



REVIEW

A Nexus for Gene Expression—Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life

Finn Werner

RNAP Laboratory, Institute of Structural and Molecular Biology, Division of Biosciences, University College London, Gower Street, London WC1E 6BT, UK

Received 13 December 2011;
received in revised form
10 January 2012;
accepted 13 January 2012
Available online
27 January 2012

Edited by B. Connolly

Keywords:
transcription;
RNA polymerase;
Spt4/5;
NusG;
evolution

Evolutionary related multisubunit RNA polymerases (RNAPs) transcribe the genomes of all living organisms. Whereas the core subunits of RNAPs are universally conserved in all three domains of life—indicative of a common evolutionary descent—this only applies to one RNAP-associated transcription factor—Spt5, also known as NusG in bacteria. All other factors that aid RNAP during the transcription cycle are specific for the individual domain or only conserved between archaea and eukaryotes. Spt5 and its bacterial homologue NusG regulate gene expression in several ways by (i) modulating transcription processivity and promoter proximal pausing, (ii) coupling transcription and RNA processing or translation, and (iii) recruiting termination factors and thereby silencing laterally transferred DNA and protecting the genome against double-stranded DNA breaks. This review discusses recent discoveries that identify Spt5-like factors as evolutionary conserved nexus for the regulation and coordination of the machineries responsible for information processing in the cell.

© 2012 Elsevier Ltd. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

Introduction—A Conserved RNAP Regulated by Divergent Factors

Life is strictly dependent on the faithful replication of the genetic information and its expression through the synthesis of RNA and protein. This is reflected in the evolutionary conservation of the machineries that carry out these functions and in their regulation. Replication, transcription, and translation are not independent of another but are coordinated in a dynamic fashion that adjusts the needs of one with the yield of the other, and overall enables maximum energy efficiency and protection of genome integrity.

Transcription of all cellular genomes is carried out by multisubunit RNA polymerases (RNAPs), all of

which are derived from one ancestral enzyme present in the last universal common ancestor (LUCA) (Fig. 1).^{1,2} This evolutionary relationship is reflected in the homology of RNAP subunit sequence and structure, overall RNAP ternary architecture, and molecular mechanisms by which RNAPs transcribe RNA in a DNA-template-dependent manner^{3,4} (Fig. 1 and Table 1). Interestingly, the same conservation does not extend to the basal transcription factors that assist RNAP during the transcription cycle.^{2,5} For example, whereas all RNAPs face the similar mechanistic challenges during transcription initiation (such as RNAP recruitment, DNA strand separation, initiation of RNA polymerisation, and promoter escape), the bacterial and archaeal/eukaryotic RNAPs utilise nonhomologous factors to facilitate these processes (Table 1). The bacterial RNAP utilises sigma factors, whereas eukaryotic RNAPII uses the basal factors TBP (*TATA binding protein*) or TFIID [TBP plus TAFs (*TBP-associated factors*)], TFIIA (transcription factor *I*A), TFIIB, TFIIIE, TFIIF, and TFIIF.^{6,7} The

E-mail address: f.werner@ucl.ac.uk.

Abbreviations used: RNAP, RNA polymerase; DSB, double-stranded DNA break; LUCA, last universal common ancestor; TEC, transcription elongation complex; ops, operon polarity suppressor.

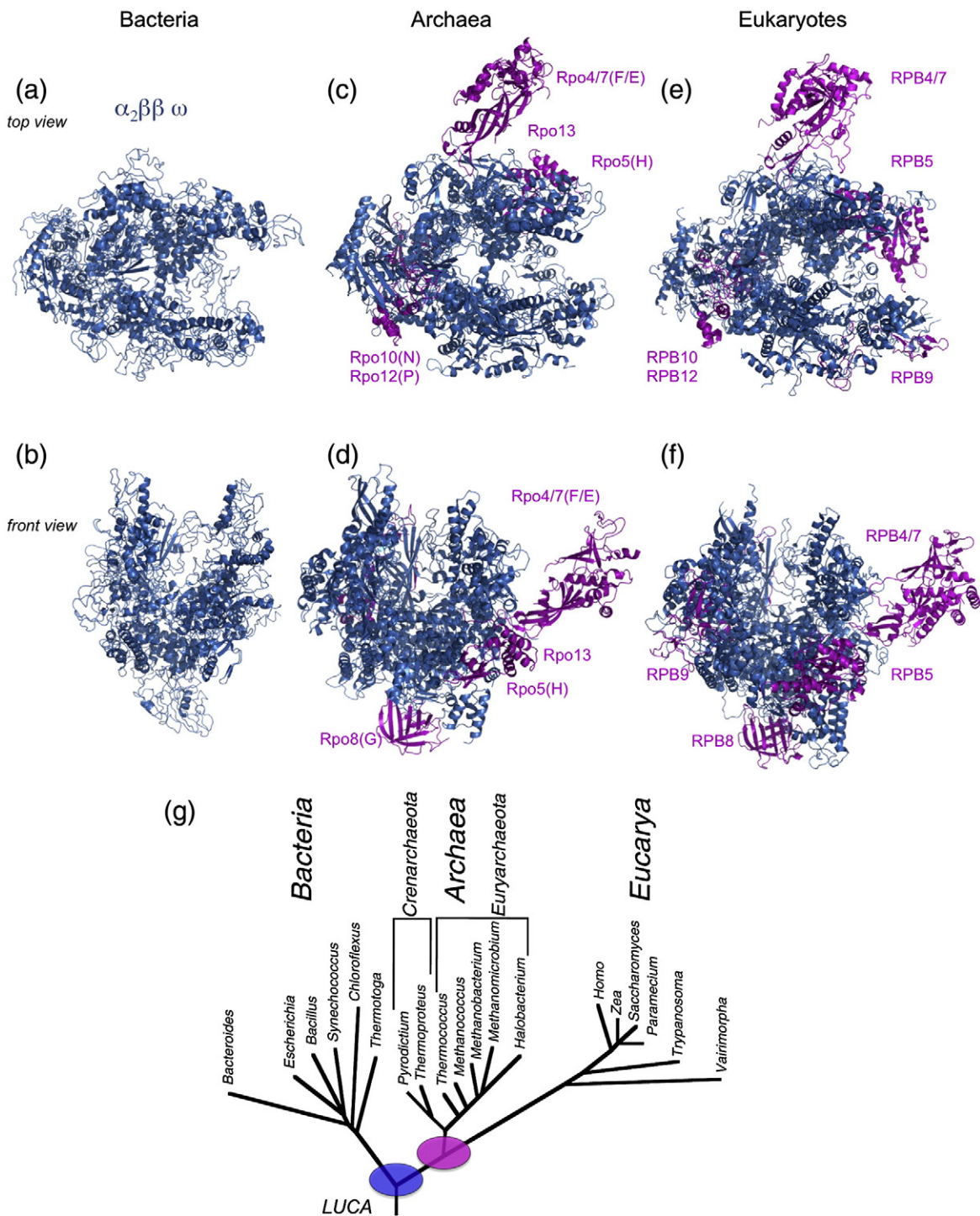


Fig. 1. Structure of multisubunit RNAPs in the three domains of life. Representative RNAP structures are shown in top views (a, c, and e) and in front views (b, d, and f): bacterial [a and b; Protein Data Bank (PDB) ID 1I6V], archaeal (c and d; PDB ID 2WAQ), and eukaryotic (e and f; PDB ID 1NT9) RNAPs. The universally conserved core subunits are shown in blue, and the subunits specific for archaeal and eukaryotic RNAPs are highlighted in magenta. (g) The universal Tree of Life; the blue circle indicates that ancestral versions of the core RNAP subunits were present in the LUCA of all life, and the magenta circle indicates that the archaeo–eukaryotic subunits were present before the split of the archaeal and eukaryotic domains of life.

archaeal RNAP is a streamlined version of the RNAPII system and utilises TBP, TFB (homologous to TFIIB), and TFE (homologous to TFIIE alpha).⁸

The fact that bacteria and archaea/eukaryotes make use of nonhomologous factors for initiation strongly suggests that the ancestral RNAP of LUCA

Table 1. Evolutionary conservation of RNAP subunits and transcription factors in the three domains of life

