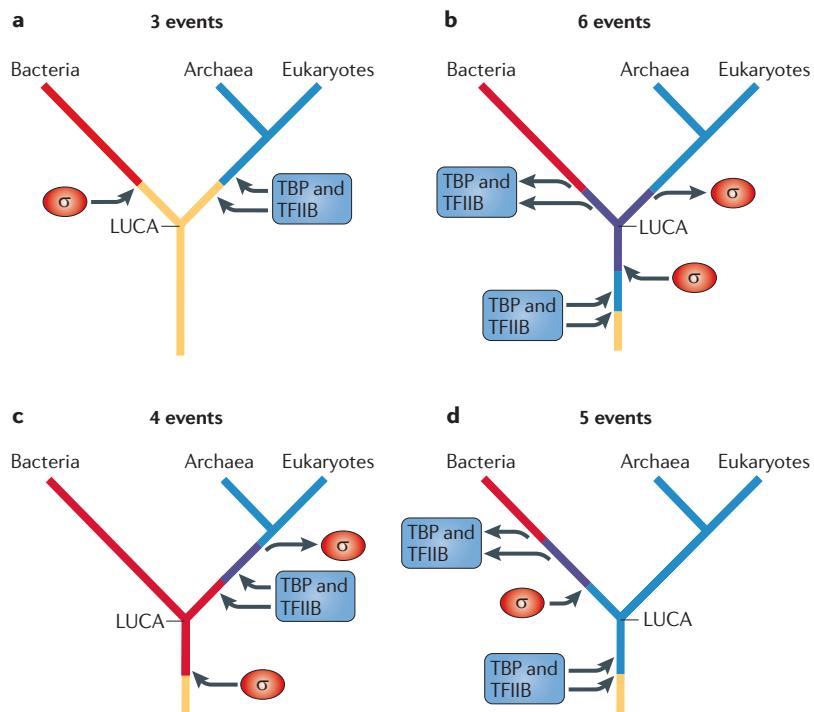


Figure 8 | Emergence of transcription initiation factors in the three domains of life. The bacterial RNA polymerase (RNAP) strictly depends on σ -factors for transcription initiation, whereas archaeal and eukaryotic RNAPs require the initiation factors TATA box-binding protein (TBP) and transcription factor IIB (TFIIB). There are four potential scenarios for the evolution of these factors. As the first scenario requires the fewest independent evolutionary events, it is the simplest explanation for the occurrence of transcription initiation factors in all extant life. **a** | As there are no σ -factor homologues in archaea and eukaryotes, and no TBP and TFIIB homologues in bacteria, it is likely that the ancestral RNAP of the last universal common ancestor (LUCA) used neither σ -factors nor TBP or TFIIB factors, and that these factors evolved independently in the bacterial and archaeal–eukaryotic lineages, respectively, after their split. **b** | An alternative scenario is that the RNAP of the LUCA used both σ -factors and TBP and TFIIB factors in parallel, and then lost the relevant factors in each lineage. **c** | A third scenario is that the LUCA used σ -factors, and then the archaeal–eukaryotic lineage lost these factors and gained TBP and TFIIB factors. **d** | The final scenario is that the LUCA used TBP and TFIIB factors, and that these were then lost in the bacterial lineage and σ -factors were gained.

box ensemble of extant archaea and eukaryotes and the σ -factor–DNA (-35 and -10 elements) ensemble of extant bacteria. Interestingly, single-subunit RNAPs, such as the T7 RNAP, can initiate transcription at a promoter in a sequence-dependent manner, without additional factors⁵.

