The EMBO Journal (2011) 30, 1302–1310 | © 2011 European Molecular Biology Organization | Some Rights Reserved 0261-4189/11
www.embojournal.org

THE EMBO JOURNAL

EMBO open

Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity

Fuensanta W Martinez-Rucobo, Sarah Sainsbury, Alan CM Cheung and Patrick Cramer*

Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Munich, Germany

Related RNA polymerases (RNAPs) carry out cellular gene transcription in all three kingdoms of life. The universal conservation of the transcription machinery extends to a single RNAP-associated factor, Spt5 (or NusG in bacteria), which renders RNAP processive and may have arisen early to permit evolution of long genes. Spt5 associates with Spt4 to form the Spt4/5 heterodimer. Here, we present the crystal structure of archaeal Spt4/5 bound to the RNAP clamp domain, which forms one side of the RNAP active centre cleft. The structure revealed a conserved Spt5–RNAP interface and enabled modelling of complexes of Spt4/5 counterparts with RNAPs from all kingdoms of life, and of the complete yeast RNAP II elongation complex with bound Spt4/5. The N-terminal NGN domain of Spt5/NusG closes the RNAP active centre cleft to lock nucleic acids and render the elongation complex stable and processive. The C-terminal KOW1 domain is mobile, but its location is restricted to a region between the RNAP clamp and wall above the RNA exit tunnel, where it may interact with RNA and/or other factors.

The EMBO Journal (2011) **30,** 1302–1310. doi:10.1038/emboj.2011.64; Published online 8 March 2011
Subject Categories: chromatin & transcription; structural biology
Keywords: gene regulation; gene transcription; multiprotein complex structure; RNA polymerase elongation; transcription elongation factor

Introduction

Structural studies of cellular RNA polymerases (RNAPs) from all three kingdoms of life revealed a conserved enzyme architecture and active centre (Zhang *et al*, 1999; Cramer *et al*, 2000, 2001, 2008; Vassylyev *et al*, 2002; Hirata *et al*, 2008; Korkhin *et al*, 2009; Grohmann and Werner, 2011). In contrast, RNAP-associated factors are not conserved between bacterial, archaeal, and eukaryotic lineages, except for the transcription elongation factor Spt5 that is called NusG in bacteria. NusG consists of an N-terminal (NGN) domain and a flexibly linked C-terminal Kyrides–Onzonis–Woese (KOW) domain (Knowlton *et al*, 2003; Mooney *et al*, 2009). Archaeal Spt5 is highly homologous to NusG and its NGN domain associates with the zinc-binding protein Spt4 to form the Spt4/5 heterodimer (Hirtreiter *et al*, 2010; Klein *et al*, 2011). Eukaryotic Spt4/5 is called DSIF in metazoans and is very similar to the archaeal heterodimer, except that Spt5 contains an additional acidic N-terminal region and an additional 3–4 C-terminal KOW domains that are followed by a C-terminal repeat region (CTR) (Hartzog *et al*, 1998; Wada *et al*, 1998; Guo *et al*, 2008).

The core function of NusG and Spt4/5 is to stimulate transcription elongation and RNAP processivity, and this function resides in the conserved NGN domain (Burova *et al*, 1995; Chen *et al*, 2009). The NGN domain binds the conserved coiled coil of the RNAP clamp (Sevostyanova *et al*, 2008; Mooney *et al*, 2009; Hirtreiter *et al*, 2010; Sevostyanova and Artsimovitch, 2010). NusG and Spt4/5 also have additional roles in transcription-coupled processes. In bacteria, NusG is required for ρ factor-dependent transcription termination (Sullivan and Gottesman, 1992; Cardinale *et al*, 2008), and it couples transcription to translation (Burmann *et al*, 2010; Proshkin *et al*, 2010). In eukaryotes, Spt4/5 is involved in mRNA 5′-capping (Wen and Shatkin, 1999), promoter-proximal gene regulation by the negative elongation factor NELF (Palangat *et al*, 2005), transcription-coupled DNA repair (Jansen *et al*, 2000), organism development (Guo *et al*, 2000), and recruitment of activation-induced cytidine deaminase to DNA during antibody diversification (Pavri *et al*, 2010). Spt4/5 is present on all transcribed yeast genes, and is apparently a general component of the elongation complex (Mayer *et al*, 2010).

To understand the mechanisms used by the Spt5/NusG elongation factor, structural details of its interaction with RNAP are required. Here, we present the crystal structure of a conserved complex of Spt4/5 from the archaeon *Pyrococcus furiosus* (*Pfu*) with the RNAP clamp domain. This structure leads to a reliable atomic model of the eukaryotic RNAP II elongation complex with Spt4/5, suggests the molecular basis of transcription processivity, and provides a framework for further studies of elongation-coupled processes. The Spt5 NGN domain binds over the RNAP active centre cleft between

*Corresponding author. Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, Munich 81377, Germany. Tel.: +49 89 2180 76965; Fax: +49 89 2180 76998; E-mail: cramer@lmb.uni-muenchen.de

Received: 4 February 2011; accepted: 17 February 2011; published online: 8 March 2011

the clamp on one side and the protrusion on the other side, enclosing the DNA–RNA hybrid and maintaining the transcription bubble. After our work had been completed, an electron microscopic reconstruction of an RNAP–Spt4/5 complex was reported that provided a medium-resolution view of the Spt4/5-containing RNAP elongation complex and resulted in similar overall conclusions (Klein *et al*, 2011).

Results

A recombinant RNAP clamp that binds Spt4/5

In long-standing efforts we could prepare milligram quantities of complexes of recombinant *Saccharomyces cerevisiae* Spt4/5 with endogenous yeast RNAP II and of *Pfu* Spt4/5 with the highly homologous endogenous *Pfu* RNAP (Materials and methods; Figure 1A and B). This demonstrated that recombinantly expressed Spt4/5 binds to endogenously purified RNAP, but these preparations never co-crystallized. We thus considered determining the structure of the isolated RNAP clamp domain in complex with Spt4/5, which could enable accurate modelling of the RNAP–Spt4/5 complex. We chose to prepare the *Pfu* complex because the *Pfu* RNAP clamp contains several shorter loops and was thus predicted to exhibit less surface flexibility. Based on the free RNAP II structure (Armache *et al*, 2005), we designed a fusion protein of the three RNAP polypeptide parts that constitute the clamp domain. We fused residues 1053–1115, 3–318, and 334–371 of the three largest *Pfu* RNAP subunits B, A′, and A″, respectively, separated by short linkers (Figure 1C). After expression of the fusion protein in bacteria, a soluble recombinant clamp domain was obtained (rClamp, Figure 1D). The rClamp protein was correctly folded since it formed a stable, apparently stoichiometric complex with recombinant Spt4/5 (Figure 1D).