	Bacter	Archaea	Eukaryotes			Plants	
			RNAPII	RNAPIII	RNAPI	RNAPIV	RNAPV
RNAP subunits	β'	Rpo1 (A)	RPB1	C160	A190	NRPD1	NRPE1
	β	Rpo2 (B)	RPB2	C128	A135	NRPD/E2	NRPD/E2
	α	Rpo3 (D)	RPB3	AC40	AC40	RPB3(+1)	RPB3(+1)
	α	Rpo11 (L)	RPB11	AC19	AC19	RPB11	RPB11
	ω	Rpo6 (K)	RPB6	RPB6	RPB6	RPB6(+1)	RPB6(+1)
		Rpo5 (H)	RPB5	RPB5	RPB5	RPB5(+3)	NRPE5
		Rpo8* (G)	RPB8	RPB8	RPB8	RPB8(+1)	RPB8(+1)
		Rpo10 (N)	RPB10	RPB10	RPB10	RPB10	RPB10
		Rpo12 (P)	RPB12	RPB12	RPB12	RPB12	RPB12
		Rpo4 (F)	RPB4	C17	A14	NRPD/E4	NRPD/E4
		Rpo7 (E)	RPB7	C25	A43	NRPD7(+1)	NRPE7
			RPB9	C11	A12	NRPD9b	RPB9
	transcription factors			<i>TFIIFα</i>	C53	A49	
			<i>TFIIFβ</i>	C37	A34.5		
		<i>TFE α</i>	<i>TFIIEα</i>	C82			
		<i>TFEβ/C34*</i>	<i>TFIIEβ</i>	C34			
				C31			
		<i>TBP</i>	<i>TBP</i>	<i>TBP</i>	<i>TBP</i>		
		<i>TFB</i>	<i>TFIIB</i>	<i>Brf-1</i>	<i>TAF1B</i>		
			<i>TFIIA</i>				
			<i>TFIIH</i>				
		<i>TFS</i>	<i>TFIIS</i>		<i>TFIIS</i>		
		<i>Spt4</i>	<i>Spt4</i>		<i>Spt4</i>		
		<i>NusG</i>	<i>Spt5</i>	<i>Spt5</i>	<i>Spt5</i>		
		<i>NusA</i>	<i>NusA</i>				
	<i>rho</i>						
	<i>Sigma</i>						
	<i>Gre</i>						

Homologous polypeptides are summarised in rows; RNAP subunits and transcription factor discussed in the text are organised in columns. The universally conserved RNAP subunits are highlighted in blue, and the subunits specific for archaea and eukaryotes are highlighted in magenta. Note that some of the subunits of RNAPI and RNAPIII are paralogous to RNAPII transcription factors. The only RNAP-associated transcription factor that is universally conserved in all three domains of life is Spt5/NusG (highlighted in green). The combination of TBP and TFB-like proteins is required for transcription of the archaeal and all eukaryotic RNAPs (highlighted in yellow). Polypeptides marked with asterisks (*) are only present in some species, and the number in brackets after the RNAPIV and RNAPV subunits indicates multiple genes.

did not utilise any of these factors but rather initiated transcription either factor-independently, or aided by a factor that subsequently was lost in evolution.⁷ There is a precedence for the former mechanism since (evolutionary unrelated) single subunit RNAPs (e.g., bacteriophage T7 RNAP) can initiate transcription by directly recognising the promoter DNA in a sequence-specific manner.⁹

During elongation, multisubunit RNAPs frequently pause and move in a retrograde direction along the template by backtracking, which can regulate transcription and contribute to its fidelity, but also be detrimental to productive transcription.^{10–12} Backtracked complexes are rescued by transcript cleavage factors that associate with RNAP and retune its active site, which results in a cleavage of the transcript and a new RNA 3' terminus competent for catalysis and elongation.^{11,13–15} Similar to the initiation phase of transcription, the challenges that RNAPs need to overcome during elongation are identical, but the transcript cleavage factors that solve the problem are not homologous.¹⁶ In bacteria, Gre factors stimulate transcript cleavage, and in archaea/eukaryotes, TFIIS/TFS fulfil this function. Gre and TFIIS/TFS are not homologous and adopt different structures; however, they interact with RNAP in the same manner by inserting two juxtaposed acidic residues into the active centre through the NTP entry pore.^{17,18} The lack of *bona fide* homology suggests—despite the compelling similarity of the mechanism—that the LUCA RNAP did not utilise any of the factors.

During termination of transcription, the stable elongation complex has to dissociate in order to release the transcript and template and make RNAP available for the next round of transcription. The elongation complex has to undergo substantial conformational changes during termination (such as an opening of the RNAP clamp) that are not necessarily energetically favourable. In bacteria, this energy is provided by RNA secondary structure formation by Watson–Crick base pairing of intrinsic terminators or by the rho helicase by virtue of ATP hydrolysis.^{19,20} Termination by eukaryotic RNAPII is more complex and involves both polyadenylation of the transcript and an exonuclease (Xrn2/Rat1) that, despite being unrelated to rho, translocates along the RNA towards the RNAP and terminates transcription reminiscent of rho.^{21,22} Eukaryotic RNAPIII and archaeal RNAP are able to terminate transcription independently of RNA secondary structures or factors.^{23–26} Thus, similar to the transcription initiation and elongation cleavage factors, none of the termination factors are evolutionary conserved across the three domains of life.

In summary, it appears that a highly conserved RNAP system is almost exclusively aided by evolutionary nonrelated transcription factors. One important exception to this observation is Spt5/

NusG, and the key functions that this factor have in gene expression might explain its universal evolutionary conservation.²⁷

Structure and Organisation of Spt5-Like Factors and Their Complexes with RNAP

Eukaryotic and archaeal Spt5 and their bacterial homologue NusG display an extensive sequence and structural homology and associate with their cognate RNAPs via evolutionary conserved binding sites.²⁷ The organisation of Spt5-like proteins is modular, and all factors encompass an NGN (NusG N-terminal) domain and one (bacteria and archaea) or multiple (eukaryotes) KOW domain(s)^{27–29} (Fig. 2a). The NGN domain interacts with RNAP and stimulates its processivity,^{27,28,31} whereas the KOW domains serve as a recruitment platform for accessory factors (Table 2).^{28,32} NusG from bacteria is a monomeric transcription factor. Spt5 from eukaryotes and archaea forms a heterodimeric complex with Spt4, a small protein that stabilises the Spt5 NGN domain structure but, otherwise, has an unknown function.²⁷ The evolutionary conservation of Spt5 and NusG is apparent from their amino acid sequence but becomes even more prominent when comparing the structures of the NGN domains (Fig. 2).^{27,28,33} Mutational analysis of the Spt5 and NusG NGN domains have identified a conserved hydrophobic depression that functionally interacts with the RNAP clamp coiled coil (“CC”, also referred to as clamp helices “CH” in bacteria) on one side of the DNA binding channel.^{27,28,31} The X-ray structure of a complex between a recombinant fragment of the *Pyrococcus furiosus* RNAP clamp with Spt4/5 confirms this binding site.³⁴ As shown for the bacterial NusG paralogue RfaH, NGN domains also make contact with the RNAP beta gate loop (“betaGL”), which is located opposite to the clamp CC across the DNA binding channel³⁵ (Fig. 3). Via these two interactions, the NGN domains of Spt5-like factors (i) lock the DNA–RNA hybrid into the active site and (ii) separate the upstream and downstream portions of the template (Fig. 4). A cryo-electron microscopy structure of the entire archaeal RNAP–Spt4/5 complex confirms that the NGN domain of Spt5 and, by inference, its homologues NusG and RfaH span across the DNA binding channel of RNAP.³⁸

Molecular Basis of the Processivity Function of Spt5-Like Factors

Spt5-like factors bind to the RNAP clamp, a structurally well conserved flexible module that