Owing to the extensive structural and functional homology between bacterial NusG, archaeal Spt5 and eukaryotic SPT5, and the highly conserved binding sites on RNAP, it is almost certain that a NusG- and Spt5-like transcription factor associated with the RNAP in the LUCA, modulated its properties and possibly regulated gene expression. This hypothesis implies that the regulation of evolutionarily ancient RNAPs could predominantly have targeted the elongation phase of transcription instead of initiation. It is unclear to what extent the ancestral forms of NusG-like factors regulated transcription per se. However, regulation could have occurred by counteracting sequence-specific pausing, by modulating

the elongation rates, by affecting the likelihood of entering the paused state or by decreasing the pause duration. All of these phenomena have been observed with NusG in *E. coli*¹⁰⁵. Alternatively or in addition to regulating transcription by RNAP in the LUCA in a gene-specific manner, the NusG-like factor could have acted as a general processivity factor, possibly even by improving the poor utilization of RNA templates by RNAP before the emergence of DNA as the main coding molecule¹⁶. It was recently demonstrated that NusG plays a crucial part in coupling transcription and translation *in vivo* by connecting elongating RNAPs and ribosomes, and this further underpins the crucial role of NusG–Spt5–SPT5 family factors in the regulation of gene expression and their possible very early origin^{106,107}.

Conclusions and questions

RNAPs have a fundamental role in the biology of all living organisms. In recent years, the study of RNAPs has made tremendous progress, which is partially due to a number of technological breakthroughs. X-ray crystallography of RNAPs and of complexes with basal transcription factors and nucleic acid scaffolds has given us structural information at the atomic level, which has helped to formulate theories on the molecular mechanisms of transcription^{9,34,108}. Single-molecule studies using either fluorescence-based methods^{61,109,110} or optical tweezers¹⁰⁵ have enabled a functional characterization of RNAP at the single-molecule level, which is essential if we are to characterize and refine our understanding of the mechanisms. Systems biology approaches including ‘deep’ sequencing transcriptomics and ChIP-seq (chromatin immunoprecipitation followed by sequencing) whole-genome occupancy studies have reported on the composition of all transcripts synthesized by RNAPs (including their accurate 5' and 3' termini)¹¹¹ and the genomic locations of RNAP and transcription factors¹¹², respectively. The field of gene expression is currently in an exciting phase, during which it will be possible to collate all the available information — from the atomic scale to a global scale — in order to achieve a genuinely comprehensive understanding of transcription in all three domains of life. However, many important questions still remain unanswered. How is the subcellular organization of RNAP in ‘transcription factories or foci’ linked to the regulation of transcription¹¹³? How do multiple RNAPs transcribing the same template affect each other^{114,115}? How does the coupling of transcription and translation regulate gene expression in the Archaea and the Bacteria^{106,107}? How are termination and initiation linked during re-initiation of the same RNAP on the same template^{116,117}? Are the recently discovered plant complexes RNAPIV and RNAPV genuine RNA polymerases and, if so, what is the nature of their templates and which transcription factors regulate their activities¹¹⁸? Not much is known about the unorthodox transcription machineries in ‘simple’ eukaryotes such as protists — how do they work, and what can they tell us about the evolution of transcription¹¹⁹? Thus, even though parts of RNAPs are in some cases understood on an atomic level, there is still much to be learned about these important enzymes.