Structure of RNAP clamp–Spt4/5 complex

The purified rClamp–Spt4/5 complex could be crystallized and its X-ray structure determined at 3.3 Å resolution (Materials and methods). For structure solution, we combined experimental phases obtained from anomalous diffraction of four zinc ions (three in the clamp and one in Spt4) with model phases obtained by molecular replacement with the *S. cerevisiae* clamp structure (Armache *et al*, 2005). The *Methanococcus janaschii* Spt4/5 structure (Hirtreiter *et al*, 2010) was then fitted into the experimentally phased electron density map alongside the clamp structure, and after repeated cycles of rebuilding and refinement, an atomic model of the complex was refined that only lacked the Spt5 KOW domain, which was disordered (Table I; Figure 2). In the rClamp–Spt4/5 complex, the structures of free Spt4/5 and the clamp in free RNAP are essentially unaltered, except for minor local conformational changes.

Table I Diffraction data and refinement statistics

Data collection	
Space group	C2
Unit cell axes (Å)	168.6, 107.0, 51.1
Unit cell β angle (deg)	97.1
Wavelength (Å)	1.2783
Resolution range (Å)	53.5–3.30 (3.48–3.30)[a]
Unique reflections	13 603[b] (1965)[a]
Completeness (%)	99.8 (99.6)[a]
Redundancy	5.7 (5.9)[a]
Mosaicity (deg)	0.32
R_{sym} (%)	14.0 (83.5)[a]
$I/\sigma(I)$	7.4 (2.1)[a]
Refinement	
Non-hydrogen atoms	4187
RMSD bonds	0.010
RMSD angles	1.4
R_{cryst} (%)	19.3
R_{free} (%)	24.7
Preferred (%)[c]	88.3
Allowed (%)[c]	97.6
Disallowed (%)[c]	2.4

[a]Values in parentheses are for highest resolution shell.
[b]Friedel pairs not merged.
[c]Ramachandran plot statistics from MolProbity (Chen *et al*, 2010).

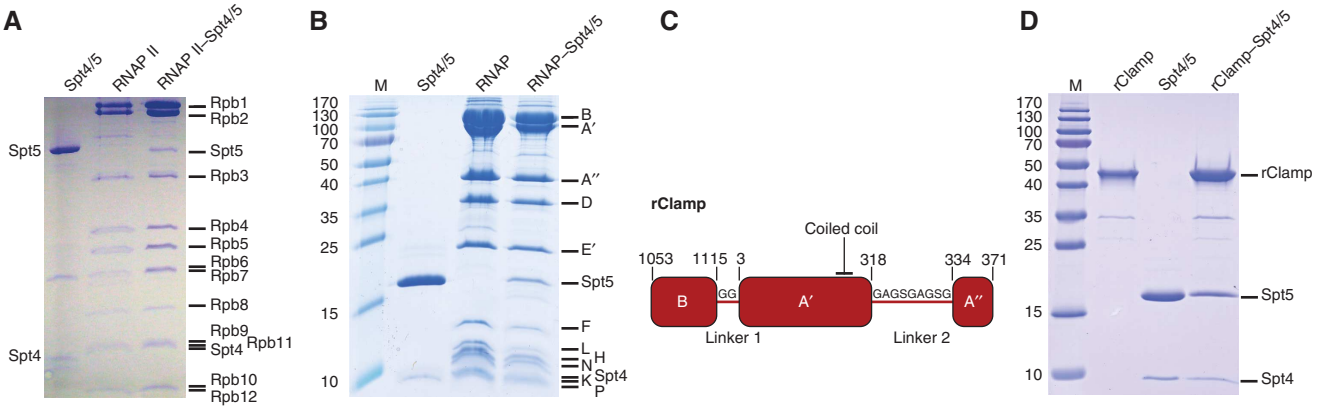

Figure 1 Spt4/5 binds endogenous RNAPs and a recombinant RNAP clamp. (**A**) Yeast (*S. cerevisiae*) complexes. SDS–PAGE analysis (Coomassie staining) of purified recombinant Spt4/5 (containing full-length Spt4 and Spt5 residues 283–853), endogenous RNAP II, and the RNAP II–Spt4/5 complex after size exclusion chromatography. The identity of the bands was confirmed by mass spectrometry. (**B**) Archaeal (*P. furiosus*, *Pfu*) complexes. SDS–PAGE analysis (Coomassie staining) of purified recombinant Spt4/5 (containing full-length Spt4 and Spt5), endogenous RNAP, and the RNAP–Spt4/5 complex after size exclusion chromatography. (**C**) Schematic of the fusion protein used to prepare the recombinant *Pfu* clamp domain. Polypeptide regions of the three largest RNAP subunits B, A′, and A″ were fused by two short linkers. Bordering residue numbers are indicated. (**D**) SDS–PAGE analysis (Coomassie staining) of recombinant *Pfu* clamp, Spt4/5, and clamp–Spt4/5 complexes.

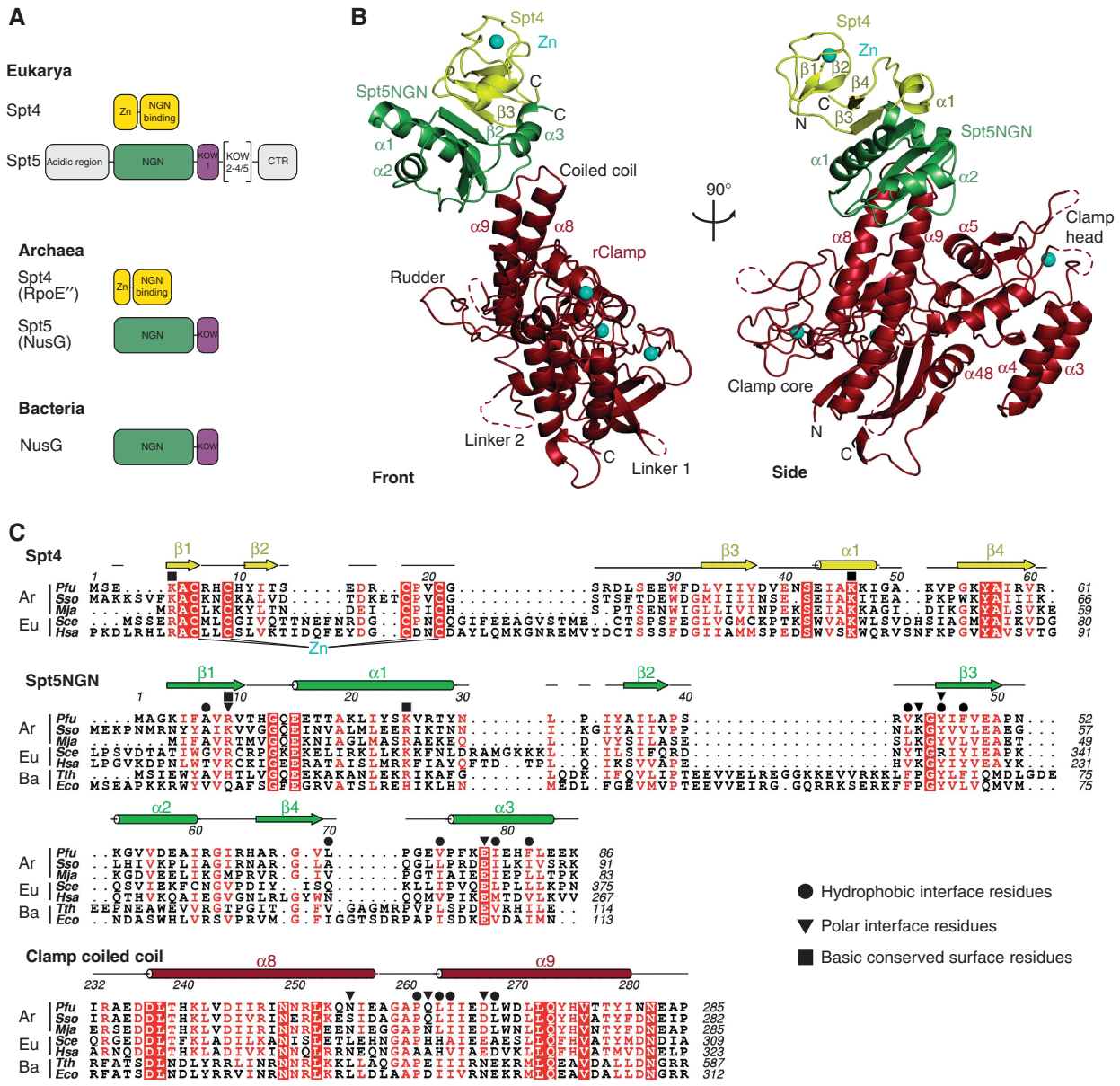

Figure 2 Crystal structure of *Pfu* RNAP clamp–Spt4/5 complex. (**A**) Schematic of domain architecture of Spt4/5 and NusG in the three kingdoms of life. Zn, zinc-binding motif; NGN, NusG N-terminal domain; KOW, Kyrides–Onzonis–Woese domain; CTR, C-terminal repeat region. (**B**) Ribbon model of the *Pfu* RNAP clamp–Spt4/5 complex crystal structure. The views correspond approximately to the front and side views of RNAP II used before (Kettenberger *et al*, 2004) and are related by a 90° rotation around a vertical axis. Spt4, the Spt5 NGN domain and the RNAP clamp are in yellow, green, and red, respectively, and these colours are used throughout. The clamp secondary structure elements are numbered in accordance with the RNAP II structure. (**C**) Amino-acid sequence alignments of Spt4, the Spt5 NGN domain, and the clamp coiled coil. Sequences from the archaea *P. furiosus* (*Pfu*), *S. solfataricus* (*Sso*), *M. janaschii* (*Mja*), the eukarya *S. cerevisiae* (*Sce*) and *H. sapiens* (*Hsa*), and the bacteria *T. thermophilus* (*Tth*) and *E. coli* (*Eco*) were used. Secondary structure elements are indicated as arrows (β-strands) or rods (α-helices). Loops are indicated with solid lines. Hydrophobic and polar residues that are part of the Spt5 NGN–clamp coiled-coil interface are marked with black dots and triangles, respectively. Conserved and positively charged surface residues on Spt4/5 are marked with black squares.

Conserved clamp–Spt5 interaction

The structure revealed that the NGN domain of Spt5 binds to the RNAP clamp coiled coil as predicted (Mooney *et al*, 2009; Hirtreiter *et al*, 2010). The clamp–Spt5 interface comprises the tip and one side of the coiled coil, and a hydrophobic concave surface patch on the Spt5 NGN domain (Figure 3). The interaction involves the clamp coiled-coil residues 255–268 from the *Pfu* RNAP subunit A′, which correspond to residues 279–292 of *S. cerevisiae* RNAP II subunit Rpb1 and residues 282–295 of *Escherichia coli* RNAP subunit β′ (Figures 2C and 3). The interaction patch on the Spt5 NGN domain involves 11 residues in three different regions of the primary sequence that cluster on the domain surface (Figure 2C). A structure-based alignment of NGN domains from eukaryotic, archaeal, and bacterial homologues revealed that the surface patch is generally conserved, including most hydrophobic residues (Figures 2C and 3). These results are consistent with mutagenesis data that indicated that the concave patch on the NGN domain interacts with the clamp (Mooney *et al*, 2009; Hirtreiter *et al*, 2010). The conservation of the clamp–Spt5 interface indicates that our structure is a good model for all complexes of Spt5/NusG with RNAPs, and suggests a general

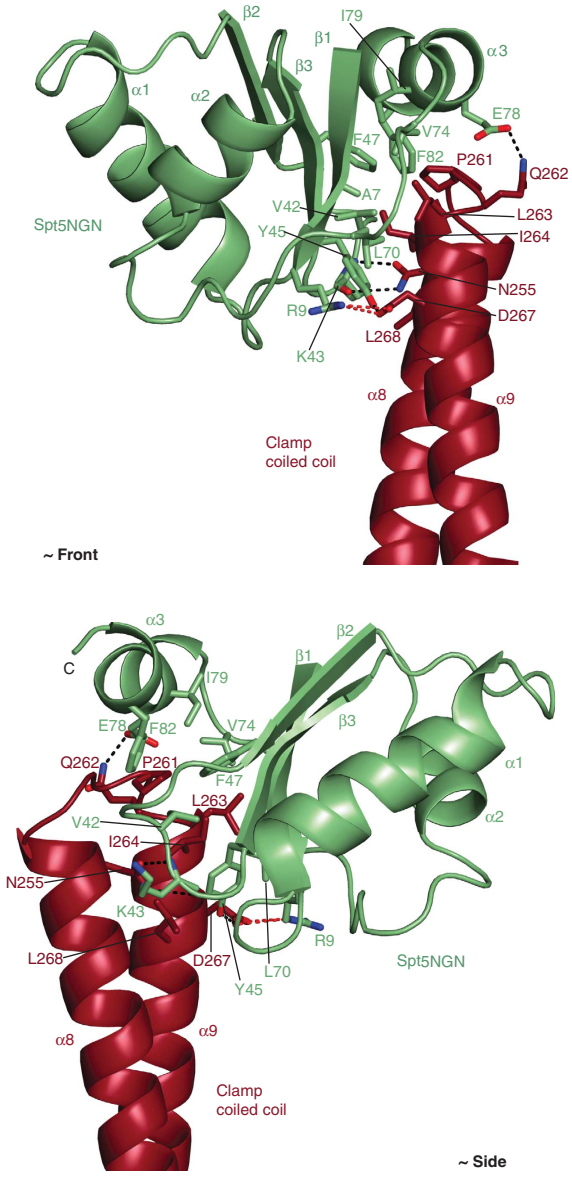

Figure 3 The Spt5 NGN–RNAP clamp coiled-coil interface. Zoom-in view of the structure in Figure 2 showing the Spt5 NGN domain (green) and the coiled coil of the RNAP clamp (red) as ribbon models. Residues that form the Spt5–clamp interface are depicted as stick models and labelled. Hydrogen bonds and salt bridges are shown as black and red dashed lines, respectively. The views correspond approximately to the front and side views used in Figure 2, albeit rotated slightly around a vertical axis to allow for optimal viewing of the interface.

architecture of the Spt5/NusG-containing RNAP elongation complex, the minimal physiological form of the elongation complex.

Spt5/NusG closes the RNAP active centre cleft

To obtain a model of the archaeal RNAP–Spt4/5 complex, we superimposed the clamp domain in our structure with the clamp in the structure of free *Sulfolobus solfataricus* RNAP (Hirata *et al*, 2008). To obtain models of the bacterial RNAP–NusG complex and the eukaryotic RNAP II–Spt4/5 complex, we repeated the superposition with the structures of *Thermus thermophilus* RNAP (Vassylyev *et al*, 2007) and *S. cerevisiae* RNAP II (Armache *et al*, 2005), respectively, and then replaced the archaeal Spt4/5 by *T. thermophilus* NusG (Reay *et al*, 2004) or yeast Spt4/5 (Guo *et al*, 2008) via superposition of their NGN domains. The resulting three models of corresponding complexes from all three kingdoms of life were free of steric clashes, and even a non-conserved domain in the bacterial RNAP (Chlenov *et al*, 2005) could be accommodated (Figure 4; Supplementary data). The models showed that the NGN domain resides above the RNAP active centre cleft, essentially closing the cleft (Figure 4). In the bacterial model, the NGN domain reaches over the cleft and resides in contact distance to the RNAP lobe and protrusion (Figure 4). In the archaeal and eukaryotic models, a contact of the NGN domain with the protrusion and lobe may also be possible if the clamp closes slightly further. Spt4 points away from the RNAP surface, consistent with its non-essential nature in eukaryotes and with the lack of an Spt4 homologue in bacteria.

The NGN domain locks nucleic acids in the cleft

We next modelled the RNAP II–Spt4/5 complex with the DNA template/non-template duplex and the RNA product by including the nucleic acids from the complete elongation complex (Kettenberger *et al*, 2004; Andrecka *et al*, 2009). To obtain a model that was free of clashes, only a minor shift of the upstream DNA was required (Figure 5). The model shows that Spt4/5 is positioned on the elongation complex such that the nucleic acids, in particular the DNA–RNA hybrid and the DNA strands forming the transcription bubble, are locked in the enzyme active centre cleft (Figure 5). Upstream and downstream DNA are thus kept separated by Spt4/5. The DNA upstream duplex and non-template strand within the bubble run along a positively charged surface of Spt4/5 (Figure 5B). At least part of this surface of the NGN domain is positively charged in all species investigated, even though only two basic residue positions are conserved (Figure 2C). The modelling is consistent with biochemical data, showing that NusG binds near the upstream edge of the transcription bubble (Sevostyanova and Artsimovitch, 2010) and that the NusG paralogue RfaH maintains the upstream bubble (Belogurov *et al*, 2010).

Restricted location of the KOW domain above exiting RNA

We next investigated the possible location of the universally conserved and flexible KOW domain located just C-terminal of the NGN domain (Figure 2A). The last ordered residue in Spt5 (residue E85) is located between the top of the clamp and wall, about 55 Å above the RNA exit tunnel (Figure 6). Since the linker from this last ordered residue to the KOW domain is restricted to a length of 11–13 residues over species, the location of the KOW domain is restricted to a sphere of a maximum radius of ~45 Å. The sphere encompasses the region between Spt4 and the RNAP clamp, wall, and Rpb4/7 subcomplex (Figure 6). To model possible locations of the KOW domain, we superimposed the NGN domains in available NusG/Spt5 structures that contain both domains (Steiner *et al*, 2002; Knowlton *et al*, 2003; Klein *et al*, 2011), with the NGN domain in our elongation complex model. The resulting positions of the KOW domain fall within the sphere and are realistic, as no clashes with RNAP were observed with the exception of one structure (PDB code 1M1G) (Figure 6 and data not shown). None of the modelled

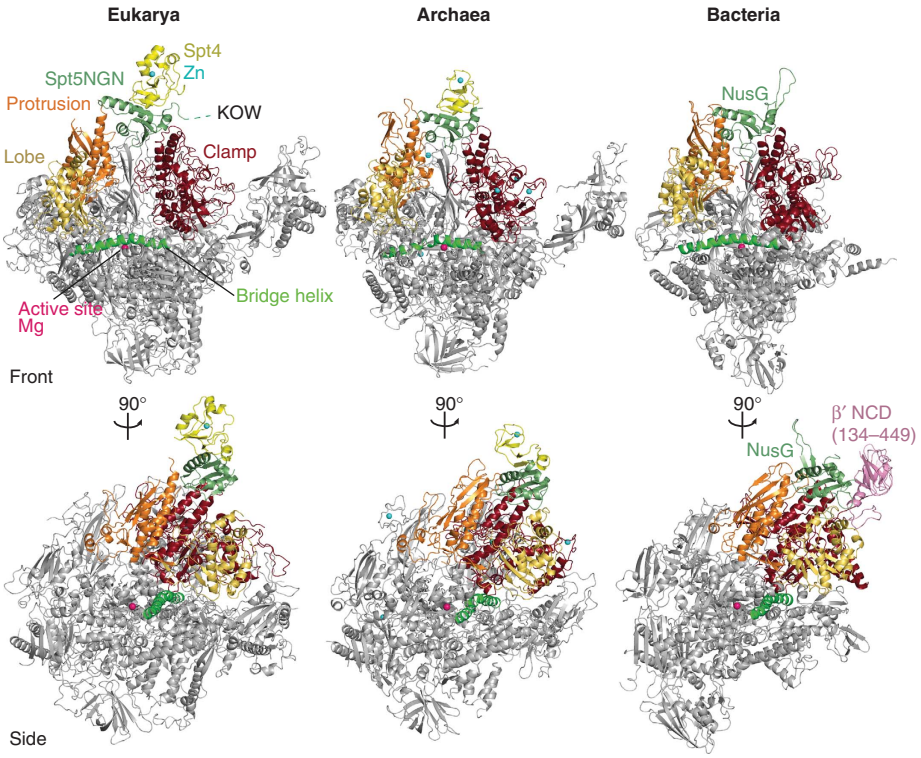

Figure 4 Models of RNAP complexes with Spt4/5 or NusG. Ribbon models for the eukaryotic RNAP II–Spt4/5 complex (left), the archaeal RNAP–Spt4/5 (middle), and the bacterial RNAP–NusG complex (right). For each model, the front (top) and side views (bottom) are shown that are also used in Figure 2. Key elements of RNAP are labelled, including a non-conserved domain (NCD) in the largest bacterial RNAP subunit (omitted from the front view). Compare text for details on model generation.

KOW domains are close to exiting RNA. Modelling also showed that the KOW domain cannot reach the RNA exit tunnel, even when the linker between the NGN and KOW domains is fully extended. The KOW domain may however contact RNA that has emerged well beyond the exit tunnel and has grown to 25–30 nucleotides in length.

Highly extended eukaryote-specific Spt5 regions

We finally considered the possible location of eukaryote-specific Spt5 regions located C-terminal to the KOW1 domain (Figure 2A). Because of a very long linker between KOW domains 1 and 2 (118 and 114 residues in yeast and human Spt5, respectively), KOW domain 2 and subsequent regions could reach any position on the Pol II surface. If all linkers between the KOW domains and the CTR would be fully extended, the C-terminus of Spt5 would be located around 2000 Å away from the RNAP II surface. This corresponds to twice the length of a fully extended C-terminal repeat domain (CTD) of yeast RNAP II (Cramer *et al*, 2001). In human Pol II, the Spt5 C-terminus could reach up to 2600 Å from the Pol II surface, and this would be about 1.6 times the length of a hypothetical totally extended CTD.

Discussion

Here, we report the crystal structure of a recombinant RNAP clamp in complex with Spt4/5, the only universally conserved transcription factor. The structure revealed a conserved clamp–Spt5 interface and enabled accurate modelling of RNAP complexes with Spt4/5 counterparts from all three kingdoms of life. These results represent a significant advance in our understanding of transcription complex architecture since atomic details of RNAP interactions with transcription factors are to date limited to the bacterial factors σ70 and Gfh1 (Vassylyev *et al*, 2002; Murakami *et al*, 2002a; Tagami *et al*, 2010), and the eukaryotic factors TFIIS and TFIIB (Kettenberger *et al*, 2003; Bushnell *et al*, 2004; Kostrewa *et al*, 2009; Liu *et al*, 2010).

Our work revealed that the Spt5 NGN domain resides above DNA and RNA bound in the RNAP active centre cleft, and provides an explanation for the universal function of NusG and Spt4/5 in transcription processivity during RNA elongation. The NGN domain locks nucleic acids in the cleft, preventing their dissociation and increasing elongation complex stability. In addition, interaction between the positively charged Spt5 surface and the negatively charged DNA non-template strand may prevent collapse of the transcription bubble. Many of these conclusions could be drawn from a recently published electron microscopic reconstruction of an RNAP–Spt4/5 complex (Klein *et al*, 2011) and are confirmed here and extended based on high-resolution data.

The data further provide insights into the initiation–elongation transition. During initiation, straight promoter DNA is melted and loaded into the cleft to trigger RNA synthesis. This results in upstream and downstream DNA duplexes that extend from RNAP at approximately right angles (Figure 5). Subsequent Spt4/5 binding may render the initiation–elongation transition irreversible, because it sterically enforces the nucleic acid arrangement and prevents RNA release and reassociation of DNA strands. Premature binding of Spt4/5 during initiation is likely prevented by initiation factors that occupy overlapping binding sites on

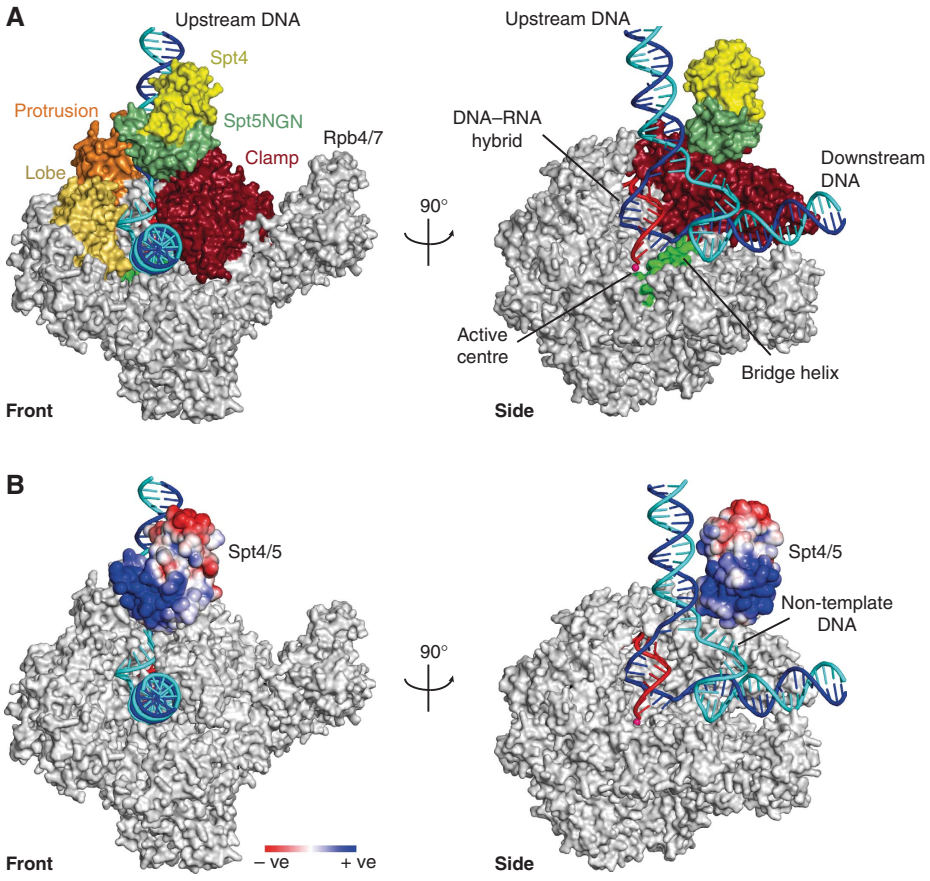

Figure 5 Model of the complete yeast RNAP II–Spt4/5 elongation complex. (**A**) Model of the complete yeast RNAP II elongation complex with bound Spt4/5. Proteins are shown as molecular surfaces with key domains highlighted in colour and labelled. Nucleic acids are shown as ribbon models with the DNA template, DNA non-template, and the RNA in blue, cyan, and red, respectively. The views are as in Figures 2 and 4. For details on how the model was prepared please compare text. (**B**) The model shown in (**A**) but with the RNAP II surface in silver and the Spt4/5 solvent-accessible surface coloured according to electrostatic charge distribution (blue, positive charge; red, negative charge) as calculated by APBS (Baker *et al*, 2001).

the clamp, in particular the bacterial factor σ70 (Vassylyev *et al*, 2002; Murakami *et al*, 2002a, b) and the archaeal/eukaryotic factor TFE/TFIIE (Chen *et al*, 2007). In addition, the results indicate that any model for transcription termination must explain how Spt4/5 is released from RNAP, to set free the nucleic acids.

Additional mechanisms contribute to Spt4/5 function during elongation and remain to be explored on a structural level. First, in an intact RNAP elongation complex, the NGN domain may contact the side of the cleft opposite the clamp, in particular the lobe and/or protrusion. This may involve subtle alterations in clamp position that could alter catalytic properties of RNAP allosterically (Hirtreiter *et al*, 2010) consistent with normal mode analysis (Yildirim and Doruker, 2004). Second, the conserved KOW1 domain adjacent to the NGN domain may contact DNA and/or exiting RNA, provided that the RNA has reached a length of 25–30 nucleotides, and such contacts could contribute to elongation complex stability and may also involve the RNAP II subcomplex Rpb4/7 (Ujvari and Luse, 2006; Cheng and Price, 2008; Missra and Gilmour, 2010). The KOW1 domain also contacts the bacterial termination factor ρ and may mediate ρ action on nearby exiting RNA. Consistent with this model, RNAP contacts are limited to the NGN domain and the KOW domain is mobile. Third, additional KOW domains that are present in eukaryotic Spt5 could contact exiting RNA and could reach anywhere on the RNAP II surface to assist in eukaryote-specific functions. For example, they could reach to the foot domain of RNAP II that was implicated in mRNA capping (Suh *et al*, 2010). Finally, the CTR that is present in eukaryotic Spt5 is subject to phosphorylation, and contributes to the recruitment of the PAF complex, which in turn recruits factors involved in chromatin modification and mRNA maturation (Liu *et al*, 2009; Zhou *et al*, 2009).

Materials and methods

Construct design and cloning

The full-length *Pfu* genes encoding Spt4 and Spt5 and the *S. cerevisiae* genes encoding full-length Spt4 and Spt5 residues 283–853 were cloned into a pET24d-derived bicistronic vector. This vector contained a second ribosomal-binding site introduced between the *Sal*I and *Not*I sites by the primer *GTCGAC*AATAATTT TGTTTAACTTTAAG**AAGGAG**ATATACATATG*GCGGCCGC* (*Sal*I and *Not*I sites are in italics, the ribosomal-binding site is bold, and *Nde*I site is underlined). The Spt4 and Spt5 genes were cloned in the vector flanked by the sites *Nco*I/*Eco*RI and *Nde*I/*Not*I, respectively. Spt5 contained a C-terminal hexahistidine tag. The DNA encoding for the rClamp was cloned into a pOPINE vector, with the three fragments connected by two linkers composed of glycines (G), alanines (A), and serines (S) (Figure 1C).

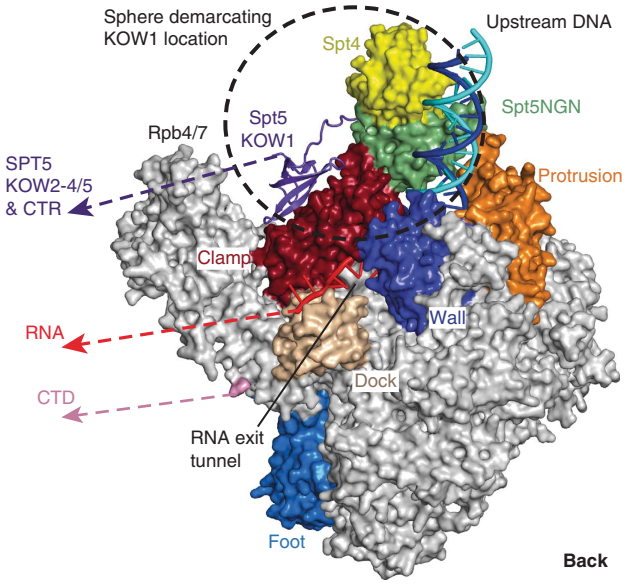

Figure 6 Relative positions of elements important for additional Spt5/NusG functions. The complete yeast RNAP II elongation complex with bound Spt4/5 (shown in Figure 5) is viewed from the back (related to the front view of Figures 2, 4, and 5 by a 180° rotation around a vertical axis). The KOW1 domain of Spt5 was positioned onto the model by superposing a representative NusG crystal structure (PDB code 1NPP) via its NGN domain and thus positioning its KOW domain relative to the Spt5 NGN. Important RNAP II elements are labelled. Exiting RNA, the C-terminal KOW domains and C-terminal repeat region of Spt5 (CTR), and the RNAP II largest subunit C-terminal domain (CTD) extend towards the same side around the Rpb4/7 subcomplex. For details compare text.

Preparation of protein complexes

All proteins were expressed in Rosetta (DE3) pLys S (Novagen) grown in LB medium at 37°C to an OD_{600} of 0.6. Expression was induced with 0.5 mM IPTG for 16 h at 20°C. Cells were lysed by sonication in buffer L1 (50 mM Tris pH 7.5, 300 mM KCl, and 3 mM DTT) for the rClamp and in buffer L2 (25 mM HEPES pH 7.5, 500 mM KOAc, 10 mM Imidazole, 0.1 mM $ZnCl_2$, 10% Glycerol, and 3 mM DTT) for Spt4/5. After centrifugation at 15 000 *g* for 20 min, the supernatant was loaded onto a 1.5-ml Ni-NTA column (Qiagen) for rClamp and onto two 1 ml Ni-NTA columns for Spt4/5. Each column was equilibrated with the respective L1 or L2 buffer. The rClamp column was washed with 10 ml and each Spt4/5 column with 5 ml of L1 (rClamp) or L2 (Spt4/5) buffer plus 100–300 mM imidazole. For archaeal proteins, a heat step (70°C, 10 min) was used to remove the *E. coli* contaminant proteins. The samples were centrifuged at 14 000 *g* for 10 min and the supernatant was applied to a Superdex 75 10/300 column (GE Healthcare) equilibrated in buffer GF1 (20 mM HEPES pH 7.0, 200 mM KCl, 5 mM DTT, and 10% Glycerol) for rClamp and in buffer GF2 (20 mM HEPES pH 7.5, 200 mM KCl, 0.1 mM ZnCl2, 20 mM Imidazole, 2.5 mM DTT, and 10% Glycerol) for Spt4/5. *S. cerevisiae* RNAP II and *Pfu* RNAP were prepared as described (Kusser *et al*, 2007; Sydow *et al*, 2009). To form the archaeal complexes, a three-fold molar excess of Spt4/5 was added to the rClamp or RNAP. For the eukaryotic complex, a 30-fold molar excess of Spt4/5 was added to the Pol II-nucleic acid complex (Kettenberger *et al*, 2004). Proteins were incubated for 1 h at 20°C. The archaeal proteins were then incubated at 70°C for 10 min. The samples were centrifuged at 14 000 *g* for 5 min prior to loading to a Superose 12 10/300 column (GE Healthcare).

Crystal structure determination

The *Pfu* rClamp–Spt4/5 complex was concentrated to 4 mg/ml. Crystals grew within 3–4 days at 20°C in hanging drops over a reservoir solution containing 10% PEG 8000, 100 mM Na/K Phosphate pH 6.2, 150 mM Guanidine hydrochloride, and 200 mM NaCl. The crystals were cryo-protected by stepwise transfer to their mother liquor supplemented with increasing concentrations of glycerol (7, 14, and 22%) and were flash frozen in liquid nitrogen. Crystals were mounted at 100 K on beamline X06SA of the Swiss Light Source, Villigen. We collected 360° of data in 0.25° increments on a PILATUS 6 M detector (DECTRIS) at the K-absorption edge of zinc. Diffraction images were integrated and scaled with XDS/XSCALE (Kabsch, 2010) or MOSFLM/CCP4 (CCP4, 1994; Leslie, 2006), to a high-resolution limit of 3.3 Å. Molecular replacement was carried out with PHASER (McCoy *et al*, 2005) using a search model from yeast RNAP II (Armache *et al*, 2005) truncated to the clamp coiled-coil domain. PHASER located the search model but revealed poor density for Spt4/5. A SAD phasing approach was then pursued where the intrinsic zinc sites were located using an anomalous difference Fourier map with phases calculated from the molecular replacement search solution. Three sites were located within the clamp coiled-coil domain and used as input sites in SAD phasing with autoSHARP (Global Phasing Limited), which found an additional fourth site corresponding to the zinc ion in Spt4. However, the resulting maps were inadequate for building until the partial model phases from PHASER were combined with SAD phases. Subsequent density modification by autoSHARP showed clear density for Spt5 and weak density for Spt4. To reduce potential model bias during phasing, a polyalanine model was built into the initial map and refined before side chain modelling. Model building and refinement were carried out with COOT (Emsley and Cowtan, 2004) and autoBUSTER (Global Phasing Limited), respectively.

Molecular modelling and figure preparation

Superpositions and molecular modelling was carried out with COOT (Emsley and Cowtan, 2004), and figures were prepared with PYMOL. Sequence alignments were edited with ALINE (Bond and Schuttelkopf, 2009).

