Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp

Carrie Bernecky[1], Jürgen M. Plitzko[2], and Patrick Cramer[1]*

[1]Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Biology, Am Faßberg 11, 37077 Göttingen, Germany. [2]Max-Planck-Institute for Biochemistry, Department of Molecular Structural Biology, Am Klopferspitz 18, 82152 Martinsried, Germany. Correspondence should be addressed to P.C. (pcramer@mpibpc.mpg.de).

Transcribing RNA polymerase II associates with the conserved elongation factor DSIF. DSIF renders the elongation complex stable and functions during RNA polymerase II pausing and RNA processing. We combined cryo-electron microscopy and X-ray crystallography to determine the structure of the mammalian RNA polymerase II-DSIF elongation complex at a nominal resolution of 3.4 Å. Human DSIF has a modular structure with two domains forming a DNA clamp, two domains forming an RNA clamp, and one domain buttressing the RNA clamp. The clamps maintain the transcription bubble, position upstream DNA, and retain the RNA transcript in the exit tunnel. The mobile C-terminal region of DSIF is located near exiting RNA, where it can recruit factors for RNA processing. The structure provides insight into the roles of DSIF during mRNA synthesis.

RNA polymerase II (Pol II) transcribes protein-coding genes in eukaryotic cells and requires accessory factors to elongate the messenger RNA (mRNA) chain[1]. The essential elongation factor DSIF (DRB Sensitivity Inducing Factor)[2] plays multiple roles during transcription[3,4] and is partially conserved in bacteria and archaea[5,6]. DSIF is globally required for normal RNA synthesis[7], binds to the Pol II elongation complex (EC) after transcription initiation[8], stimulates pre-mRNA capping[9], and enables promoter-proximal pausing of Pol II[10]. Contrary to observations from the bacterial and archaeal systems[11,12], stable association of DSIF with Pol II ECs and the negative role of metazoan DSIF require an exiting RNA transcript[13-16]. DSIF stabilizes a productive Pol II EC *in vitro*[14] and facilitates processive elongation[7] and termination at the 3'-end of genes *in vivo*[17,18].

DSIF comprises two subunits, SPT5 and SPT4. SPT5 contains a disordered N-terminal acidic region that is followed by an NGN (NusG N-terminal) domain[19], several KOW (Kyrpides, Ouzounis, Woese) motifs[20], and the mobile CTR (C-terminal repeat region). The CTR is phosphorylated by P-TEFb during the transcription cycle and coordinates transcription with RNA processing[3]. Previous studies have revealed that the NGN domain binds Spt4[21] and is located above the active center cleft of archaeal[22,23], yeast[23,24], and human[25] RNA polymerases. However, the precise location of the other Spt5 regions is unknown. Here we provide the structure of the complete Pol II-DSIF EC.

The structure reveals the intricate domain architecture of DSIF, the location of DSIF domains on the EC surface, the contacts between DSIF, Pol II, and nucleic acids, and provides insights into how DSIF accomplishes its functional roles.

RESULTS

Cryo-EM provides a model of the Pol II-DSIF EC

We prepared a mammalian Pol II EC on a synthetic DNA-RNA scaffold containing a 9-base pair DNA-RNA hybrid, 11 DNA mismatches to mimic the transcription bubble, and a total of 50 nucleotides of RNA (Methods). We then added human DSIF and another elongation factor (NELF) and purified the resulting complex. Cryo-electron microscopy (cryo-EM) analysis revealed an intact Pol II-DSIF EC that lacked NELF, which apparently dissociated during grid preparation, as is often observed with transient macromolecular complexes.

From a total of 2549 cryo-EM micrographs, we obtained a reconstruction at a nominal resolution of 3.4 Å (**Supplementary Fig. 1, Supplementary Fig. 2a, b**). In this reconstruction, high-quality density was observed for KOW5, but the remaining DSIF domains were only visible after filtering the volume to 15 Å (**Supplementary Fig. 2c**), reflecting mobility of the factor on the Pol II surface. We therefore carried out focused 3D classification and obtained several particle classes that showed improved resolution for particular DSIF domains (**Supplementary Fig. 1, Supplementary Fig. 2d-h**).

These reconstructions enabled building of a model for the Pol II-DSIF EC (**Fig. 1, Fig. 2a, Supplementary Fig. 3a, Table 1, Supplementary Video 1**). First, we placed our previous Pol II EC structure[25] into the density and extended it to include 20 nucleotides of RNA and the nontemplate DNA single strand in the bubble region. Second, the human NGN-SPT4 structure[26] was fitted into the density. Third, we generated models for the KOW1-Linker1 (KOW1-L1) and KOW2-KOW3 (KOW2-3) domains based on X-ray structures of the *S. cerevisiae* homologs[27] and unambiguously placed these into the density. Finally, the density for the KOW5 domain revealed side chains and was overall of such high quality that we could build an atomic model for this domain. These efforts led to a partial model of the Pol II-DSIF EC.

Crystallography completes the Pol II-DSIF EC structure

Comparison of the cryo-EM density with the partial model of the Pol II-DSIF EC showed that one additional density remained between the Pol II clamp, stalk, and the placed KOW2-3 domain. This additional density corresponded to KOW4 and a preceding region in SPT5 and could guide the design of a protein construct for crystallization (SPT5 residues 536-646) (Methods). We solved the crystal structure for this SPT5 variant at ~1.6 Å resolution (Methods, **Table 2**).

The crystal structure revealed a tandem KOW domain comprising KOW4 and a preceding, previously undetected KOW motif that we named KOWx (**Fig. 2b**). The

KOWx-4 crystal structure was placed into the single remaining cryo-EM density, completing the Pol II-DSIF EC model. The structure was finalized using flexible fitting and real space refinement, as appropriate for the local resolution we obtained (Methods). The final structure lacked only the N- and C-terminal mobile regions in SPT5, and the flexible KOW4-KOW5 linker region.

Overall architecture of the Pol II-DSIF EC

The final Pol II-DSIF EC structure (**Fig. 1b**) reveals five DSIF domains arrayed over the Pol II surface spanning from the DNA cleft to the RNA exit tunnel, and explains previous crosslinking data[14,24,25]. DSIF forms multiple interactions with the Pol II clamp, protrusion, wall, stalk, and dock domains, and with both DNA and RNA. Thereby DSIF recognizes the EC with RNA located in the RNA exit tunnel. In the following, we describe the location of these five DSIF domains on the EC surface, their interactions with Pol II and nucleic acids, and implications for their functions during transcription.

DSIF forms a DNA clamp

The Pol II-DSIF EC structure suggests that the NGN-SPT4 and KOW1-L1 domains together form a 'DNA clamp'. The NGN-SPT4 domain binds between the tips of the clamp and protrusion of a closed active center cleft (**Fig. 3**). Compared to its archaeal counterpart, which was previously positioned on the clamp[23], the NGN-SPT4 domain is slightly rotated and shifted (**Supplementary Fig. 4a**). A helix in the NGN domain (SPT5 residues 195-202) contacts the tip of the Pol II protrusion (**Supplementary Fig. 4b**), and mutation of this helix in the *S. cerevisiae* homologue reduces DSIF activity[14]. The NGN domain also contacts the nontemplate DNA strand in the transcription bubble. This is consistent with reported crosslinking to DNA in this region[14]. Additionally, the SPT4 residue Arg11 protrudes towards upstream DNA.

The KOW1-L1 domain binds between the clamp and wall. It contacts the zipper loop protruding from the clamp and positions the upstream DNA duplex. The domain surface contacting upstream DNA was identified as a nucleic acid interaction site biochemically[27]. The position of upstream DNA agrees with that observed in the free mammalian cryo-EM EC structure[25], but deviates from the path along the RPB2 wall that was observed in a yeast EC crystal structure[28]. In the latter structure, the upstream DNA formed an approximately 30° wider angle with respect to the downstream DNA. When restricted by KOW1-L1, upstream DNA orientations deviating only by ~5° are observed. An arginine- and lysine-rich loop within the KOW1-L1 domain prevents upstream DNA from directly contacting the wall domain (**Supplementary Fig. 4c**).

DSIF forms an RNA clamp

The structure further reveals that DSIF domains KOWx-4 and KOW5 form an 'RNA clamp' (**Fig. 4a, b**). KOWx-4 is buttressed by KOW2-3, which binds between the RPB4-

RPB7 stalk and the Pol II clamp. KOWx-4 contacts the wall, stalk and dock domains of Pol II. Whereas KOWx contacts the flap loop on the wall, KOW4 binds over the dock, thereby forming a bridge over the RNA exit tunnel (**Fig. 4c**). The KOWx-KOW4 linker is positively charged and contacts exiting RNA around register -15 relative to the Pol II active site (register +1). Additionally, KOW4 Arg 619 is positioned to interact with RNA at register -17 to -18. These additional RNA contacts are consistent with the observed RNase protection of 2-4 additional nucleotides of exiting RNA by DSIF[14].

The KOW5 domain binds on the outside of Pol II next to the exiting RNA in a pocket lined by RPB3, RPB11, RPB12, the dock domain, and the lower part of the wall (**Fig. 3a**). The observation that KOW5 displays the best cryo-EM density of all SPT5 domains may reflect its firm binding to Pol II, consistent with the observation that strong DSIF binding to the EC requires a region encompassing KOW5[14,29]. The arrangement of KOWx through KOW5 around the RNA exit channel explains the previously observed crosslinking of DSIF to nascent RNA[15]. The linker connecting KOW4 to KOW5 passes exiting RNA and contains an arginine-rich stretch adjacent to KOW5. Taken together, the KOW5 domain and the KOWx-4 domain, buttressed by the KOW2-3 domain, form an 'RNA clamp' that may function to retain RNA in the exit tunnel.

Nucleic acid contacts contribute to DSIF affinity for the EC

In order to test the importance of the DNA and RNA clamps for binding of DSIF to the Pol II EC, we mutated two DSIF regions that contact nucleic acids. To impair the DNA clamp, we truncated five residues of the NGN domain helix α3 and loop α3-β4 that were in close proximity to the nontemplate strand to alanine residues (**Supplementary Fig. 5a, b**). To impair the RNA clamp, we mutated four positively charged residues within the KOWx-KOW4 linker to either alanine or glutamate residues (**Supplementary Fig. 5a, b**). We then tested the ability of the wild type and mutant DSIF proteins to bind a stabilized EC containing a 20-nucleotide RNA and a DNA mismatch bubble via an electrophoretic mobility shift assay (EMSA). Both DSIF mutants displayed reduced affinity for the EC, consistent with these regions contributing to binding of DSIF to the EC (**Supplementary Fig. 5c**).

A tandem KOW domain at the C-terminal end of SPT5

We did not observe cryo-EM density for SPT5 beyond the KOW5 domain, indicating that the C-terminal region of SPT5 is mobile. However, bioinformatics analysis of the SPT5 sequence suggested the existence of an additional structured region in DSIF at the very C-terminal end of SPT5, following the CTR[19,30]. We could indeed solve the crystal structure of this C-terminal region at 1.1-Å resolution (Methods, **Table 3**). The structure reveals another tandem KOW domain that we refer to as 'KOW6-7' (**Supplementary Fig. 3b**). This domain is apparently only present in metazoan and plant DSIF homologs and corresponds to a region in DSIF that has a role in the development of the fruit fly

Drosophila melanogaster[31] and the zebrafish *Danio rerio*[32]. The sites of mutation leading to developmental defects are located within KOW6, and the mutations are predicted to destabilize the KOW6-7 structure. Both mutations have been associated with a defect in the negative regulatory roles of DSIF[31,32]. Thus, this previously unannotated tandem KOW domain in SPT5 carries out an important function *in vivo*, and the molecular mechanisms involved remain to be discovered.

DISCUSSION

We have used a combination of cryo-EM, X-ray crystallography, and molecular modeling to obtain the structure of the Pol II-DSIF EC. The results show that DSIF is a highly modular factor that adopts its structure upon binding to the Pol II EC containing a transcription bubble and the RNA transcript in the exit tunnel. DSIF contacts all EC-specific nucleic acid elements, including the nontemplate DNA in the transcription bubble, the upstream DNA, and the exiting RNA transcript. The structure thus suggests that DSIF maintains the EC with a DNA clamp and an RNA clamp.

The DNA clamp formed by DSIF apparently contributes to the maintenance of a closed active center cleft, because the NGN-SPT4 domain bridges between both sides of the cleft and restricts the location of the non-template strand in the transcription bubble. Additionally, KOW1-L1 bridges the clamp and wall domains, closing over the upstream DNA and helping to position upstream DNA. The DNA clamp is very likely important for the maintenance of a processive EC because clamp opening is predicted to weaken Pol II contacts with the DNA-RNA hybrid and is associated with polymerase pausing[33]. Disruption of the DNA clamp by mutation of the NGN domain abrogates the *in vitro* activity of DSIF in stabilizing the productive EC, and deletion of KOW1-L1 in yeast is lethal[14]. The function of the DNA clamp to maintain the EC is also consistent with a role of the NGN domain in stabilizing a minimal transcription bubble in the bacterial system[34].

The RNA clamp formed by DSIF likely contributes to maintenance of the EC. The KOW4-KOW5 linker is part of the RNA clamp and contributes to the affinity of DSIF for the EC. Several arginine residues in this linker can be methylated, and inhibition of methylation increases DSIF association with Pol II[35]. Maintenance of the RNA in the exit channel is important for EC stability and processivity, because this prevents RNA from competing with the nontemplate DNA strand for the upstream template strand, which can lead to the formation of R-loops in the wake of Pol II that destabilize the EC[36]. The RNA clamp may also restrain conformational changes in Pol II because it bridges between the mobile polymerase modules ‘core’ and ‘shelf’[37] that are known to move with respect to each other when multisubunit RNA polymerases adopt off-line states[38-40].

The RNA clamp also has roles in polymerase pausing, because a SPT5 region encompassing KOW4 and KOW5 is required for promoter-proximal pausing of Pol II[29].

The KOW4-KOW5 linker can be phosphorylated by the positive transcription elongation factor b (P-TEFb)[41] that is known to release Pol II from pause sites. These results suggest that the RNA clamp has positive and negative effects on transcription elongation, and that these depend on posttranslational modifications. A regulatory role for the RNA clamp, in contrast to a more general stabilizing function of the DNA clamp, is consistent with evolutionary considerations. Bacterial and archaeal DSIF homologues[5,6] comprise the critical regions forming the DNA clamp, but lack the regions forming the RNA clamp.

Our results further show that the observed DSIF binding to Pol II requires that transcription initiation factors have been released. Superposition of our Pol II-DSIF EC onto a structure of the human Pol II initiation complex[42] reveals many putative clashes of DSIF domains with initiation factors. The NGN-SPT4 domain would clash with the winged helix domains in TFIIE and in the small TFIIF subunit. The KOW1-L1 domain would clash with the N-terminal TFIIB cyclin domain, and the KOW2-3 domain would clash with the TFIIEα subunit. A minor clash would occur between the KOWx-4 domain and the TFIIB zinc ribbon, which resides in the RNA exit tunnel. Competition between the archaeal homologs of DSIF and TFIIE for binding the polymerase has been described[43].

Finally, the Pol II-DSIF EC structure also shows that the mobile CTR of SPT5 is positioned near exiting RNA. This location for the CTR is consistent with the function of the CTR in recruiting factors for 5'-RNA capping[44] and 3'-RNA processing[18]. RNA 3'-processing is coupled to transcription termination, when DNA is displaced from Pol II. Termination requires that the DNA and RNA clamps of DSIF are released. We speculate that when the transcribing Pol II-DSIF EC reaches the polyadenylation site at the end of a gene, the KOW5 domain remains bound to position 3'-processing factors, whereas the other DSIF domains are released from Pol II, thereby loosening the DNA and RNA clamps and rendering the polymerase prone to terminate.

When our manuscript was under review, a publication appeared that described the structure of a yeast (*K. pastoris*) Pol II EC with bound DSIF[45]. Since the DSIF domains observed in our structure are all highly conserved in yeast, the two structures should be highly similar. Indeed, the published yeast structure observed the domains NGN-Spt4, KOW1-L1, and KOW5 in locations consistent with our human structure presented here. However, the yeast structure did not reveal the KOW2-3 domain, and the location of KOW4 differs from that observed in our structure. We note that we unambiguously placed both KOW2-3 and KOW4 in our cryo-EM density map. Generation of the complete model required a new X-ray crystallographic structure of the KOWx-4 domain that is conserved in yeast but was not available when the yeast Pol II-DSIF EC was reported[45]. The conserved Pol II-DSIF EC structures provide the foundation for a mechanistic investigation of eukaryotic transcription elongation and its regulation.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ACKNOWLEDGEMENTS

We thank C. Plaschka and D. Tegunov for helpful discussions regarding electron microscopy; H. Hillen for assistance with X-ray data collection and phasing; H. Hillen, C. Dienemann, S. Sainsbury, C. Engel, S. Neyer, Y. Xu, and other members of the Cramer laboratory for helpful discussions regarding X-ray crystallography; S. Neyer, C. Dienemann, and D. Tegunov for assistance with EM and X-ray validation figures; J. Wawrzinek for setting initial crystallization screens; S. Bilakovic for the modified pET-DUET-1 vector; J. Kellermann for N-terminal sequencing; other members of the Cramer laboratory for help and discussions. Part of this work was performed at the Swiss Light Source (SLS) at the Paul Scherrer Institut, Villigen, Switzerland. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935), the Advanced Grant 'TRANSREGULON' from the European Research Council (grant agreement No 693023), and the Volkswagen Foundation.

AUTHOR CONTRIBUTIONS

C.B. designed and carried out experiments and data analysis. J.M.P. provided access to a high-end EM facility and advised on microscope setup. P.C. supervised research. C.B. and P.C. prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Kwak, H. & Lis, J.T. Control of transcriptional elongation. *Annu Rev Genet* **47**, 483-508 (2013).
2. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**, 343-56 (1998).
3. Hartzog, G.A. & Fu, J. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim Biophys Acta* **1829**, 105-15 (2013).
4. Yamaguchi, Y., Shibata, H. & Handa, H. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta* **1829**, 98-104 (2013).
5. NandyMazumdar, M. & Artsimovitch, I. Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - Shifting shapes and paradigms. *Bioessays* **37**, 324-34 (2015).
6. Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J Mol Biol* **417**, 13-27 (2012).

7. Shetty, A. et al. Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol Cell* **66**, 77-88 e5 (2017).
8. Lidschreiber, M., Leike, K. & Cramer, P. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Mol Cell Biol* **33**, 3805-16 (2013).
9. Wen, Y. & Shatkin, A.J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev* **13**, 1774-9 (1999).
10. Wu, C.H. et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev* **17**, 1402-14 (2003).
11. Hirtreiter, A. et al. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* **38**, 4040-51 (2010).
12. Liu, B. & Steitz, T.A. Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res* **45**, 968-974 (2017).
13. Cheng, B. & Price, D.H. Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res* **36**, e135 (2008).
14. Crickard, J.B., Fu, J. & Reese, J.C. Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. *J Biol Chem* **291**, 9853-70 (2016).
15. Missra, A. & Gilmour, D.S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc Natl Acad Sci U S A* **107**, 11301-6 (2010).
16. Palangat, M., Renner, D.B., Price, D.H. & Landick, R. A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc Natl Acad Sci U S A* **102**, 15036-41 (2005).
17. Baejen, C. et al. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol Cell* **66**, 38-49 e6 (2017).
18. Mayer, A. et al. The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol Cell Biol* **32**, 1321-31 (2012).
19. Ponting, C.P. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res* **30**, 3643-52 (2002).
20. Kyrpides, N.C., Woese, C.R. & Ouzounis, C.A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem Sci* **21**, 425-6 (1996).
21. Guo, M. et al. Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure* **16**, 1649-58 (2008).
22. Klein, B.J. et al. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci U S A* **108**, 546-50 (2011).
23. Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* **30**, 1302-10 (2011).
24. Li, W., Giles, C. & Li, S. Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res* **42**, 7069-83 (2014).

25. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551-4 (2016).
26. Wenzel, S., Martins, B.M., Rosch, P. & Wohrl, B.M. Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochem J* **425**, 373-80 (2009).
27. Meyer, P.A. et al. Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Mol Cell Biol* **35**, 3354-69 (2015).
28. Barnes, C.O. et al. Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Mol Cell* **59**, 258-69 (2015).
29. Qiu, Y. & Gilmour, D.S. Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. *J Biol Chem* **292**, 5555-5570 (2017).
30. Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K. & Jones, D.T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, W349-57 (2013).
31. Jennings, B.H. et al. Locus-specific requirements for Spt5 in transcriptional activation and repression in Drosophila. *Curr Biol* **14**, 1680-4 (2004).
32. Guo, S. et al. A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366-9 (2000).
33. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S.A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431-41 (2013).
34. Turtola, M. & Belogurov, G.A. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* **5**(2016).
35. Kwak, Y.T. et al. Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol Cell* **11**, 1055-66 (2003).
36. Proudfoot, N.J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).
37. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-76 (2001).
38. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502**, 650-5 (2013).
39. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol Cell* **57**, 408-21 (2015).
40. Tagami, S. et al. Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978-82 (2010).
41. Sanso, M. et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev* **30**, 117-31 (2016).
42. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-65 (2016).
43. Grohmann, D. et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* **43**, 263-74 (2011).

44. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J Biol Chem* **277**, 19639-48 (2002).
45. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* (2017).

FIGURE LEGENDS

Figure 1 Structure of the Pol II-DSIF EC. (**a**) Schematic showing the domain organization of human DSIF subunits SPT5 and SPT4. Solid and dashed black lines above the domain schematic indicate modeled regions and unmodeled regions with cryo-EM density, respectively. (**b**) Cryo-EM density for DSIF domains. Contour levels… (**c**) Two views of the structure. Pol II is shown as a semitransparent gray surface. Coloring of DSIF domains as in (b).

Figure 2 Modeling of the Pol II-DSIF EC through a combination of cryo-EM and X-ray analysis. (**a**) Cryo-EM densities for all regions of DSIF. NGN-SPT4 and KOW1-L1 are overlaid with the DSIF-EC5 density filtered to 5 Å (mesh) and filtered to 7 Å (transparent surface); KOW3 and KOWx are overlaid with the DSIF-EC2 density filtered to 4 Å and B-factor sharpened (mesh); KOW2-3 and KOWx-4 are overlaid with the DSIF-EC2 density filtered to 5 Å (transparent surface); KOW5 is overlaid with the DSIF-EC1 density filtered to 3.4 Å and B-factor sharpened. (**b**) Two views of the crystal structure of isolated KOWx-4 reveals a tandem domain connected by a linker.

Figure 3 Features of the DSIF DNA clamp. Shown are the DSIF domains NGN-SPT4 and KOW1-L1 and their interactions with DNA and Pol II domains. (**a**) Structural overview with domains constituting the DNA clamp shown in solid colors and the rest of the Pol II-DSIF EC shown as semi-transparent ribbon. The Pol II surface is indicated with a black outline. (**b**) Top view. The path of the protrusion tip is indicated with a dashed line. Residues of human DSIF homologous to those essential for crosslinking of *S. cerevisiae* Spt5 to the nontemplate DNA are in firebrick red. (**c**) Side view. The path of the flexible, positively charged loop in the KOW1-L1 domain is indicated with a dashed line. Residues of the NGN domain in close proximity to the nontemplate strand which were mutated to generate a NGN mutant form of DSIF are shown in black.

Figure 4 Features of the DSIF RNA clamp. Shown are the DSIF domains KOW2-3, KOWx-4, and KOW5 (ribbon models) and their interactions with RNA (red) and Pol II domains. (**a**) Structural overview with the domains encircling the RNA shown in solid colors, KOW2-3 shown in intermediate transparency, and the rest of the Pol II-DSIF EC shown as semi-transparent ribbon. The Pol II surface is indicated with a black outline. (**b**)

Top view. Pol II, upstream DNA, and the SPT5 domains NGN and KOW1-L1 are shown in surface representation. Pol II is colored by subunit. (**c**) Back view with the exiting RNA oriented toward the reader. All elements except for the KOWx-4 and KOW5 domains of DSIF are shown in surface representation. The positively charged KOWx-KOW4 linker residues that were mutated to generate the RNA exit contact mutant forms of DSIF are shown in black.

Table 1 Cryo-EM data collection, refinement and validation statistics

	DSIF-EC1 (Pol II core + KOW5)	**DSIF-EC2 (Stalk + KOW2-4)**	**DSIF-EC3 (DSIF architecture)**	**DSIF-EC4 (DNA clamp, upstream 1)**	**DSIF-EC5 (DNA clamp, upstream 2)**
Data collection and processing					
Magnification	37037	37037	37037	37037	37037
Voltage (kV)	300	300	300	300	300
Electron exposure (e–/Å^2)	33	33	33	33	33
Defocus range (μm)	–0.6 to –3.6	–0.6 to –3.6	–0.6 to –3.6	–0.6 to –3.6	–0.6 to –3.6
Pixel size (Å)	1.35	1.35	1.35	1.35	1.35
Initial particle images (no.)	687,928	687,928	687,928	687,928	687,928
Final particle images (no.)	659,282	139,075	101,140	76,394	67,969
Map resolution[1] (Å)	3.4	3.6	3.7	3.7	3.8
Map resolution range[2] (Å)	2.9-7 (3.4)	3.1-7 (4-5)	3.1-8 (5-7)	3.1-8 (5-7)	3.2-7 (5-7)
EMDB code	EMD-3815	EMD-3816	EMD-3817	EMD-3818	EMD-3819
Refinement	**Pol II core + KOW5**	**Pol II stalk + KOW3-x**	**Full model[3]**		
Initial models used (PDB code)	5FLM, 2E70	2C35, 5OHO	5FLM, 2E70, 2C35, 5OHO, 3H7H		
Map sharpening *B* factor (Å^2)	-128	-112	0		
Model composition					
Non-hydrogen atoms	31,230	3,199	37,988		
Protein residues	3,662	406	4,513		
Nucleic acid residues	97		97		
R.m.s. deviations					
Bond lengths (Å)	0.01	0.01	0.01		
Bond angles (°)	1.01	1.04	1.06		
Validation					
MolProbity score	1.90	2.30	1.88		
Clashscore	11.65	12.09	11.19		
Poor rotamers (%)	0.1	2.6	0.3		
Ramachandran plot					
Favored (%)	95.4	94.2	95.5		
Allowed (%)	4.5	4.3	4.3		
Disallowed (%)	0.1	1.5	0.2		
PDB code			5OIK		

Table 2 Data collection, phasing and refinement statistics for KOWx-4

	Native	Se-Met-SAD
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	50.3, 54.4, 73.1	50.2, 54.9, 74.0
α, β, γ (°)	90	90
Resolution (Å)	44–1.6 (1.7–1.6)*	44–1.6 (1.7–1.6)*
R_{merge} (%)	4.7 (133.6)	5.4 (174.7)
$I / \sigma I$	15.9 (0.9)	15.9 (0.8)
Completeness (%)	99.7 (98.7)	99.7 (99.0)
Redundancy	6.8 (6.5)	6.9 (6.9)
Refinement		
Resolution (Å)	44–1.6	
No. reflections	50,852	
R_{work} / R_{free}	17.5/22.1	
No. atoms	2,135	
Protein	1,856	
Glycerol/Cl^-	25	
Water	254	
B-factors		
Protein	40.5	
Glycerol/Cl^-	55.0	
Water	54.4	
R.m.s deviations		
Bond lengths (Å)	0.006	
Bond angles (°)	0.862	
PDB code	5OHO	

* Values in parentheses are for highest-resolution shell. Friedel mates were not averaged.

Table 3 Data collection, phasing and refinement statistics for KOW6-7

	Native	S-SAD
Data collection		
Space group	$C222_1$	$C222_1$
Cell dimensions		
a, *b*, *c* (Å)	35.1, 75.7, 96.8	35.1, 75.8, 97.0
α, β, γ (°)	90	90
Resolution (Å)	48–1.1 (1.16–1.1)*	48–2.3 (2.44–2.3)*
R_{merge} (%)	3.2 (16.3)	3.5 (4.8)
$I / \sigma I$	27.7 (4.3)	39.5 (21.5)
Completeness (%)	91.7 (56.9)	95.8 (74.7)
Redundancy	6.2 (2.7)	5.4 (3.1)
Refinement		
Resolution (Å)	48–1.1	
No. reflections	93,225	
R_{work} / R_{free}	12.2/14.0	
No. of non-hydrogen atoms	1,226	
Protein	981	
Na^+, Cl^-	2	
Water	243	
B-factors		
Protein	16.8	
Na^+, Cl^-	26.8	
Water	38.0	
R.m.s deviations		
Bond lengths (Å)	0.010	
Bond angles (°)	1.17	
PDB code	5OHQ	

* Values in parentheses are for highest-resolution shell. Friedel mates were not averaged.

METHODS

Cloning and recombinant expression. Human DSIF and NELF variants were recombinantly expressed in *E. coli*. Gene optimized human SPT4 and SPT5 (Life Technologies) were cloned into a modified pETDuet-1 vector, in which the first cassette contained SPT4 with an N-terminal His_{10}-Arg_8-SUMO tag and 3C protease site. DSIF variants were generated by PCR amplification of the full vector. In the NGN variant, R246, L247, Y249, W250, and N251 were mutated to alanine. In the KOWx-KOW4 linker variant, R577, K578, K579, and R582 were mutated to alanine ('A' mutant) or glutamate ('E' mutant). The SPT5 variants KOWx-4 (residues 536-646) and KOW6-7 (residues 979-1087) were cloned into the pOPINB vector, containing an N-terminal His_6 tag and 3C protease site. Gene optimized human NELF subunits[46] were cloned into a pETDuet-1 vector. NELF A (with an N-terminal His_6 tag and 3C protease site) and C were cloned into the first open reading frame, and NELF B and E into the second. Internal ribosome entry sites were introduced between each pair of subunits.

Plasmids were transformed into *E. coli* BL21(DE3)RIL cells and cells were grown at 37°C in LB to an OD_{600} of 0.5-0.6. Expression of full length DSIF variants was induced by addition of 1 mM IPTG and 10 μM $ZnCl_2$ for 3-4 hours at 37°C. Expression of NELF and truncated versions of SPT5 was induced by the addition of 1 mM IPTG for 3-4 hours at 37°C. To generate selenomethionine-labeled KOWx-4, we used feedback inhibition of methionine synthesis. A 10 mL LB starter culture of *E. coli* was used to inoculate a 1 L culture of methionine-depleted minimal media (SelenoMet™, Molecular Dimensions) supplemented with 40 mg/L L-methionine. Cells were harvested at an OD_{600} of 1, washed twice with phosphate-buffered saline, and used to inoculate 8 L of methionine-depleted minimal media supplemented with 40 mg/L L-selenomethionine. The culture was grown at 37°C for 3 hours to an OD_{600} of 0.6. The temperature was reduced to 20°C and media were supplemented with 100 mg/L L-lysine, L-threonine, L-phenylalanine and 50 mg/L L-leucine, L-isoleucine, and L-valine. The culture was grown for an additional 30 minutes prior to induction of expression with 0.3 mM IPTG for 18h. Cells were harvested by centrifugation and stored at −80°C.

Protein preparation. Unless otherwise stated, all steps were completed at 4°C and all buffers contained 1 mM DTT. Bovine Pol II was prepared from calf thymus as described[25], except that buffers included protease inhibitor concentrations of 1 mM PMSF, 2 mM benzamidine, 0.6 μM leupeptin, and 2 μM pepstatin ('1x protease inhibitor'). *E. coli* pellets from the expression of full length DSIF variants were lysed by sonication in buffer A (50 mM HEPES pH 7.5 (25°C), 500 mM NaCl, 10% (v/v) glycerol, 10 μM $ZnCl_2$) supplemented with 150 mM imidazole and 1x protease inhibitor. Lysate was clarified by centrifugation and filtration through a 0.45-μm filter and applied to a 5-mL HisTrap HP column (GE Healthcare Life Sciences). Protein was eluted with

buffer A containing 400 mM imidazole, then subjected to 3C protease cleavage during overnight dialysis into buffer B (50 mM HEPES pH 7.5 (25°C), 300 mM NaCl, 10% (v/v) glycerol, 10 µM $ZnCl_2$) containing 130 mM imidazole. Cleaved DSIF was loaded a second time onto the HisTrap column to remove the cleaved tag and uncleaved complex. The eluate was applied to a Mono Q 5/50 GL anion exchange column (GE Healthcare Life Sciences), which was washed with buffer C (50 mM HEPES pH 7.5 (25°C), 10% (v/v) glycerol, 10 µM $ZnCl_2$) containing 350 mM NaCl. DSIF was eluted from the column in buffer C using a linear gradient from 350-1000 mM NaCl. A truncated form of SPT5 lacking the N-terminal region (N-terminal sequence MKKYAK, Edman sequencing) did not bind the Mono Q column. Residual nucleic acids and truncated SPT5 was removed by chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences) equilibrated in buffer D (20 mM HEPES pH 7.5 (25°C), 500 mM NaCl, 10% (v/v) glycerol, 10 µM $ZnCl_2$). Peak fractions were concentrated using a 50-kDa cutoff Amicon spin concentrator (EMD Millipore) to a concentration of 4-5 mg/mL. Yield was approximately 0.8 mg per L of culture.

Cells expressing KOWx-4 were lysed by sonication in buffer E (50 mM HEPES pH 7.5 (25°C), 300 mM NaCl, 10% (v/v) glycerol) supplemented with 50 mM imidazole and 1x protease inhibitor. Clarified lysate was applied to a 5-mL HisTrap HP column, which was washed with buffer E containing 100 mM imidazole and eluted with buffer E containing 250 mM imidazole. Eluate was subjected to 3C protease cleavage during overnight dialysis into buffer E supplemented with 45 mM imidazole. The cleaved product was passed over the HisTrap column, which was then washed with buffer E supplemented with 75 mM imidazole to remove uncleaved protein, cleaved tag, and 3C protease. The eluate was concentrated using a 3-kDa cutoff Amicon spin concentrator and applied to a HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences) equilibrated in buffer F (10 mM HEPES pH 7.5, 100 mM NaCl). Peak fractions were concentrated to ~35 mg/mL. Yields were approximately 10 mg and 1 mg per L of culture for native and selenomethionine-containing protein, respectively.

Cells expressing KOW6-7 were lysed as above in buffer E supplemented with 10 mM imidazole and 1x protease inhibitor. Clarified lysate was loaded onto a HisTrap column and eluted with buffer E containing 100 mM imidazole. The eluate was subjected to 3C protease cleavage overnight during dialysis into buffer G (50 mM HEPES pH 7.5 (25°C), 100 mM NaCl, 10% (v/v) glycerol). The sample was then passed over a 5-mL HiTrap Q HP column (GE Healthcare Life Sciences) equilibrated in buffer G. The column flow-through material was adjusted to contain 300 mM NaCl and 10 mM imidazole. The sample was passed over the HisTrap column, concentrated using a 3-kDa cutoff Amicon spin concentrator and applied, 10 mg at a time, to a Superdex 75 10/300 GL column (GE Healthcare Life Sciences) equilibrated in buffer F. Peak fractions were concentrated to 75 mg/mL using a 3-kDa cutoff Amicon concentrator. Yield was approximately 40 mg per L culture.

E. coli pellets from the expression of NELF were suspended in buffer H (50 mM HEPES pH 7.5 (25°C), 300 mM NaCl, 10% (v/v) glycerol, 2.5 mM $MgCl_2$, 0.1 mM $CaCl_2$, 30 mM imidazole) and lysed using a French press. Clarified lysate was loaded onto a HisTrap column and the column was washed at room temperature with buffer E supplemented with 30 mM imidazole, 5 mM ATP, and 2 mg/mL denatured *E. coli* protein. The column was cooled to 4°C and washed with several column volumes of buffer E containing 30 mM imidazole prior to complex elution with buffer E containing 300 mM imidazole. Complex-containing fractions were pooled and subjected to 3C protease cleavage during overnight dialysis into buffer E containing 22 mM imidazole. The sample was applied to the HisTrap column, followed by slow dilution of the unbound sample to 150 mM NaCl using buffer I (50 mM HEPES pH 7.5 (25°C), 10% (v/v) glycerol). The sample was applied to a Mono Q 5/50 column and eluted with a linear gradient from 150-1000 mM NaCl in buffer I, removing excess NELF A. Complex-containing fractions were combined and subjected to chromatography on a HiLoad 16/600 Superdex 200 column equilibrated in buffer J (50 mM HEPES pH 7.5 (25°C), 300 mM NaCl, 10% (v/v) glycerol). Peak fractions were combined and concentrated to 5-6 mg/mL using a 50 kDa cutoff Amicon concentrator. Yield was approximately 1 mg per L of culture.

Pol II-DSIF EC preparation. Pol II ECs were assembled by adding a 2-fold molar excess of pre-annealed template DNA (5'-GATCAAGCTCAAGTACTTAAGCCTGGTCTATACTAGTACTGCC-3') and RNA (5'-UAACGAGAUCAUAACAUUUGAACAAGAAUAUAUAUACAUAAAGACCAGGC-3') to Pol II. The reaction was incubated on ice for 5 minutes and at 25°C for 20 minutes. A 4-fold molar excess of nontemplate DNA (5'-GGCAGTACTAGTATTCTAGTATTGAAAGTACTTGAGCTTGATC-3') was added and the reaction was incubated for 20 minutes at 25°C. DSIF was added in 2-fold molar excess over Pol II and incubated with ECs at 25°C for 5 minutes. NELF was added in 1.5-fold molar excess over Pol II and incubated at 25°C for 20 minutes. The sample was applied to a Superdex 200 increase 3.2/300 GL column (GE Healthcare Life Sciences) equilibrated in buffer K (5 mM HEPES pH 7.25 (25°C), 150 mM NaCl, 10 μM $ZnCl_2$, 1 mM DTT). Peak fractions were collected and crosslinked for 20 minutes with 0.1% glutaraldehyde on ice in the dark. The reaction was quenched by addition of lysine pH 7.8 to a concentration of 50 mM. After 25 minutes, the sample was applied to the Superdex 200 increase 3.2/200 GL column, and peak fractions were used for cryo-EM.

Electron microscopy. Sample (4 μL) was applied to glow-discharged Quantifoil R 2/1 holey carbon grids, which were blotted and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) operated at 4°C and 100% humidity. Micrographs were collected

with an FEI Titan Krios equipped with a Gatan K2 Summit detector as described[25]. The TOM toolbox[47] was used to collect 2549 movie images (nominal magnification of ×37,000) in 'super-resolution mode' (0.675 Å per pixel), with a dose rate on the camera of ~6 electrons per pixel per second. One movie micrograph was collected per hole, with even illumination of the hole and a beam size of 2.3 µm. The total exposure time per movie was 9.9 s, with a total dose of 33 e- A^{-2} fractionated into 33 frames. The defocus range was from −0.6 µm to −3.6 µm. Movies were aligned and binned to the physical pixel size of 1.35 Å as described[25].

Image processing was performed using RELION 1.4[48]. Initial contrast transfer function (CTF) parameters were estimated using CTFFIND4[49]. An initial data set of 12,749 particles (box size 256 pixels) was selected using semi-automated picking in EMAN2[50]. Twelve classes were selected from unsupervised 2D classification of the initial data set and used as references for RELION autopicking of all 2549 micrographs, with a lower picking threshold applied to close-to-focus micrographs (0.6-1 µm). The autopicked particles were subjected to manual screening, resulting in a data set of 687,928 particles. Three-dimensional classification was carried out as outlined in Supplementary Figure 1. Classification using a mask encompassing the Pol II EC separated out particles with a nearly invisible clamp domain and weak DNA density. Refinement of the remaining 662,347 particles resulted in a reconstruction in which the Pol II EC and DSIF KOW5 were visible at high resolution, and the remaining DSIF domains were visible after filtering to 15 Å resolution.

Due to movement of nucleic acids and DSIF domains, 3D classification proved challenging. Consequently, we used a strategy of focused classifications within various regions of interest. One 3D classification used a first DSIF-EC reconstruction as a starting point and a mask encompassing DSIF KOW2-4. In order to further improve the density for the Pol II EC with KOW5, we subjected the data set to additional screening of micrographs according to CTF quality, as well as to per-particle CTF fitting with a tilted plane geometry as previously described[51]. Micrographs in which the CTF fit did not correlate well with the observed power spectrum until at least 0.8 of Nyquist were removed, resulting in a data set of 659,282 particles ('DSIF-EC1'). All other 3D classifications were carried out using this data set as a starting point. Through the various focused classifications, different conformations of DSIF domains, nucleic acids, the Pol II stalk domain, and the Pol II clamp domain could be observed (**Supplementary Fig. 1**).

X-ray crystallography. Native KOWx-4 (36 mg/mL) and selenomethionine-labeled KOWx-4 (25 mg/mL) were subjected to crystallization directly after purification. A Crystal Gryphon LCP robot (ARI) was used to set sitting crystallization drops containing 100 nL protein solution plus 100 nL reservoir solution in 96-well INTELLI-PLATEs (ARI). Native X-ray diffraction data were obtained from a crystal grown with a reservoir solution of 0.1 M sodium cacodylate pH 6.5 and 1.0 M tri-sodium citrate dihydrate.

Transfer into cryo-protectant solution (0.1 M sodium cacodylate pH 6.5, 1.0 M tri-sodium citrate dihydrate, and 27.5% (v/v) glycerol) was completed in four steps before flash freezing in liquid nitrogen. Selenomethionine diffraction data were obtained from an isomorphous crystal grown over a reservoir solution of 0.1 M bis-tris propane pH 6.5, 0.2M tri-sodium citrate dihydrate, and 18% (w/v) PEG 3350. Crystals were transferred to a cryoprotectant solution of 0.1 M bis-tris propane pH 6.5, 0.2 M tri-sodium citrate dihydrate, 20% (w/v) PEG 3350, and 25% (v/v) glycerol in two steps, then flash frozen in liquid nitrogen. KOW6-7 was thawed directly before crystallization in EasyXtal 15-Well Tool X-Seal plates (Qiagen). Drops contained 1 µL 75 mg/mL KOW6-7 and 1 µL reservoir solution over 600 µL of the same reservoir solution (0.1 M bis-tris propane pH 7 and 3 M sodium formate). Crystals were transferred to cryoprotectant solution (0.1 M bis-tris propane pH 7, 3 M sodium formate, 10% (v/v) glycerol) in four steps.

KOWx-4 data and KOW6-7 native X-ray data were collected at the Swiss Light Source (Villigen, Switzerland) on beamline PX1 with an EIGER 16M detector (Dectris). KOW6-7 sulfur data were collected on beamline PXIII using a PILATUS 2M detector (Dectris). KOWx-4 and KOW6-7 native data were collected at wavelengths of 1.000031 Å and 0.999987 Å, respectively. Data were processed with XDS[52], and space groups were confirmed using POINTLESS[53]. The KOWx-4 structure was phased by single wavelength anomalous diffraction (SAD) using the selenomethionine data collected at a wavelength of 0.971761 Å, whereas the KOW6-7 structure was phased using single wavelength anomalous dispersion of S atoms (S-SAD) using the native protein crystal collected at a wavelength of 2.07505 Å. In both cases, an anomalous data set and a native data set (for phase extension) were used in conjunction with the SHELX C/D/E pipeline[54]. Models were built in COOT[55] and subjected to iterative rebuilding and refinement using PHENIX[56]. Sodium ions were placed automatically by PHENIX. KOWx-4 was subjected to TLS refinement with 10 groups. For KOW6-7, individual anisotropic B-factors were refined, and stereochemistry and adp weights were optimized. The final models displayed good stereochemistry with no Ramachandran outliers. For KOWx-4, 96.9% of residues fell within the preferred regions of the Ramachandran plot, whereas for KOW6-7, 97.7% fell within preferred regions.

Modeling of the Pol II-DSIF EC. We first improved the Pol II EC model by adjusting it in COOT using the DSIF-EC1 density. Upstream DNA was adjusted using the map from class 3 of the upstream DNA mask focused classification, and additional nucleotides of exiting RNA were built using the map from class 7 of the DSIF KOW2-4 mask focused classification (**Supplementary Fig. 1**). We then fitted the previously determined human KOW5 NMR structure (PDB 2E70), adjusted it in COOT, and refined this along with the rest of the Pol II EC model in real space using PHENIX. The RPB2 flap loop was modeled as stubbed amino acids using the unsharpened DSIF-EC2 map, then refined in real space using PHENIX. The crystal structure of human RPB4 and RPB7 (PDB 2C35)[57]

was docked into the DSIF-EC2 map in UCSF Chimera[58] and regions near the Pol II core were adjusted in COOT. Chain A of the KOWx-4 structure was fitted into the DSIF-EC2 map. Alternative side chain conformations were removed. The rotamer for SPT5 R577 was changed to a preferred rotamer consistent with the density. Additionally, SPT5 residues 627 and 616-622 were replaced with their counterparts from chain B of the crystal structure. A model for human KOW2-3 (SPT5 416-520) was generated using Modeller[59] with chain B of the yeast KOW2-3 structure (PDB 4YTL)[27] as a reference. KOW3 was fit into the DSIF-EC2 map (filtered to 4 Å and B-factor sharpened) and adjusted in COOT. The KOW3-KOWx linker was built into the DSIF-EC2 map. RPB4, RPB7, KOW3, and the KOW3-KOWx linker were refined into the DSIF-EC2 map (filtered to 4 Å and B-factor sharpened) in real space using PHENIX, with reference restraints using the human RPB4-RPB7 crystal structure. KOW2 was flexibly fit into the DSIF-EC2 map (filtered to 5 Å, without B-factor sharpening) using Rosetta Relax[60].

A model for the human KOW1-L1 domain was generated using Modeller, with the yeast crystal structure of the homologous domain (PDB 4YTK)[27] as a reference. The model was rigid body fitted into the DSIF-EC3 map after removal of residues 319-334. The KOW1-L1 domain and upstream DNA were flexibly fitted into the DSIF-EC3 map using Rosetta. One prominent geometry outlier in the X-ray structure of the NGN domain (PDB 3H7H)[26] was corrected by converting a cis peptide bond at chain A residue 104 to a trans peptide bond, consistent with the 1.6 Å electron density. This model was rigid body fitted into the DSIF-EC3 density (filtered to 5 Å, without B-factor sharpening) using Chimera. Finally, the three missing amino acids connecting the NGN domain and KOW1 were modeled using COOT and geometry-regularized. To illustrate the path of the nontemplate DNA, phosphate atoms were placed in the center of the corresponding density (DSIF-EC5 map). Although significant density for the stacked bases was observed, the position of the phosphate backbone was ambiguous and as such we were unable to confidently model all atoms of the nontemplate DNA. Relative B-factors were assigned to each residue of the model with PHENIX real space ADP refinement into the DSIF-EC3 map.

The structure was validated with Molprobity[61], EMRinger[62], and by calculating the FSC of regions of the final model versus the corresponding map. EMRinger scores of 3.39 for the EC-KOW5 model and the DSIF EC1 map, and 1.54 for the Pol II stalk plus KOW3-x model and DSIF-EC2 map (4 Å), supported correct peptide backbone placement. Model versus map validation was carried out using the two maps for which side chain density was observed, DSIF-EC1 and DSIF-EC2. The Pol II core-DNA-RNA model was correlated against the DSIF-EC1 density, and the Pol II stalk plus KOW3 through KOWx were correlated against the DSIF-EC2 density. Masks for the correlation were generated using a map calculated from the model regions of interest, which was filtered to 15 Å and extended with a soft edge. Local resolution was estimated as

described[25], except that a FSC cutoff of 0.2 and a voxel size of 30 were used. Local filtering was done as described[63]. Figures were generated using UCSF Chimera[58,64].

Electrophoretic mobility shift assay (EMSA). ECs were prepared by adding pre-annealed template DNA and 5'-6-FAM-labeled RNA (DNA sequences as above; RNA sequence 5'-UAUAUGCAUAAAGACCAGGC-3') to a 1.5-fold molar excess of concentrated Pol II, to ensure that most RNA was bound by Pol II. The sample was incubated for 10 minutes on ice, followed by 10 minutes at 30°C. A 2-fold molar excess of nontemplate DNA (relative to RNA) was added and the reaction was incubated for an additional 10 minutes at 30°C. The concentrated ECs were then diluted and split into aliquots, and DSIF was added at the indicated final concentrations. Samples were incubated for 10 minutes at 30°C and incubated at room temperature for 40 minutes. Final concentration of ECs (RNA) in the reaction was 100 nM. Final reaction buffer concentrations were 20 mM HEPES pH 7.5 (25°C), 75 mM NaCl, 3 mM $MgCl_2$, 10% (v/v) glycerol, 1 µM $ZnCl_2$, and 1 mM DTT. Reactions were subjected to native polyacrylamide gel electrophoresis at 4°C on a NativePAGE Bis-Tris gel (Invitrogen). Fluorescence signal was visualized using a Typhoon FLA 9500 (GE Healthcare Life Sciences). Assays were performed in triplicate.

A Life Sciences Reporting Summary for this article is available online.

Data availability. Electron microscopy densities were deposited in the EM Data Bank under the accession codes EMD-3815, EMD-3816, EMD-3817, EMD-3818, and EMD-3819. The Pol II-DSIF EC model coordinates were deposited in the PDBe under the accession code 5OIK, and X-ray model coordinates and structure factors are available under accession codes 5OHO and 5OHQ.

METHODS-ONLY REFERENCES

46. Vos, S.M. et al. Architecture and RNA binding of the human negative elongation factor. *Elife* **5**(2016).
47. Korinek, A., Beck, F., Baumeister, W., Nickell, S. & Plitzko, J.M. Computer controlled cryo-electron microscopy--TOM(2) a software package for high-throughput applications. *J Struct Biol* **175**, 394-405 (2011).
48. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).
49. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).
50. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46 (2007).
51. Danev, R., Tegunov, D. & Baumeister, W. Using the Volta phase plate with defocus for cryo-EM single particle analysis. *bioRxiv* (2016).

52. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
53. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).
54. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-85 (2010).
55. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
56. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
57. Meka, H., Werner, F., Cordell, S.C., Onesti, S. & Brick, P. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res* **33**, 6435-44 (2005).
58. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
59. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **47**, 5 6 1-32 (2014).
60. Tyka, M.D. et al. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-18 (2011).
61. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
62. Barad, B.A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* **12**, 943-6 (2015).
63. Plaschka, C. et al. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376-80 (2015).
64. Sanner, M.F., Olson, A.J. & Spehner, J.C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305-20 (1996).