1. Werner, F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* **16**, 247–250, (2008).
2. Werner, F. Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* **65**, 1395–1404 (2007).
3. Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
4. Cheetham, G. M. & Steitz, T. A. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **10**, 117–123 (2000).
5. Steitz, T. A. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr. Opin. Struct. Biol.* **19**, 683–690 (2009).
6. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Ångström resolution. *Science* **292**, 1863–1876 (2001).
7. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
8. Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**, 851–854 (2008).
9. Korkhin, Y. *et al.* Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e102 (2009).
10. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
11. 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.* **38**, 585–596 (2010).
12. Batada, N. N., Westover, K. D., Bushnell, D. A., Levitt, M. & Kornberg, R. D. Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc. Natl Acad. Sci. USA* **101**, 17361–17364 (2004).
13. Iyer, L. M., Koonin, E. V. & Aravind, L. 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 (2003).
14. Grohmann, D., Werner, F. Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase. *Res. Microbiol.* 21 Sep 2010 (doi:10.1016/j.resmic.2010.09.002).
15. Wassarman, K. M. & Saecher, R. M. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**, 1601–1603 (2006).
16. Lehmann, E., Brueckner, F. & Cramer, P. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* **450**, 445–449 (2007).
17. Severinov, K. *et al.* Structural modules of the large subunits of RNA polymerase. Introducing archaeabacterial and chloroplast split sites in the β and β' subunits of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **271**, 27969–27974 (1996).
18. Zakhrova, N. *et al.* Fused and overlapping *rpoB* and *rpoC* genes in helicobacters, campylobacters, and related bacteria. *J. Bacteriol.* **181**, 3857–3859 (1999).
19. Severinov, K., Mooney, R., Darst, S. A. & Landick, R. Tethering of the large subunits of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **272**, 24137–24140 (1997).
20. Sweetser, D., Nonet, M. & Young, R. A. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl Acad. Sci. USA* **84**, 1192–1196 (1987).
21. Lane, W. J. & Darst, S. A. Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* **395**, 671–685 (2010).
22. Lane, W. J. & Darst, S. A. Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.* **395**, 686–704 (2010).
23. Ruprich-Robert, G. & Thuriaux, P. Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases. *Nucleic Acids Res.* **38**, 4559–4569 (2010).
24. Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).
25. Salgado, P. S. *et al.* The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* **4**, e434 (2006).
26. Orlicky, S. M., Tran, P. T., Sayre, M. H. & Edwards, A. M. Dissociable Rpb4-Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.* **276**, 10097–10102 (2001).
27. Grohmann, D., Hirtreiter, A. & Werner, F. RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly *in vitro*. *Biochem. J.* **421**, 339–343 (2009).
28. Grohmann, D., Werner, F. Hold on! RNAP interactions with the nascent RNA modulate transcription elongation and termination. *RNA Biol.* **7**, 310–315 (2010).
29. Naji, S., Grunberg, S. & Thomm, M. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J. Biol. Chem.* **282**, 11047–11057 (2007).
30. Ouhammouch, M., Werner, F., Weinzierl, R. O. & Geiduschek, E. P. A fully recombinant system for activator-dependent archaeal transcription. *J. Biol. Chem.* **279**, 51719–51721 (2004).
31. Werner, F. & Weinzierl, R. O. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell Biol.* **25**, 8344–8355 (2005).
32. Werner, F. & Weinzierl, R. O. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell* **10**, 635–646 (2002).
33. Werner, F., Eloranta, J. J. & Weinzierl, R. O. Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res.* **28**, 4299–4305 (2000).
34. Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
35. Reich, C. *et al.* The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable *in vivo* and *in vitro*. *Mol. Microbiol.* **71**, 989–1002 (2009).
36. Grunberg, S., Reich, C., Zeller, M. E., Bartlett, M. S. & Thomm, M. Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes. *Nucleic Acids Res.* **38**, 1950–1963 (2010).