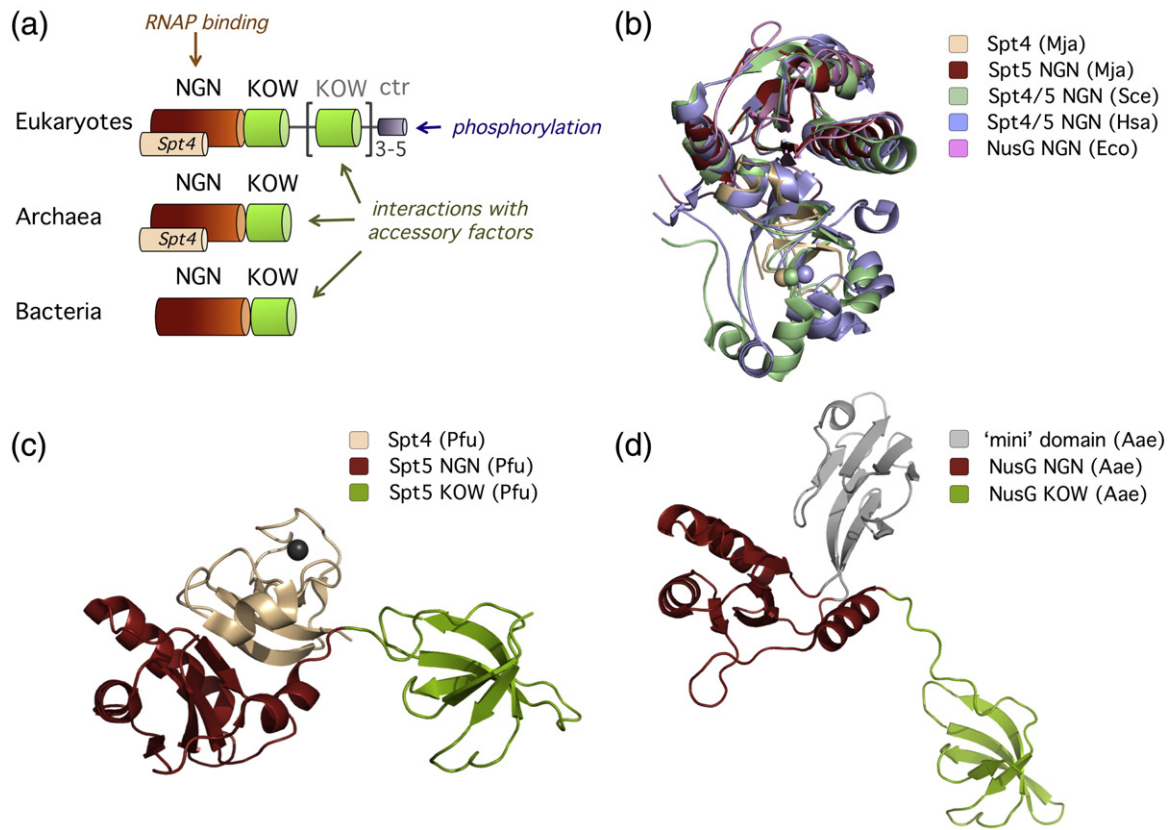


Fig. 2. Structure and domain organisation of Spt5-like factors. Organisation of Spt4/5 and NusG (a). Spt5 consists of an NGN domain (highlighted in firebrick red) and one or more KOW (Kyrpidis, Ouzounis, Woese) domains (green), only eukaryotic Spt5 contains two C-terminal repeats (ctr). Eukaryotic and archaeal Spt5 form a complex with Spt4 (wheat); the zinc ion coordinated by Spt4 is illustrated as a sphere. (b) A structural alignment of Spt4/5 NGN from archaea (*Methanocaldococcus jannaschii*, Mja) and eukaryotes (*Homo sapiens*, Hsa; *S. cerevisiae*, Sce) and the NusG NGN domain from bacteria (*E. coli*, Eco) prepared using VMD (<http://www.ks.uiuc.edu/Research/vmd/>). The structure of Mja Spt4/5NGN was solved using a dimeric complex of Spt4 (wheat) and Spt5 NGN (firebrick red), and the Sce (mint green) and Hsa (light blue) Spt4/5 NGN structures were solved by crystallising a fusion protein of Spt4 and Spt5. (c) The X-ray structure of Spt4/5 from *P. furiosus* (Pfu; PDB ID 3P8B) and (d) the X-ray structure of NusG from *Aquifex aeolicus* (Aae; PDB ID 1NPR) that contains a “mini”-domain (coloured light grey) inserted into the NGN domain at a position similar to Spt4.³⁰ The mini-domain is not present in all bacterial NusG variants.

closes over the DNA binding channel by a rotation of up to 30° relative to the RNAP core. The position and possibly movement of the clamp can be altered by the archaeal/eukaryote-specific RNAP stalk (consisting of the Rpo4/7 and RPB4/7 subunits; Fig. 1c–f), which has been hypothesised to play a role during the loading of the template into the active site (open complex formation) and which is important to maintain high transcription processivity.³⁹ The translocation/elongation mechanism of RNAPs is facilitated by the bridge helix and trigger loop in the active site.⁴⁰ Since both elements are connected to the inside of the clamp, even a slight reposition of the latter can have substantial impact on the elongation properties of the transcription elongation complex (TEC);⁴¹ that is, Spt5-like factors could alter the catalytic properties of RNAP via an allosteric mechanism.⁴²

By binding across the DNA binding channel of RNAP, Spt4/5 (and NusG and RfaH) has the potential to affect RNAP in two opposite ways (Fig. 4). When Spt4/5 associates with elongating RNAPs, it could deny the dissociation of RNAP-bound DNA and could thereby increase the stability of the elongation complex and stimulate transcription processivity.^{38,43} When Spt4/5 associates with free RNAP, it could deny association of promoter DNA with the DNA binding channel of RNAP, prevent RNAP recruitment, and thereby repress transcription (Fig. 4). The stimulatory property for transcription processivity of RNAP by NusG in bacteria and Spt4/5 in archaea is well documented,^{27,28} and a recent article also supports the second, inhibitory, mechanism. Preincubation of Spt4/5 with archaeal RNAP efficiently inhibits RNAP recruitment and transcription initiation in a minimal *in vitro* transcription system.⁴⁴

Table 2. The KOW domains of Spt5 and NusG interact with a plethora of factors

Domain	Factor	Sequence	Interactor	Function
Bacteria	NusG	Nonspecific	S10/ribosome rho	Coupling of transcription and translation termination; silencing of foreign DNA
	RfaH	ops	S10/ribosome	
Archaea	Spt5	Nonspecific	S10/ribosome	Coupling of transcription and translation
Eukaryotes	Spt5	Nonspecific	NELF	
			P-TEFb, Bur-1/Bur-2	Phosphorylation of RNAP, NELF, and Spt4/5
			FACT, Spt6	
			capping enzyme	Chromatin remodelling
			Cap methyl transferase	
				RNA processing

In prokaryotes, the KOW domain of NusG (and Spt5) interacts with ribosomal protein S10 and thereby physically connects RNAPs and ribosomes. It also interacts with the termination factor rho and functionally acts as its cofactor. In eukaryotes, the multiple KOW domains of Spt5 facilitate interactions with RNAPI and RNAPII, with the transcript and with a wide range of accessory factors involved in the regulation of transcription and transcript maturation.

The first swap: initiation factor TFE and elongation factor Spt4/5

The archaeal RNAP requires only two basal transcription factors for promoter-directed transcription, TBP and TFB.⁸ TBP recognises the TATA element of the promoter and bends the DNA by approximately 90°;⁴⁵ this DNA–TBP complex is recognised by TFB, which in turn recruits RNAP and transcription initiation commences. In contrast, the Spt4/5–RNAP complex cannot be recruited to the DNA–TFB–TFB complex, and transcription initiation is repressed.⁴⁴

How is this repression overcome in order to assure efficient transcription in the presence of Spt4/5 in the cell?

The answer to this question is the transcription initiation factor TFE. TFE is homologous to TFIIE alpha in the eukaryotic RNAPII system⁴⁶ and of C82 in RNAPI.⁴⁷ TFE associates with RNAP and stabilises the open transcription initiation complex^{48,49} by two mechanisms involving an allosteric effect on the RNAP clamp and by interacting with the nontemplate strand of the promoter.^{44,50} In addition, TFE can be cross-linked downstream of promoter sequences, which suggests that it can be retained on RNAP elongation complexes after promoter escape.⁵⁰ However, the binding site of TFE or the molecular mechanisms underlying its activities have remained obscure until recently. Grohmann *et al.* used a Förster resonance (fluorescence resonance energy transfer) system to identify the TFE binding site on RNAP in the complete archaeal transcription initiation complex.⁴⁴ Both the TFE winged helix domain and the Spt5 NGN domain interact with the RNAP clamp coiled coil.^{27,44} The two binding sites are overlapping, and correspondingly, Spt4/5 and TFE compete for RNAP binding. Importantly, the relative binding affinities of the factors are context dependent: TFE prevails over Spt4/5 in the initiation complex, whereas Spt4/5 prevails over TFE in the elongation complex.⁴⁴ As a result, TFE can efficiently

prevent the inhibitory effect of Spt4/5 on transcription initiation.⁴⁴

What is the function of the transcription factor swapping during transcription?

All RNAPs bind DNA in a nonsequence-specific manner, which can critically quench or reduce the concentration of “free” RNAP available for promoter-directed transcription. Association of RNAP with Spt4/5 has the potential to reduce this effect, and since the repression by Spt4/5 is negated by TFE at the promoter, the overall outcome is an increase in promoter specificity of RNAP. In addition, or alternatively, the transition between the initiation and elongation phases of transcription—referred to as promoter escape—may be aided by the factor swap. One of the intrinsic conflicts during transcription initiation is the apparent need for high-affinity interactions between RNAP and promoter DNA to ensure efficient RNAP recruitment and the ability to readily release the promoter-bound RNAP necessary to enable a high firing rate of the promoter.² This is made possible by the binding mode of initiation factors to RNAP where multiple contacts are made with a combined high affinity.⁵¹ Conformational changes in the initiation complex during promoter escape disrupt the individual interactions in a stepwise manner. The distinct conformational states of RNAP in the initiation and elongation complexes may be stabilised by associated transcription factors, and the energy barriers for the changes between the conformational states can be lowered by association with transcription factors. Thus, Spt4/5 displacing TFE could stimulate promoter escape (Fig. 5). Alternatively, Spt4/5 could ensure that RNAP-bound TFE during early elongation is efficiently displaced and replaced by Spt4/5, resulting in a high-processivity RNAP–Spt4/5 TEC. Whole genome occupancy profiling of RNAPII and transcription initiation and elongation factors in *Saccharomyces cerevisiae* has revealed a

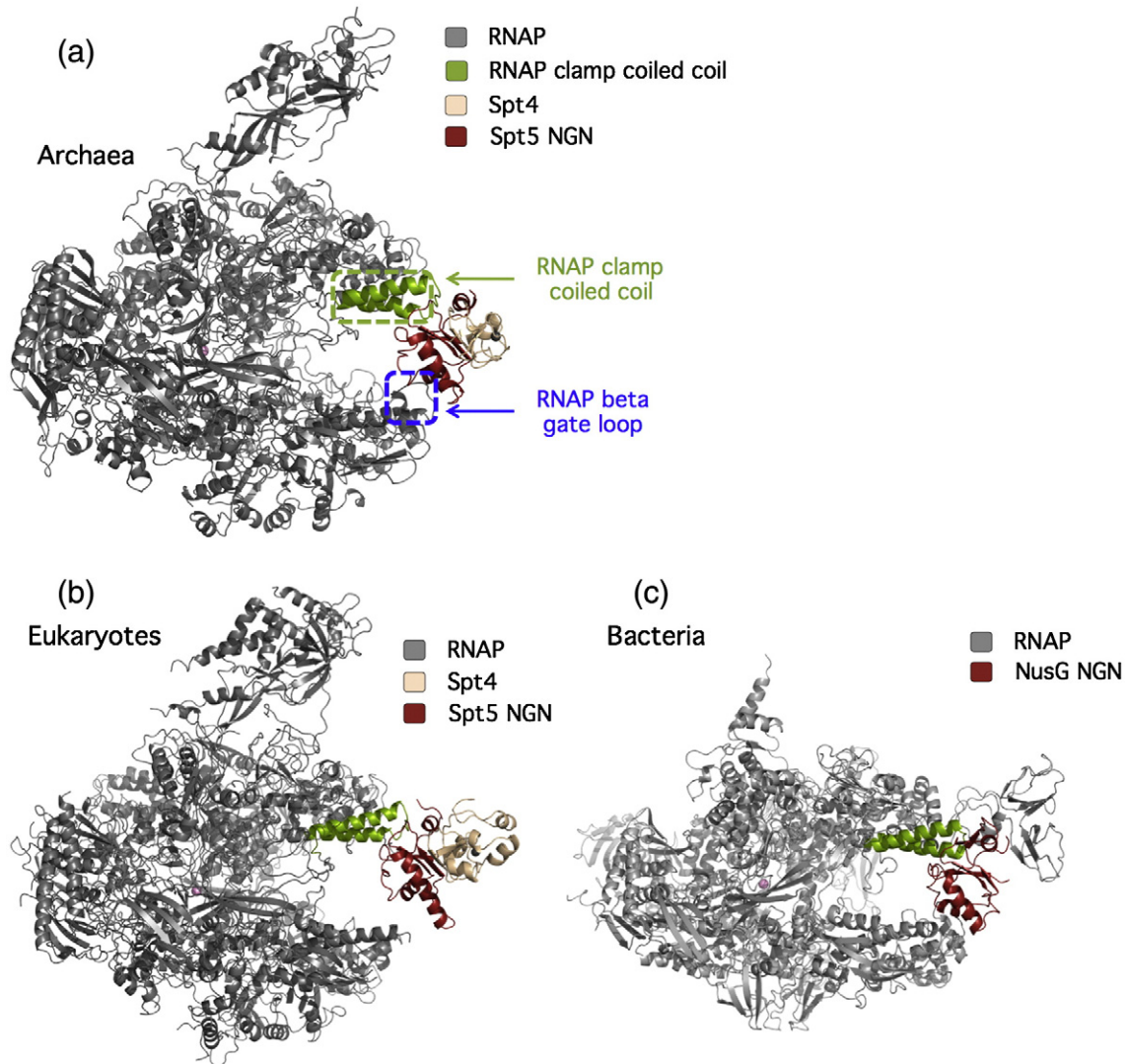


Fig. 3. Structure of RNAP–Spt4/5 complexes in the three domains of life. (a and b) Models of RNAP–Spt4/5 complexes in archaea and eukaryotes, respectively; (c) a model of the RNAP–NusG complex in bacteria. The Spt5 NGN domain is located across the DNA binding channel by interacting with the RNAP clamp coiled coil (also known as clamp helices) on one side of the cleft and with the beta gate loop on the opposite side (highlighted with blue broken rectangles). The X-ray structure of a complex encompassing the archaeal RNAP clamp and Spt4/5 NGN from *P. furiosus* was fitted into the structures of RNAPs (modified from Ref. 34). The Spt5 NGN domain is highlighted in firebrick red; Spt4, in wheat. In the model of the eukaryotic complex, the Spt5 NGN domain does not appear to make contacts with the wall of the DNA binding channel opposite to the clamp because the beta gate loop in the parental RNAPII structure is disordered.

relatively sharp transition between initiation and elongation factors (including Spt4/5) at approximately 150 bp downstream of the transcription start site.⁵² Whereas this position is too far downstream to qualify as promoter escape, it supports the notion of an efficient swap between initiation and elongation factors. Interestingly, the bacterial transcription initiation factor sigma, despite not being evolutionary related to TFE, also interacts with the RNAP CH.^{53,54} Sigma can, like TFE, remain associated with RNAP during early elongation and induce pausing.⁵⁵ Importantly,

NusG and RfaH, like Spt5, interact with the RNAP clamp CH, and both can compete with sigma for RNAP binding⁵⁶ (Fig. 5).

In summary, the mechanism of competition binding between initiation and elongation factors for RNAP is universally conserved in evolution.

The second swap: elongation and termination (co-) factor NusG and RfaH

In the bacterial system, NusG is an essential general transcription factor that has pleiotropic

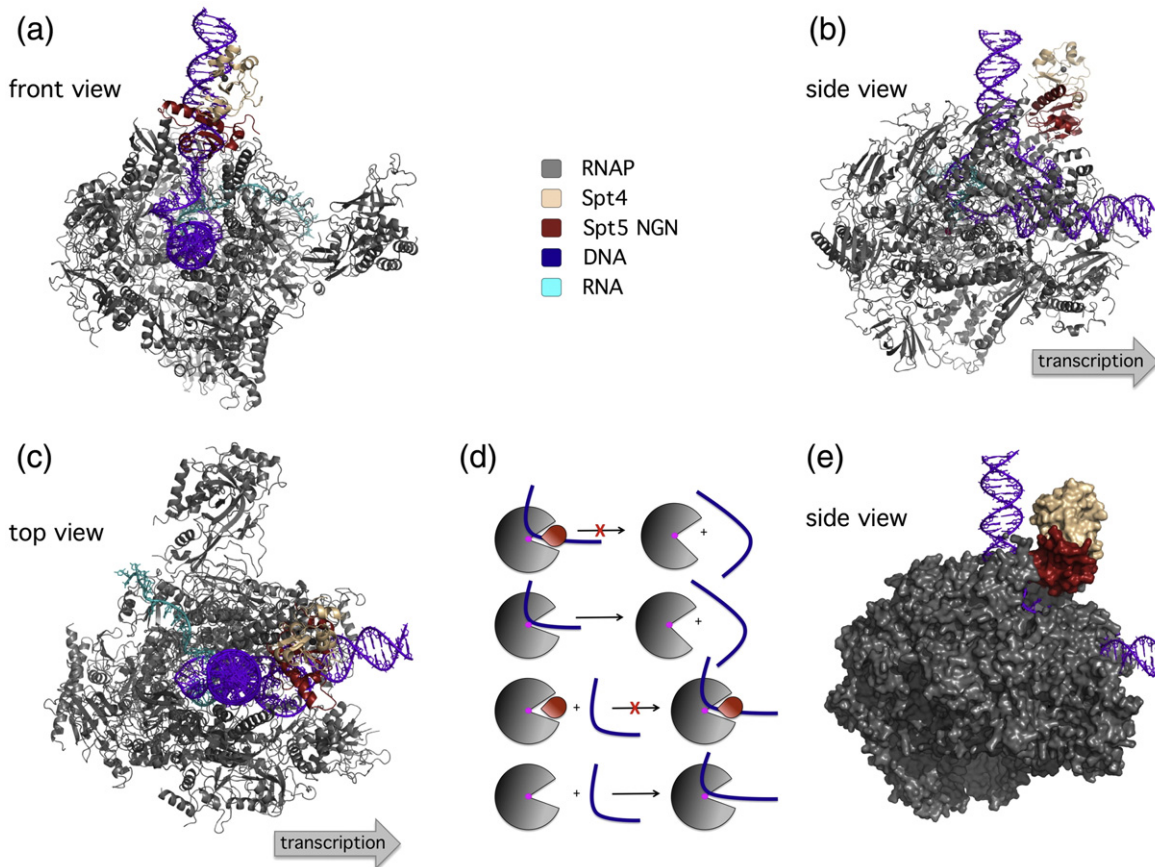


Fig. 4. Spt4/5 locks the template in the DNA binding channel of RNAP in the TEC. Front (a), side (b and e), and top (c) views of the eukaryotic RNAPII-Spt4/5 TEC. This model was built by combining structural information of the archaeal RNAP clamp-Spt4/5 complex with the X-ray structure and single-molecule fluorescence resonance energy transfer data from the yeast RNAPII-DNA-RNA elongation complex^{36,37} (modified from Ref. 34). (d) The Spt4/5 complex (red wedge) locks the DNA-RNA hybrid and the DNA strands forming the transcription bubble into the active site, it thereby denies both dissociation and association of the DNA template (blue line) from the RNAP. As a result, when Spt4/5 associates with the RNAP-DNA-RNA elongation complex, it increases its stability; by contrast, when Spt4/5 associates with free RNAP, it suppresses nonspecific DNA binding. (e) A surface representation of the complex that emphasises how the template is locked into the RNAP [same orientation as in (b)].

effects on transcription elongation.^{28,57–59} NusG is a component of the antitermination complex that converts RNAP into a termination-resistant form, which is able to read through early intrinsic termination signals in, for example, ribosomal RNA (*rrn*) operons and during late gene expression of the bacteriophage lambda.⁶⁰ NusG stimulates processivity by suppressing transcription pausing (also referred to as antipausing), but it also acts in concert with rho to facilitate termination.^{10,59,61,62} rho itself interacts with RNAP and is recruited to the TEC via physical interactions with the NusG KOW domain (Figs. 5 and 7). Thus, NusG association with RNAP enhances elongation but also directly contributes to termination. The whole genome occupancy profile of NusG demonstrates that it is bound to TECs on nearly all genes, protein encoding and noncoding transcription units.⁶³ NusG appears to

be recruited to genes in a stochastic fashion downstream of the promoter, and it is overrepresented on longer operons. Whether the latter is due to the fact that TEC on long genes have more opportunities to recruit NusG or whether NusG association with RNAP enables the TEC to transcribe longer genes is unclear.⁶³ RfaH is a nonessential sequence-specific paralogue of NusG that induces the expression of genes downstream of the operon polarity suppressor (*ops*) DNA element.⁶⁴ Transcription of the *ops* sequence triggers recruitment of RfaH to TEC and directs the transcription of a small subset of operons involved in bacterial virulence.⁶⁵ On the sequence level, RfaH and NusG are closely related, and this homology extends to the structural level with respect to their NGN domains.⁶⁶ However, whereas the NusG C-terminus is a typical KOW domain entirely consisting of

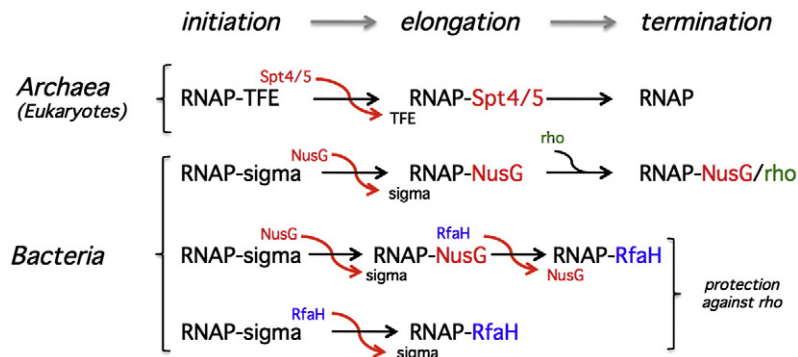


Fig. 5. Transcription factor swapping during the transcription cycle. In the archaeal and eukaryotic systems, the binding sites for the initiation factor TFE (TFII α) and the elongation factor Spt4/5 overlap on the RNAP clamp in mutual exclusive manner. During promoter escape or early elongation, Spt4/5 displaces TFE from the RNAP. In perfect analogy, the bacterial sigma initiation factor can be displaced *in vitro* by the Spt5

homologue and paralogue NusG and RfaH, respectively. NusG is able to recruit the rho factor, which results in transcription termination. In contrast, RfaH could also displace NusG *in vivo*, which protects the elongation complex against the recruitment of rho and termination and thereby facilitates the expression of distal genes.

beta sheets, the RfaH C-terminal domain in full-length RfaH is a radically different all-alpha helical fold.⁶⁶ RfaH, despite being present in the cell at much lower concentration than NusG, excludes NusG from binding RNAP once the TEC has transcribed through the ops element.⁶⁷ Whole genome occupancy profiles demonstrate that RNAP and NusG peaks overlap just downstream of the promoter on the *rfb* operon—indicative of NusG recruitment immediately after transcription initiation where the swap of sigma and NusG has occurred.³⁵ A second swap is executed with a sharp transition at approximately 1 kb when NusG is replaced by RfaH. This reduces transcription termination facilitated by rho by at least two independent mechanisms, a stimulation of transcription processivity by converting RNAP into a high-processivity (pause-resistant state) TEC similar to NusG and by preventing the recruitment of rho via the NusG KOW domain.³⁵ The latter is demonstrated by the occupancy profile of rho on the *rfb* operon. The net outcome of the second swap (between NusG and RfaH) in the very early stages of transcription elongation is the efficient transcription of genes distal to the ops sequence.

In summary, both NusG and RfaH interact with the RNAP using a conserved mechanism, but the two factors have opposing regulatory roles. NusG in concert with rho represses the expression of “foreign” genes (see below), while RfaH inhibits rho action on RNAP and thereby enhances the expression of “foreign” genes such as the *rfb* operon.^{35,67}

A Paradigm Shift—Eukaryotic Spt4/5 and Promoter Proximal Pausing

In eukaryotes, Spt4/5 is associated with RNAPII on most, if not all, genes;⁵² Spt5 is an essential gene in *S. cerevisiae*, whereas Spt4 is dispensable at

permissive temperatures.⁶⁸ This is in good agreement with results from the archaeal system, which demonstrate that Spt4 has a stabilising effect on the Spt5 NGN domain but is not strictly required for its binding to RNAP nor its stimulatory effect on processivity.²⁷ Eukaryotic Spt5 variants contain four to six copies of KOW domains (in comparison to one in archaea and bacteria) and two C-terminal repeat regions that are subject to phosphorylation but not required for cell viability in yeast.^{69,70} The additional KOW domains serve as binding or recruitment sites for a plethora of factors including NELF (negative elongation factor) and kinases such as P-TEFb (positive transcription elongation factor *b*) and Bur-1/Bur-2, factors that are involved in chromatin remodelling such as Spt6 and FACT, and RNA processing factors such as the mRNA capping enzyme and the cap methyl transferase³² (Table 2).