Database accession numbers

Coordinates and structure factors for the *Pfu* RNAP clamp–Spt4/5 complex have been deposited at the protein data bank under accession number 3QQC.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Sandra Schildbach and members of the Cramer laboratory for help. We thank the Crystallization Facility at the Max-Planck-Institute for Biochemistry in Martinsried. Part of this work was performed at the Swiss Light Source (SLS) at the Paul Scherrer Institut, Villigen, Switzerland. FWMR and SS were supported by the Alexander-von-Humboldt Stiftung. PC was supported by the Deutsche Forschungsgemeinschaft, SFB646, TR5, FOR1068, NIM, the Bioimaging Network BIN, and the Jung-Stiftung.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Andrecka J, Treutlein B, Arcusa MA, Muschielok A, Lewis R, Cheung AC, Cramer P, Michaelis J (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

Armache K-J, Mitterweger S, Meinhart A, Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex Rpb4/7. *J Biol Chem* **280:** 7131–7134

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to micro-

tubules and the ribosome. *Proc Natl Acad Sci USA* **98:** 10037–10041

Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* **76:** 286–301

Bond CS, Schuttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* **65:** 510–512

Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rosch P (2010) A NusE:NusG complex links transcription and translation. *Science* **328:** 501–504

Burova E, Hung SC, Sagitov V, Stitt BL, Gottesman ME (1995) Escherichia coli NusG protein stimulates transcription elongation rates *in vivo* and *in vitro*. *J Bacteriol* **177:** 1388–1392

Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303:** 983–988

Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* **320:** 935–938

CCP4 (1994) The CCP4 Suite: programs for protein crystallography. *Acta Cryst* **D50:** 760–763

Chen H, Contreras X, Yamaguchi Y, Handa H, Peterlin BM, Guo S (2009) Repression of RNA polymerase II elongation *in vivo* is critically dependent on the C-terminus of Spt5. *PLoS One* **4:** e6918

Chen H-T, Warfield L, Hahn S (2007) The positions of TFIIF and TFIIE in the RNA polymerase II transcription initiation complex. *Nat Struct Mol Biol* **8:** 696–703

Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12–21

Cheng B, Price DH (2008) Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res* **36:** e135

Chlenov M, Masuda S, Murakami KS, Nikiforov V, Darst SA, Mustaev A (2005) Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. *J Mol Biol* **353:** 138–154

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF, Vannini A (2008) Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37:** 337–352

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288:** 640–649

Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292:** 1863–1876

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Grohmann D, Werner F (2011) Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase. *Res Microbiol* **162:** 10–18

Guo M, Xu F, Yamada J, Egelhofer T, Gao Y, Hartzog GA, Teng M, Niu L (2008) Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure* **16:** 1649–1658

Guo S, Yamaguchi Y, Schilbach S, Wada T, Lee J, Goddard A, French D, Handa H, Rosenthal A (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408:** 366–369

Hartzog GA, 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

Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451:** 851–854

Hirtreiter A, Damsma GE, Cheung AC, Klose D, Grohmann D, Vojnic E, Martin AC, Cramer P, Werner F (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* **38:** 4040–4051

Jansen LE, den Dulk H, Brouns RM, de Ruijter M, Brandsma JA, Brouwer J (2000) Spt4 modulates Rad26 requirement in transcription-coupled nucleotide excision repair. *EMBO J* **19:** 6498–6507

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125–132

Kettenberger H, Armache K-J, Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114:** 347–357

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

Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci USA* **108:** 546–550

Knowlton JR, Bubunenko M, Andrykovitch M, Guo W, Routzahn KM, Waugh DS, Court DL, Ji X (2003) A spring-loaded state of NusG in its functional cycle is suggested by X-ray crystallography and supported by site-directed mutants. *Biochemistry* **42:** 2275–2281

Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NG (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol* **7:** e102

Kostrewa D, Zeller ME, Armache KJ, Seizl M, Leike K, Thomm M, Cramer P (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462:** 323–330

Kusser A, Bertero M, Naji S, Becker T, Thomm M, Beckmann R, Cramer P (2007) Structure of an archaeal RNA polymerase. *J Mol Biol* **376:** 303–307

Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62:** 48–57

Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD (2010) Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327:** 206–209

Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J, Shokat KM, Hahn S (2009) Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol* **29:** 4852–4863

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

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61:** 458–464

Missra A, Gilmour DS (2010) 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

Mooney RA, Schweimer K, Rosch 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

Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002a) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296:** 1285–1290

Murakami KS, Masuda S, Darst SA (2002b) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **296:** 1280–1284

Palangat M, Renner DB, Price DH, Landick R (2005) A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc Natl Acad Sci USA* **102:** 15036–15041

Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, Resch W, Yamane A, Reina San-Martin B, Barreto V, Nieland TJ, Root DE, Casellas R, Nussenzweig MC (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143:** 122–133

Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328:** 504–508

Reay P, Yamasaki K, Terada T, Kuramitsu S, Shirouzu M, Yokoyama S (2004) Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from Thermus thermophilus. *Proteins* **56:** 40–51

Sevostyanova A, Artsimovitch I (2010) Functional analysis of Thermus thermophilus transcription factor NusG. *Nucleic Acids Res* **38:** 7432–7445

Sevostyanova A, Svetlov V, Vassylyev DG, Artsimovitch I (2008) The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *Proc Natl Acad Sci USA* **105:** 865–870

Steiner T, Kaiser JT, Marinkovic S, Huber R, Wahl MC (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J* **21:** 4641–4653

Suh MH, Meyer PA, Gu M, Ye P, Zhang M, Kaplan CD, Lima CD, Fu J (2010) A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. *J Biol Chem* **285:** 34027–34038

Sullivan SL, Gottesman ME (1992) Requirement for E. coli NusG protein in factor-dependent transcription termination. *Cell* **68:** 989–994

Sydow JF, Brueckner F, Cheung AC, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P (2009) Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34:** 710–721

Tagami S, Sekine S, Kumarevel T, Hino N, Murayama Y, Kamegamori S, Yamamoto M, Sakamoto K, Yokoyama S (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468:** 978–982

Ujvari A, Luse DS (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* **13:** 49–54

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **417:** 712–719

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448:** 157–162

Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H (1998) Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. *EMBO J* **17:** 7395–7403

Wen Y, Shatkin AJ (1999) Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev* **13:** 1774–1779

Yildirim Y, Doruker P (2004) Collective motions of RNA polymerases. Analysis of core enzyme, elongation complex and holoenzyme. *J Biomol Struct Dyn* **22:** 267–280

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution [see comments]. *Cell* **98:** 811–824

Zhou K, Kuo WH, Fillingham J, Greenblatt JF (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci USA* **106:** 6956–6961