37. Bartlett, M. S., Thomm, M. & Geiduschek, E. P. Topography of the euryarchaeal transcription initiation complex. *J. Biol. Chem.* **279**, 5894–5903 (2004).
38. Kang, X. *et al.* Structural, biochemical, and dynamic characterizations of the hRPB8 subunit of human RNA polymerases. *J. Biol. Chem.* **281**, 18216–18226 (2006).
39. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl Acad. Sci. USA* **95**, 14699–14704 (1998).
40. Briand, J. F. *et al.* Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III. *Mol. Cell Biol.* **21**, 6056–6065 (2001).
41. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaeabacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361 (2008).
42. Walmacq, C. *et al.* Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J. Biol. Chem.* **284**, 19601–19612 (2009).
43. Ziegler, L. M., Khaperskyy, D. A., Ammerman, M. L. & Ponticelli, A. S. Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF. *J. Biol. Chem.* **278**, 48950–48956 (2003).
44. Campbell, E. A., Westblade, L. F. & Darst, S. A. Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. *Curr. Opin. Microbiol.* **11**, 121–127 (2008).
45. Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
46. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
47. MacLellan, S. R., Guariglia-Oropeza, V., Gaballa, A. & Helmann, J. D. A two-subunit bacterial σ-factor activates transcription in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **106**, 21323–21328 (2009).
48. MacLellan, S. R., Wecke, T. & Helmann, J. D. A previously unidentified σ factor and two accessory proteins regulate oxalate decarboxylase expression in *Bacillus subtilis*. *Mol. Microbiol.* **69**, 954–967 (2008).
49. Wigñeshweraj, S. *et al.* Modus operandi of the bacterial RNA polymerase containing the σ⁵⁴ promoter-specificity factor. *Mol. Microbiol.* **68**, 538–546 (2008).
50. Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1422 (2000).
51. Schwartz, E. C. *et al.* A full-length group I bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem. Biol.* **15**, 1091–1103 (2008).
52. Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
53. Campbell, E. A. *et al.* Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol. Cell* **9**, 527–539 (2002).
54. Chen, H. T. & Hahn, S. Binding of TFIIB to RNA polymerase II: mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol. Cell* **12**, 437–447 (2003).
55. Chen, H. T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169–180 (2004).
56. Sevostyanova, A., Svetlov, V., Vassilyev, D. G. & Artsimovitch, I. The elongation factor RfaH and the initiation factor σ bind to the same site on the transcription elongation complex. *Proc. Natl. Acad. Sci. USA* **105**, 865–870 (2008).
57. Renfrow, M. B. *et al.* Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J. Biol. Chem.* **279**, 2825–2831 (2004).
58. Bushnell, D. A., Westover, K. D., Davis, R. E. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**, 983–988 (2004).
59. Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
60. Kulbachinskiy, A. & Mustaev, A. Region 3.2 of the σ subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* **281**, 18273–18276 (2006).
61. Kaparidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
62. Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts *in vivo*. *Science* **324**, 927–928 (2009).
63. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285–1290 (2002).
64. Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666–1677 (1996).
65. Okhuma, Y. *et al.* Structural motifs and potential σ homologies in the large subunit of human general transcription factor TFIIE. *Nature* **354**, 398–401 (1991).
66. Sumimoto, H. *et al.* Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. *Nature* **354**, 401–404 (1991).
67. Garrett, K. P. *et al.* The carboxyl terminus of RAP30 is similar in sequence to region 4 of bacterial σ factors and is required for function. *J. Biol. Chem.* **267**, 23942–23949 (1992).
68. Flores, O. *et al.* The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc. Natl. Acad. Sci. USA* **88**, 9999–10003 (1991).
69. Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274**, 35668–35675 (1999).
70. Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
71. Geiger, S. R. *et al.* RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell* **39**, 583–594 (2010).
72. Selth, L. A., Sigurdsson, S. & Svejstrup, J. Q. Transcript elongation by RNA polymerase II. *Annu. Rev. Biochem.* **79**, 271–293 (2010).