The classical paradigm of transcription control—regulation is chiefly controlled by the recruitment of the RNAPII to the promoter—was established using the yeast transcription machinery. Early observations of the regulation of the *hsp70* heat shock gene in a metazoan system, *Drosophila melanogaster*, suggested that its transcription was *not* regulated at the level of RNAPII recruitment. Rather, TECs were paused (also referred to as stalled or poised) proximal to the promoter approximately 40 bp downstream of the transcription start site.^{71,72} The complexes were catalytically competent but not able to penetrate into the downstream gene. Signalling events induced by heat shock lead to the activation of the kinase P-TEFb that phosphorylates multiple components of the elongation complex including Spt4/5, NELF, and RNAPII⁷³ (Fig. 6). This results in the release of the poised elongation complexes and robust expression of the heat shock genes. The regulatory advantage of these mechanisms is a fast induction achieved by uncoupling of RNAP recruitment and

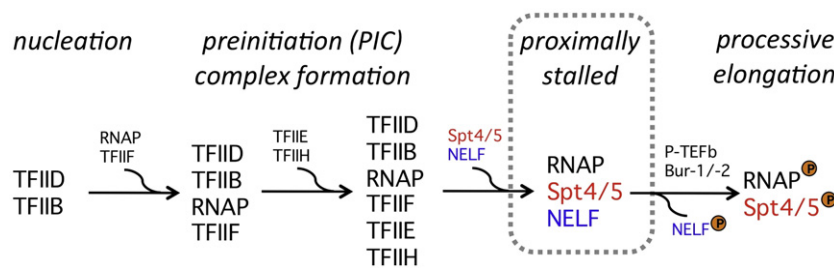


Fig. 6. Spt4/5 facilitates promoter proximal pausing in metazoans. Binding of TFIID to the TATA element of a eukaryotic class II promoter triggers a recruitment cascade that results in the initiation of transcription. However, the association of Spt4/5 and NELF with RNAPII in the early elongation complex leads to promoter proximally stalled elongation complexes. Activation of kinases including P-TEFb and Bur-1/Bur-2 leads to the phosphorylation of the RNAPII C-terminal domain, Spt4/5, and NELF, which releases the stalled complexes and induces robust gene expression.⁷⁴ Many class II genes in metazoans harbour promoter proximally stalled initiation complexes, and it is possible that this mechanism operates on a global level.⁷⁵ Phosphorylation events are highlighted with an orange “P”.

transcription.⁷⁶ In the last 5 years, it has become apparent that many, if not most, genes in metazoans harbour promoter proximally stalled RNA-P-II and are therefore likely to be subject to this regulation.^{74,75} Promoter proximal stalling could have other functions than regulation such as mRNA quality control and coupling between transcription and mRNA processing.⁷⁷ A delay of elongation complexes could thus allow the recruitment of RNA modification enzymes such as the capping enzyme, and elongation would commence only if the capping is completed. Both Spt4/5 and NELF are required for promoter proximal pausing, but the molecular mechanisms are still not understood. Spt4/5 is a key candidate for directly inhibiting RNAP activity because its binding to RNAP alters the DNA-binding properties of the enzyme and possibly induces conformational alterations in the active centre (see above) that, in theory, could pause transcription elongation.⁴⁴ The significance of the position of the paused complexes at +40 is unclear, even though it seems compelling that human RNAPII via the RPB4/7 complex is interacting with approximately 40 nucleotides of the transcript⁷⁸ and that the recruitment of eukaryotic Spt4/5 to RNAPII is enhanced, if not dependent on, interactions between the Spt5 KOW domains and the RNA transcript.⁷⁹

In summary, the Spt4/5 complex is associated with many, if not all, class II genes in eukaryotes, and it plays a pivotal role for a eukaryote-specific mechanism that regulates the processivity of transcription—promoter proximal stalling. In addition, like its prokaryotic counterpart, it is involved in the recruitment of accessory factors that regulate gene expression, alter the availability of the chromatin template, and are involved in the processing of the nascent transcript. The function of Spt4/5 in eukaryotes is not restricted to class II genes since Spt4/5 also directly associates with RNAPI and

modulates its processivity both positively and negatively.^{80,81}

Prokaryotes—NusG-Like Factors Connect RNAP, Ribosomes, and rho

In both prokaryotic domains, bacteria and archaea, transcription and translation are coupled.⁸² In bacteria, the rates of transcription and translation are correlated over a broad range of growth rates, which keeps the ratio between the two elongation rates at 3 (three nucleotides for every one amino acid).⁸³ As a result, the transcription yield is correlating with the translation needs on different genes and during varying growth rates. In effect, the elongation rate of transcription is influenced by the presence of ribosomes and their elongation rate of translation; the average elongation rate of RNAPs can be increased by cotranslating ribosomes, possibly because the latter reduce pausing and backtracking of RNAP. Ribosomes whose elongation rates have been decreased by antibiotic action in turn cause a decrease in transcription elongation rates of RNAP by virtue of their intrinsic propensity to pause and backtrack. In both situations, little mRNA is exposed and available for recognition by rho; that is, the gene is protected against rho-dependent termination.⁸³ These phenomena could be explained by the motions of two molecular motors, one of which uses the product of the other as template (mRNA). Recently, it has emerged that the NusG KOW domain physically interacts with ribosomal protein S10 (first characterised as anti-termination factor NusE) and that this interaction could tether elongating ribosomes and RNAPs⁸⁴ (Fig. 7a). In principle, this interaction could also enhance ribosome recruitment during translation initiation. Since S10 competes with rho for the binding to the NusG KOW domain, the coupling

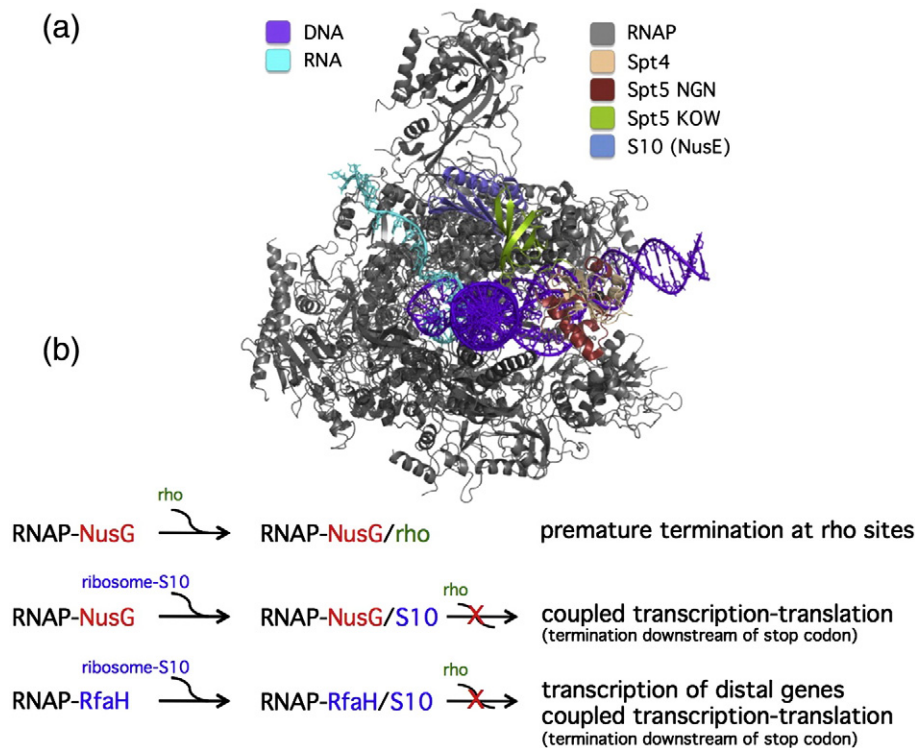


Fig. 7. Transcription and translation are coupled via NusG in prokaryotes. (a) Structural model of the RNAP-Spt4/5-S10 elongation complex. The S10 protein (light blue) of the 30S ribosomal subunit forms a complex with the KOW domain (green) of Spt5, and thus, the ribosome is ideally positioned to interact with the mRNA template (cyan). This model was built by combining the X-ray structure of the NusG KOW-S10 complex from *E. coli* (PDB ID 2KVO), the structure of the archaeal Spt4/5 complex from *P. furiosus* (PDB ID 3P8B), and a model of the RNAPII-Spt4/5 NGN complex from *S. cerevisiae*.³⁴ The components are colour coded according to the key in the figure. The RNA interacts with the RNAP stalk (Rpo4/7 or RPB4/7) during transcription elongation, but the RNA species included in the X-ray structure (PDB ID 1NT9) was too short to observe this interaction. (b) Mechanisms of coupled transcription-translation. The NusG KOW domain interacts with the rho factor and ribosomal protein S10 in a mutual exclusive manner. RNAP-NusG elongation complexes are able to recruit rho, which can lead to (pre-) mature termination at rho termination sites. Ribosomal protein S10 can bind to the NusG KOW domain and thereby prevent the recruitment of rho. Thus, the efficient coupling of transcription and translation (i.e., of RNAP and ribosome) prevents premature termination. Following translation termination, the ribosome dissociates from the transcript and makes the NusG KOW domain accessible for rho binding, which promptly leads to transcription termination. Likewise, the NusG paralogue RfaH can recruit ribosomes, protect against rho, and facilitate the expression of distal genes.

of transcription and translation protects the TEC against rho-dependent termination⁸⁴ (Fig. 7). Translation is terminated at the 3' end of operons, and S10 dissociates from RNAP-bound NusG, which makes the KOW domain available for the recruitment of rho and subsequent efficient transcription termination.⁸⁴ In summary, Spt5-like factors have the potential to coordinate transcription and translation in prokaryotes, bacteria and archaea.

The Silencing of Foreign DNA and Maintenance of Genome Stability

Bacterial genomes contain a considerable load of horizontally transferred or "foreign" DNA includ-

ing prophages, insertion sequence elements, and K-island clusters.⁸⁵ Many of these elements can be deleted to obtain reduced genome strains, which sometimes have beneficial biotechnological properties and different requirements for essential genes than their parental strains.⁸⁶ In the synthetic *Escherichia coli* strain MDS42, 14% of the genome of the parental strain MG1655 has been deleted. In contrast to MG1655, NusG (and NusA) is not essential for cell viability in MDS42.⁸⁷ Moreover, deletion of the lambdoid prophage rac from MG1655 is sufficient to make NusG dispensable.⁸⁷ Considering that MDS42 has a dramatically increased ($\sim 10^4\times$) resistance to the highly selective inhibitor of the rho termination factor (bicyclomycin), it is possible that the function of NusG that makes it indispensable in the wild-type strain is due

to its role as a rho cofactor for rho.⁸³ In other words, NusG in conjunction with rho silences *rac* gene expression, which otherwise would kill the cell. A whole genome expression analysis reveals that bicyclomycin treatment predominantly upregulates the expression of foreign DNA, which suggest that rho represses its expression.⁸⁷ Since rho to a large extent relies on NusG as cofactor, NusG may play an important role for the silencing of foreign DNA on a global level.

Transcription and replication occur simultaneously, and collisions between the TEC and the replication fork not only are inevitable but also occur on a regular basis.^{88,89} Since the replisome moves at a greater speed (20×) than the TEC, even codirectional collisions happen frequently, and these can lead to double-stranded DNA breaks (DSBs). Two recent articles show that rho-dependent termination protects against DSBs by removing arrested elongation complexes.^{90,91} In particular, the DSBs caused by codirectional collisions were dependent on backtracked TECs.⁹¹ The cell has evolved a range of strategies to prevent DSBs caused by backtracked TEC. Transcription factors that improve transcription processivity and suppress backtracking and factors that reactivate backtracked TEC by transcript cleavage reduce the occurrence of DSB.⁹¹ Efficient coupling of transcription and translation, that is, of RNAP and ribosomes, suppresses backtracking and reduces DSB.⁹¹ Transcription termination factors such as rho and its cofactors that remove backtracked and arrested TEC from the genome also protect against DSB. The common denominator of all these strategies is NusG: it increases processivity, efficiently couples transcription and translation, and promotes termination in conjunction with rho.

Conclusion and Perspective

Spt5-like transcription factors are universally conserved in evolution and serve multiple functions in the three domains of life.

What is the significance of their evolutionary ancient origin, and what was the role of the Spt5 ancestor for transcription carried out by the RNAP of the LUCA?

The most basic function of Spt5-like proteins that is common to all transcription systems is the stimulation of transcription processivity carried out by the universally conserved Spt5 NGN domain. The exact composition of early genomes is unclear; they could have consisted of either DNA or RNA, either in single-stranded or in double-stranded form. The catalytic centre of multisubunit RNAPs can utilise single-stranded DNA as template²³ and, under special

circumstances, RNA as template for transcription. For example, RNAPII can transcribe the hepatitis virus delta RNA genome, and bacterial RNAP uses the noncoding 6S RNA regulator as template for RNA-dependent RNA polymerisation.^{92,93} However, the processivity of transcription on these templates is severely impaired, possibly because the interactions between RNAP and single-stranded DNA or RNA are weaker as compared to double-stranded DNA. Likewise, ancestral versions of multisubunit RNAPs were likely capable of using nonduplex DNA as template.⁹⁴ The ancestral version of Spt5 might have been crucial for robust RNAP function using these early “poor” genome templates, and thus, Spt5 could have provided a crucial selective advantage for its host. Experimental evidence supports this notion, since archaeal Spt4/5 strongly enhances the processivity of its cognate RNAP using single-stranded template DNA strand.²⁷ Not unlike the RNAP stalk,²³ the Spt5 ancestor may have played a crucial role for the expression of long genes and, during evolution, even have “permitted” an increase in gene or operon length and thereby assisted an increase in the complexity of the genetic repertoire of the organism. Once established as an integral component of TECs, other functions including the recruitment of transcription termination factors (such as rho in bacteria), the coupling of RNAPs to ribosomes (in prokaryotes), and nascent RNA processing (in eukaryotes) might have emerged over time due to the advantageous location of Spt5 in the TEC: across the DNA binding channel of RNAP, proximal to the nascent transcript, and bound to RNAP in a reversible manner. All recruitment sites for exogenous factors are located in the Spt5 KOW domain(s). The expansion of the number of KOW domains in evolution (e.g., one in prokaryotes, four in yeast, and six in humans) seems in-line with the notion that they facilitate more elaborate, later additions to the functionality of Spt5-like factors.

Acknowledgements

Research in my laboratory is currently supported by the Wellcome Trust Investigator Award WT096553MA and the Biotechnology and Biological Sciences Research Council grant BB/H019332/1 to F.W. I am grateful to Evgeny Nudler and Irina Artsimovitch for sharing and discussing unpublished information and to Patrick Cramer for sharing the RNAP–Spt4/5 models used for the preparation of Figs. 3 and 4. I would like to thank Professor Shakeel R. Saeed from the University College London Ear Institute for saving my life and thereby making it possible to continue my research into RNAPs.

References

1. Iyer, L. M., Koonin, E. V. & 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.
2. Werner, F. & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev., Microbiol.* **9**, 85–98.
3. Lane, W. J. & Darst, S. A. (2010). Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* **395**, 671–685.
4. Lane, W. J. & Darst, S. A. (2010). Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.* **395**, 686–704.
5. Carter, R. & Drouin, G. (2009). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol. Biol. Evol.* **27**, 1035–1043.
6. Sikorski, T. W. & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr. Opin. Cell Biol.* **21**, 344–351.
7. Buratowski, S. (1994). The basics of basal transcription by RNA polymerase II. *Cell*, **77**, 1–3.
8. Grohmann, D., Hirtreiter, A. & Werner, F. (2009). Molecular mechanisms of archaeal RNA polymerase. *Biochem. Soc. Trans.* **37**, 12–17.
9. Cheetham, G. M. & Steitz, T. A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **10**, 117–123.
10. Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R. & Block, S. M. (2010). *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399**, 17–30.
11. Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. (2010). Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol. Cell*, **38**, 202–210.
12. Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E. *et al.* (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell*, **34**, 710–721.
13. Ghavi-Helm, Y., Michaut, M., Acker, J., Aude, J. C., Thuriaux, P., Werner, M. & Soutourina, J. (2008). Genome-wide location analysis reveals a role of TFIIIS in RNA polymerase III transcription. *Genes Dev.* **22**, 1934–1947.
14. Hausner, W., Lange, U. & Musfeldt, M. (2000). Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**, 12393–12399.
15. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. (2003). Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334.
16. Kettenberger, H., Armache, K. J. & Cramer, P. (2003). Architecture of the RNA polymerase II–TFIIIS complex and implications for mRNA cleavage. *Cell*, **114**, 347–357.
17. Cheung, A. C. & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, **471**, 249–253.
18. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D. & Cramer, P. (2011). Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J. Biol. Chem.* **286**, 18701–18707.
19. Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. (2007). An allosteric path to transcription termination. *Mol. Cell*, **28**, 991–1001.
20. Epshtein, V., Dutta, D., Wade, J. & Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. *Nature*, **463**, 245–249.
21. Dengl, S. & Cramer, P. (2009). *Torpedo* nuclease Rat1 is insufficient to terminate RNA polymerase II *in vitro*. *J. Biol. Chem.* **284**, 21270–21279.
22. West, S., Gromak, N. & Proudfoot, N. J. (2004). Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522–525.
23. Hirtreiter, A., Grohmann, D. & 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.
24. Santangelo, T. J., Cubonova, L., Skinner, K. M. & Reeve, J. N. (2009). Archaeal intrinsic transcription termination *in vivo*. *J. Bacteriol.* **191**, 7102–7108.
25. Santangelo, T. J. & Reeve, J. N. (2006). Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J. Mol. Biol.* **355**, 196–210.
26. Spitalny, P. & Thomm, M. (2008). A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol. Microbiol.* **67**, 958–970.
27. 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.
28. Mooney, R. A., Schweimer, K., Roesch, P., Gottesman, M. & Landick, R. (2009). Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **391**, 341–358.
29. Wenzel, S., Martins, B. M., Rosch, P. & Wohrl, B. M. (2009). Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochem. J.* **425**, 373–380.
30. Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R. & Wahl, M. C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.* **21**, 4641–4653.
31. Belogurov, G. A., Sevostyanova, A., Svetlov, V. & Artsimovitch, I. (2010). Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol. Microbiol.* **76**, 286–301.
32. Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A. *et al.* (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.* **23**, 1368–1378.

33. Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A. *et al.* (2008). Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure*, **16**, 1649–1658.
34. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310.
35. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. (2011). The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell*, **43**, 253–262.
36. Andrecka, J., Treutlein, B., Arcusa, M. A., Muschielok, A., Lewis, R., Cheung, A. C. *et al.* (2009). Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res.* **37**, 5803–5809.
37. Kettenberger, H., Armache, K. J. & Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, **16**, 955–965.
38. Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X. & Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl Acad. Sci. USA*, **108**, 546–550.
39. Grohmann, D. & Werner, F. (2010). Hold on!: RNA polymerase interactions with the nascent RNA modulate transcription elongation and termination. *RNA Biol.* **7**, 310–315.
40. Brueckner, F., Ortiz, J. & Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.* **19**, 294–299.
41. 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.
42. Svetlov, V., Belogurov, G. A., Shabrova, E., Vassylyev, D. G. & Artsimovitch, I. (2007). Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res.* **35**, 5694–5705.
43. Svetlov, V. & Nudler, E. (2011). Clamping the clamp of RNA polymerase. *EMBO J.* **30**, 1190–1191.
44. 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.
45. Grohmann, D., Klose, D., Fielden, D. & Werner, F. (2011). FRET (fluorescence resonance energy transfer) sheds light on transcription. *Biochem. Soc. Trans.* **39**, 122–127.
46. Bell, S. D., Brinkman, A. B., van der Oost, J. & Jackson, S. P. (2001). The archaeal TFIIE α homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep.* **2**, 133–138.
47. Lefevre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., Pinaud, N. *et al.* (2011). Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nat. Struct. Mol. Biol.* **18**, 352–358.
48. Naji, S., Grunberg, S. & 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.
49. Werner, F. & Weinzierl, R. O. (2005). Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell Biol.* **25**, 8344–8355.
50. Grunberg, S., Bartlett, M. S., Naji, S. & Thomm, M. (2007). Transcription factor E is a part of transcription elongation complexes. *J. Biol. Chem.* **282**, 35482–35490.
51. Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J. *et al.* (2001). Binding of the initiation factor σ^{70} to core RNA polymerase is a multistep process. *Mol. Cell*, **8**, 21–31.
52. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J. & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278.
53. Murakami, K. S. & Darst, S. A. (2003). Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**, 31–39.
54. Murakami, K. S., Masuda, S. & Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, **296**, 1280–1284.
55. Zenkin, N., Kulbachinskiy, A., Yuzenkova, Y., Mustaev, A., Bass, I., Severinov, K. & Brodolin, K. (2007). Region 1.2 of the RNA polymerase σ subunit controls recognition of the -10 promoter element. *EMBO J.* **26**, 955–964.
56. Sevostyanova, A., Svetlov, V., Vassylyev, D. G. & Artsimovitch, I. (2008). 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.
57. Burova, E. & Gottesman, M. E. (1995). NusG overexpression inhibits Rho-dependent termination in *Escherichia coli*. *Mol. Microbiol.* **17**, 633–641.
58. Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L. & Gottesman, M. E. (1995). *Escherichia coli* NusG protein stimulates transcription elongation rates *in vivo* and *in vitro*. *J. Bacteriol.* **177**, 1388–1392.
59. Pasman, Z. & von Hippel, P. H. (2000). Regulation of Rho-dependent transcription termination by NusG is specific to the *Escherichia coli* elongation complex. *Biochemistry*, **39**, 5573–5585.
60. Torres, M., Balada, J. M., Zellars, M., Squires, C. & Squires, C. L. (2004). *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.* **186**, 1304–1310.
61. Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R. & Block, S. M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell*, **125**, 1083–1094.
62. Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* **34**, 1062–1066.
63. Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z. & Landick, R. (2009). Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell*, **33**, 97–108.
64. Artsimovitch, I. & Landick, R. (2002). The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell*, **109**, 193–203.

65. Nagy, G., Dobrindt, U., Schneider, G., Khan, A. S., Hacker, J. & Emody, L. (2002). Loss of regulatory protein RfaH attenuates virulence of uropathogenic *Escherichia coli*. *Infect. Immun.* **70**, 4406–4413.
66. Belogurov, G. A., Vassilyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassilyev, D. G. & Artsimovitch, I. (2007). Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell*, **26**, 117–129.
67. Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R. & Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. *EMBO J.* **28**, 112–122.
68. Malone, E. A., Fassler, J. S. & Winston, F. (1993). Molecular and genetic characterization of *SPT4*, a gene important for transcription initiation in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **237**, 449–459.
69. Hartzog, G. A., Wada, T., Handa, H. & Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369.
70. Ivanov, D., Kwak, Y. T., Guo, J. & Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**, 2970–2983.
71. Rougvie, A. E. & Lis, J. T. (1990). Postinitiation transcriptional control in *Drosophila melanogaster*. *Mol. Cell. Biol.* **10**, 6041–6045.
72. Rougvie, A. E. & Lis, J. T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell*, **54**, 795–804.
73. Margaritis, T. & Holstege, F. C. (2008). Poised RNA polymerase II gives pause for thought. *Cell*, **133**, 581–584.
74. Core, L. J. & Lis, J. T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*, **319**, 1791–1792.
75. Core, L. J., Waterfall, J. J. & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
76. Li, J. & Gilmour, D. S. (2011). Promoter proximal pausing and the control of gene expression. *Curr. Opin. Genet. Dev.* **21**, 231–235.
77. Sims, R. J., III, Belotserkovskaya, R. & Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18**, 2437–2468.
78. Ujvari, A. & Luse, D. S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol.* **13**, 49–54.
79. Cheng, B. & Price, D. H. (2008). Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res.* **36**, e135.
80. Anderson, S. J., Sikes, M. L., Zhang, Y., French, S. L., Salgia, S., Beyer, A. L. *et al.* (2011). The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. *J. Biol. Chem.* **286**, 18816–18824.
81. Viktorovskaya, O. V., Appling, F. D. & Schneider, D. A. (2011). Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *J. Biol. Chem.* **286**, 18825–18833.
82. Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T. & Reeve, J. N. (2008). Polarity in archaeal operon transcription in *Thermococcus kodakaraensis*. *J. Bacteriol.* **190**, 2244–2248.
83. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, **328**, 504–508.
84. Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E. & Rosch, P. (2010). A NusE:NusG complex links transcription and translation. *Science*, **328**, 501–504.
85. Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J. *et al.* (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*, **409**, 529–533.
86. Posfai, G., Plunkett, G., III, Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K. *et al.* (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science*, **312**, 1044–1046.
87. Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science*, **320**, 935–938.
88. Saeki, H. & Svejstrup, J. Q. (2009). Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell*, **35**, 191–205.
89. Pomerantz, R. T. & O'Donnell, M. (2010). What happens when replication and transcription complexes collide? *Cell Cycle*, **9**, 2537–2543.
90. Washburn, R. S. & Gottesman, M. E. (2011). Transcription termination maintains chromosome integrity. *Proc. Natl Acad. Sci. USA*, **108**, 792–797.
91. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell*, **146**, 533–543.
92. Lehmann, E., Brueckner, F. & Cramer, P. (2007). Molecular basis of RNA-dependent RNA polymerase II activity. *Nature*, **450**, 445–449.
93. Wassarman, K. M. (2007). 6S RNA: a small RNA regulator of transcription. *Curr. Opin. Microbiol.* **10**, 164–168.
94. Artsimovitch, I. & Vassilyev, D. G. (2007). Merging the RNA and DNA worlds. *Nat. Struct. Mol. Biol.* **14**, 1122–1123.