73. Vassilyev, D. G. Elongation by RNA polymerase: a race through roadblocks. *Curr. Opin. Struct. Biol.* **19**, 691–700 (2009).
74. Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* **34**, 1062–1066 (2006).
75. Brueckner, F., Ortiz, J. & Cramer, P. A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.* **19**, 294–299 (2009).
76. Brueckner, F. *et al.* Structure–function studies of the RNA polymerase II elongation complex. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 112–120 (2009).
77. Ha, K. S., Touloukhon, I., Vassilyev, D. G. & Landick, R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J. Mol. Biol.* **401**, 708–725 (2010).
78. Shibata, R. *et al.* Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem. Biophys. Res. Commun.* **355**, 122–128 (2007).
79. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of a *Mycobacterium tuberculosis* NusA–RNA complex. *EMBO J.* **24**, 3576–3587 (2005).
80. Runner, V. M., Podolny, V. & Buratowski, S. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell Biol.* **28**, 1883–1891 (2008).
81. Armache, K. J., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134 (2005).
82. Deighan, P. & Hochschild, A. Conformational toggle triggers a modulator of RNA polymerase activity. *Trends Biochem. Sci.* **31**, 424–426 (2006).
83. Sigurdsson, S., Dirac-Sveistrup, A. B. & Sveistrup, J. Q. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol. Cell* **38**, 202–210 (2010).
84. Hausner, W., Lange, U. & Musfeldt, M. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**, 12393–12399 (2000).
85. Laptenco, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334 (2003).
86. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).
87. Jeon, C., Yoon, H. & Agarwal, K. The transcription factor TFIIS zinc ribbon dipeptide Asp–Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc. Natl Acad. Sci. USA* **91**, 9106–9110 (1994).
88. Borukhov, S., Lee, J. & Laptenco, O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* **55**, 1315–1324 (2005).
89. Opalka, N. *et al.* Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**, 335–345 (2003).
90. Condon, C., Squires, C. & Squires, C. L. Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59**, 623–645 (1995).
91. Arnvig, K. B. *et al.* Evolutionary comparison of ribosomal operon antitermination function. *J. Bacteriol.* **190**, 7251–7257 (2008).
92. Greenblatt, J., Nodwell, J. R. & Mason, S. W. Transcriptional antitermination. *Nature* **364**, 401–406 (1993).
93. Dodd, I. B., Shearwin, K. E. & Egan, J. B. Revisited gene regulation by bacteriophage λ. *Curr. Opin. Genet. Dev.* **15**, 145–152 (2005).
94. Hirnreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
95. Mooney, R. A., Schweiner, K., Roessch, P., Gottesman, M. & Landick, R. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **2**, 341–358 (2009).
96. Zhou, K., Kuo, W. H., Fillingham, J. & Greenblatt, J. F. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl Acad. Sci. USA* **106**, 6956–6951 (2009).
97. Nickels, B. E., Mukhopadhyay, J., Garrity, S. J., Ebright, R. H. & Hochschild, A. The σ⁷⁰ subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nature Struct. Mol. Biol.* **11**, 544–550 (2004).
98. Artsimovitch, I. Post-initiation control by the initiation factor sigma. *Mol. Microbiol.* **68**, 1–3 (2008).
99. Misra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. *Proc. Natl Acad. Sci. USA* **107**, 11301–11306 (2010).
100. Cheng, B. & Price, D. H. Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res.* **36**, e135 (2008).
101. Svetlov, V., Belogurov, G. A., Shabrova, E., Vassilyev, D. G. & Artsimovitch, I. Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res.* **35**, 5694–5705 (2007).
102. Belogurov, G. A. *et al.* Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell* **26**, 117–129 (2007).
103. Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J. Biol.* **7**, 40 (2008).
104. Peters, J. P. 3rd & Maher, L. J. DNA curvature and flexibility *in vitro* and *in vivo*. *Q. Rev. Biophys.* **43**, 23–63 (2010).
105. Herbert, K. M. *et al.* *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399**, 17–30 (2010).
106. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
107. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010).
108. Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157–162 (2007).
109. Andrecka, J. *et al.* Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res.* **37**, 5803–5809 (2009).
110. Greive, S. J. & von Hippel, P. H. Thinking quantitatively about transcriptional regulation. *Nature Rev. Mol. Cell Biol.* **6**, 221–232 (2005).
111. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
112. Mooney, R. A. *et al.* Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell* **33**, 97–108 (2009).
113. Sutherland, H. & Bickmore, W. A. Transcription factories: gene expression in unions? *Nature Rev. Genet.* **10**, 457–466 (2009).
114. Epshteyn, V., Toume, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.* **22**, 4719–4727 (2003).
115. Saeki, H. & Sveistrup, J. Q. Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell* **35**, 191–205 (2009).
116. Werner, M., Thuriaux, P. & Soutourina, J. Structure–function analysis of RNA polymerases I and III. *Curr. Opin. Struct. Biol.* **19**, 740–745 (2009).
117. Spitalny, P. & Thomm, M. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol. Microbiol.* **67**, 958–970 (2008).
118. Landick, R. Functional divergence in the growing family of RNA polymerases. *Structure* **17**, 323–325 (2009).
119. Vanhamme, L. Trypanosome RNA polymerases and transcription factors: sensible trypanocidal drug targets? *Curr. Drug Targets* **9**, 979–996 (2008).
120. Grunberg, S., Bartlett, M. S., Naji, S. & Thomm, M. Transcription factor E is a part of transcription elongation complexes. *J. Biol. Chem.* **282**, 35482–35490 (2007).

Acknowledgements

We thank S. Gribaldo from the Institute Pasteur, Paris, France, for stimulating discussions on evolution and the inspiration for figure 8. We also thank K. S. Murakami for sharing unpublished results.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Finn Werner's homepage: <http://www.smb.ucl.ac.uk/molecular-microbiology/dr-finn-werner.html>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF