Structure of activated transcription complex Pol II-DSIF-PAF-SPT6

Seychelle M. Vos[1], Lucas Farnung[1], Marc Boehning[1], Christoph Wigge[1], Andreas Linden[2,3], Henning Urlaub[2,3], 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 Biophysical Chemistry, Bioanalytical Mass Spectrometry, Am Fassberg 11, 37077 Göttingen, Germany.* [3]*University Medical Center Göttingen, Institute of Clinical Chemistry, Bioanalytics Group, Robert-Koch-Straße 40, 37075 Göttingen, Germany.*

Gene regulation involves activation of RNA polymerase II (Pol II) that is paused and bound by the DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF). Here we show that formation of an activated Pol II elongation complex (EC) *in vitro* requires the kinase function of the positive transcription elongation factor b (P-TEFb) and the elongation factors PAF1 complex (PAF) and SPT6. The cryo-EM structure of the activated *Sus scrofa/Homo sapiens* Pol II-DSIF-PAF-SPT6 EC (EC*) was determined at 3.1 Å resolution and compared to the paused Pol II-DSIF-NELF EC (PEC) structure. PAF displaces NELF from the Pol II funnel for pause release. P-TEFb phosphorylates the Pol II linker to the C-terminal domain (CTD). SPT6 binds to the phosphorylated CTD linker and opens the RNA clamp formed by DSIF. These results provide the molecular basis for Pol II pause release and elongation activation.

Transcription of metazoan protein coding genes is regulated during the elongation phase by promoter-proximal pausing and subsequent release of paused RNA polymerase (Pol) II into elongation[1]. During pausing, the Pol II elongation complex (EC) binds the DRB-sensitivity inducing factor (DSIF), which is composed of subunits SPT4 and SPT5, and the four-subunit negative elongation factor (NELF)[2,3]. The recent structure of the mammalian Pol II-DSIF EC showed that DSIF forms clamps around upstream DNA and exiting RNA[4]. Similar results were obtained for a related yeast complex[5]. Our accompanying structure of the paused Pol II-DSIF-NELF EC (PEC)[6] shows that NELF binds the polymerase funnel and the open trigger loop, an element of the Pol II active site that closes to stimulate nucleotide addition. The PEC structure adopts an inactive state with a tilted DNA-RNA hybrid that impairs binding of the next nucleoside triphosphate (NTP) substrate. These results suggest possible mechanisms for NELF-stabilized Pol II pausing, but the molecular basis of pause release and formation of an activated EC remain unknown.

Release of paused Pol II into elongation requires the positive transcription elongation factor b (P-TEFb), which comprises the kinase CDK9 and the predominant cyclin, T1[7,8]. P-TEFb phosphorylates DSIF, NELF, and the C-terminal domain (CTD) of the large Pol II subunit RPB1[9-13]. The elongation factor PAF1 complex (PAF) was recently implicated in Pol II pausing and pause release[14,15], although its role remains unclear. PAF contains the subunits PAF1, LEO1, CTR9, CDC73, and WDR61[16-18] and is required for transcription elongation through chromatin[19,20]. SPT6 is another conserved elongation factor that is also required for chromatin transcription[21,22]. SPT6 stimulates transcription elongation *in vitro*[23] and *in vivo*[21].

Here we demonstrate that P-TEFb kinase activity enables formation of a stable complex of the Pol II-DSIF EC with PAF and SPT6 *in vitro*. We determined the cryo-EM structure of the resulting 20-subunit activated Pol II-DSIF-PAF-SPT6 EC, which we call EC*. Comparison of the EC* structure with the accompanying structure of the PEC[6] elucidates how NELF is displaced for Pol II release from pausing, and how Pol II is activated for productive RNA elongation and chromatin passage.

Formation of EC*

In the accompanying paper, we used an RNA extension assay to recapitulate NELF function in stabilizing pausing of a Pol II-DSIF EC *in vitro*[6]. To understand how Pol II is released from pause sites, we extended this assay and additionally purified recombinant human P-TEFb, a catalytically inactive P-TEFb mutant (CDK9 D149N), PAF, and SPT6 (**Methods**, **Extended Data Fig. 1a-e**, **Extended Data Table 1**). ECs were formed on a DNA-RNA scaffold (‘modified pause scaffold', **Extended Data Fig. 1b**) that enabled Pol II pausing after addition of CTP and GTP, and allowed for the use of ATP solely as a kinase substrate. Incubation of the PEC with active P-TEFb and ATP (**Extended Data Fig. 1f, g**) had no effect on pausing, as previously observed[24]. However, when PAF was additionally included, RNA extension beyond the pause site was facilitated (**Extended Data Fig. 1h, i**). When both PAF and SPT6 were included, RNA extension was strongly stimulated (**Extended Data Fig. 1j, k**). These results show that PAF can reverse NELF stabilized Pol II pausing *in vitro* when active P-TEFb and ATP are present, and that elongation is further stimulated when SPT6 is additionally present.

We then carried out RNA extension assays in the absence of NELF. When DSIF, PAF, SPT6, P-TEFb, and ATP were added to the Pol II EC, RNA extension was stimulated in a time and concentration dependent manner (**Fig. 1a; Extended Data Fig. 2a-c**). Stimulation was not observed when the D149N P-TEFb mutant was used (**Fig. 1a; Extended Data Fig 2a-c**). In the absence of DSIF, PAF and SPT6 stimulated elongation only modestly when incubated with active P-TEFb and ATP, whereas incubation with PAF or SPT6 alone had no effect (**Extended Data Fig 2d, e**). This indicates a functional interaction between PAF and SPT6, consistent with previous observations[25-27]. These results demonstrate that stimulated RNA extension *in vitro* requires the presence of DSIF, PAF, SPT6, active P-TEFb, and ATP.

Based on these functional results, we tested whether we could form a stable, activated Pol II EC *in vitro* by size exclusion chromatography (**Fig. 1b**, **Methods**, **Extended Data Fig. 2f-i**). To enable subsequent structure determination, we used a nucleic acid scaffold that contained a DNA mismatch bubble (‘EC* scaffold’, **Extended Data Fig. 1c**). We found that DSIF readily bound the resulting EC, whereas PAF and SPT6 association required P-TEFb and ATP (**Fig. 1b, Extended Data Fig. 2f-j**). These biochemical investigations led to the formation of a stable, activated EC (EC*, **Fig. 1c**) that contains Pol II, DSIF, PAF, and SPT6, and elongates RNA efficiently.

Cryo-EM structure of EC*

After purification of EC* by size exclusion chromatography and mild crosslinking with glutaraldehyde, we determined its cryo-EM structure at a nominal resolution of 3.1 Å (**Fig. 2**, **Supplementary Video 1, Extended Data Fig. 2j**). 2D classification revealed densities on the Pol II surface (**Extended Data Fig. 3, 4; Extended Data Table 2**) and resulted in a 3D reconstruction from 374,964 particles. The core of Pol II extended to ~2.6 Å resolution.

Elongation factors were resolved at lower resolutions (~12 Å for the most flexible domains), and their corresponding densities were improved by focused classification and refinement (**Extended Data Figs. 3-5**, **Methods**). This led to a total of eight cryo-EM density maps that enabled us to fit available structures and homology models (**Extended Data Fig. 3; Supplementary Table 1**). Modeling was aided by lysine crosslinking data (**Extended Data Fig. 6, Supplementary Tables 2-4**). 225 unique crosslinks were detected in structured regions, of which 210 fell into the permitted 30 Å range. The remaining 15 crosslinks formed between mobile elements of the structure (**Extended Data Fig. 6; Supplementary Table 2**).

To complete the EC* structure, we determined the crystal structure of the isolated human SPT6 tandem SH2 (tSH2) domain at 1.8 Å resolution, and unambiguously docked this new structure into the corresponding density of EC* (**Fig. 3**, **Methods, Extended Data Fig. 5f, 6f, 7, Extended Data Table 3**). The resulting structure of EC* shows good stereochemistry and lacks only mobile regions, including the terminal regions of PAF1 and LEO1, most of CDC73, the acidic N-terminal region of SPT6, and the C-terminal extensions of SPT5, SPT6, and CTR9 (**Supplementary Table 1**).

PAF and SPT6 structure and contacts

DSIF, PAF, and SPT6 are modular proteins that coat the outer surface of Pol II (**Fig. 2**). DSIF domains are arrayed around the Pol II cleft and RNA exit tunnel[4]. PAF extends along the RPB2 side and docks on the Pol II funnel. PAF is anchored to the external domains of RPB2 via its PAF1-LEO1 dimerization module (**Fig. 2b, c**). The central PAF subunit CTR9 contains 19 tetratricopeptide repeats (TPRs; residues 41-750) that each form two antiparallel α-helices (**Fig. 3a, Supplementary Table 6, Extended Data Fig. 5b**). The CTR9 TPRs form a right-handed superhelix that extends from the Pol II subunit RPB11 along RPB8 via the polymerase funnel to the foot (**Fig. 3a**). The TPRs are followed by a pair of helices that create a 'vertex' and connect to a prominent 'trestle' helix in CTR9 (CTR9 residues 807-892) (**Extended Data Fig. 5c**). The trestle extends ~100 Å from the Pol II foot to subunit RPB5 where downstream DNA enters the Pol II cleft. The vertex and TPRs 13, 14, and 18 buttress the PAF subunit WDR61, which forms a seven-bladed β-propeller[28] and faces away from Pol II (**Fig. 3a, Extended Data Fig. 5d, 8a**). CDC73 is mobile except for an 'anchor helix' that binds CTR9 TPR 17 (**Fig. 2d**).

SPT6 binds the RPB4-RPB7 stalk on the RPB1 side of Pol II (**Fig. 2c**). The SPT6 core region is well resolved and resembles the structure of the yeast SPT6 core[29] (**Fig. 3b**, **Extended Data Fig. 5g, Supplementary Table 7**). Binding of the SPT6 core to the RPB4-RPB7 stalk includes an electrostatic interaction with the RPB7 β-strands C1-C3 (**Extended Data Fig. 8b, c**). These interactions of an elongation factor with RPB4-RPB7 befit a role of the RPB4-RPB7 stalk not only during transcription initiation[30] but also during elongation[31,32]. The SPT6 tSH2 domain is tethered flexibly to the SPT6 core and docks to Pol II at the site where the CTD linker emerges to connect the CTD to the RPB1 body (CTD linker) (**Fig. 2b**).

Interactions are also observed between the elongation factors. The SPT5 domain KOWx-4 contacts the SPT6 core, explaining the known SPT5-SPT6 genetic interaction[33] and a weak physical interaction between DSIF and SPT6[23,34]. Low pass filtering of the cryo-EM maps also revealed a density extending C-terminally from the SPT6 tSH2 domain to the CTR9 vertex and TPRs 18 and 19 (**Extended Data Fig. 7e**). This is consistent with known interactions between SPT6 and CTR9[25-27]. Cryo-EM density and crosslinking data further indicate that the C-terminal tail of LEO1 contacts the upstream DNA and extends to the DNA clamp formed by the SPT5

NGN and KOW1 domains (**Extended Data Fig. 5e**, **6e**), explaining reported PAF-DSIF interactions[35-39]. Finally, initiation factors and elongation factors utilize similar regions of the Pol II surface for binding. TFIIB and TFIIE occupy similar regions as DSIF, whereas TFIIF and the coactivator complex Mediator engage regions bound by PAF and SPT6, respectively (**Extended Data Fig. 8d**). Taken together, DSIF, SPT6, and PAF are interconnected, coat a considerable portion of the Pol II surface, and could block reassociation of initiation factors.

Release of NELF and paused Pol II

Comparison of the EC* structure with the PEC structure reported in the accompanying paper[6] indicates that binding of NELF and PAF to Pol II is mutually exclusive (**Fig. 2e, Supplementary Video 2**). In particular, NELF association with RPB8, the foot, and the protrusion is sterically incompatible with binding of PAF to Pol II. To test whether NELF and PAF bind in a mutually exclusive manner, we incubated the Pol II EC with DSIF, NELF, PAF, P-TEFb, and ATP. Under these conditions, a stable Pol II-DSIF-PAF EC was formed, and NELF was excluded (**Extended Data Fig. 2k, l**). Alternatively, when P-TEFb and ATP were omitted, a stable PEC was formed and PAF was excluded (**Extended Data Figure 2k, m**). These data show that P-TEFb phosphorylation allows for NELF release and PAF binding in our defined biochemical system. Together with our structural data, our results further indicate that PAF prevents NELF reassociation with the Pol II funnel.

Comparison of the PEC and EC* structures further shows a critical difference in the conformation of the DNA-RNA hybrid. The PEC structure adopts an inactive conformation with a tilted DNA-RNA hybrid that impairs NTP substrate binding[6], whereas the EC* structure adopts the active, post-translocated conformation with a free NTP-binding site (**Fig. 3c**). The trigger loop is observed in an open conformation in both the PEC and EC*. In contrast, NELF contacts the trigger loop, whereas PAF does not. The trigger loop in EC* is thus predicted to close easily after NTP binding, to stimulate nucleotide incorporation and RNA chain elongation. These observations explain how NELF is displaced when Pol II is released from a pause site, and how the pause-stabilizing effects of NELF are overcome in EC*.

Changes in the DSIF DNA-RNA clamp

Our biochemical data show that PAF and SPT6 stimulate RNA extension (**Fig. 1a**, **Extended Data Fig. 2a, d, e**), consistent with published results of a stimulatory role of SPT6 *in vitro*[23] and *in vivo*[21]. Since neither PAF nor SPT6 reach the Pol II active site, the stimulatory effect is allosteric in nature. Comparison of the EC* structure with structures of the Pol II-DSIF EC[4] and the PEC[6] revealed several conformational changes on the Pol II surface (**Fig. 4a, Supplementary Video 3**) that can explain the stimulatory effect of SPT6. The changes include repositioning of the RPB4-RPB7 stalk (**Fig. 4b**) and rearrangement of the SPT5 domains KOW2-3 and KOWx-4 (**Fig. 4c**). KOWx-4 is rotated by 50º and moves away from exiting RNA by ~12 Å. This rearrangement breaks the previously observed contacts of the KOWx-4 linker with RNA[4], and thereby opens the RNA clamp of DSIF (**Fig. 4c**).

The rearrangement of KOW2-3 and KOWx-4 disrupts their previously observed interaction with KOW1, resulting in a 40º rotation of KOW1 away from the upstream DNA. The rotation is accompanied by a movement of the upstream DNA, which is bent away from the protrusion (**Fig. 4d**). This generates a space between the upstream DNA and protrusion that is occupied by the LEO1 C-terminal extension (**Fig. 4e**, **Extended Data Fig. 5e**). The path of

the LEO1 extension is similar to that of a linker in the small subunit of the initiation factor TFIIF[40]. LEO1 may thereby stabilize the upstream DNA and KOW1 in a new position that could facilitate rewinding of upstream DNA. DNA rewinding is beneficial for elongation and is facilitated by the SPT5 homologue NusG in the bacterial system[41]. SPT4 and the SPT5 NGN domain remain fixed to keep the Pol II cleft closed and retain nucleic acids. Taken together, PAF and SPT6 alter the DSIF DNA and RNA clamps, respectively, and stabilize the EC* conformation to stimulate elongation activity.

How P-TEFb triggers EC* formation

To investigate how P-TEFb triggers conversion of the PEC to EC*, we determined P-TEFb phosphorylation sites *in vitro* (**Extended Data Fig. 9, Supplementary Table 5**). We confirmed that P-TEFb phosphorylates the Pol II CTD and also mapped phosphorylation sites on DSIF, NELF, PAF, and SPT6. We obtained 49 phosphorylation sites, of which ten percent are known P-TEFb sites[10,12,42]. Most of the phosphorylation sites are found in databases (**Methods**), demonstrating that they are present *in vivo*. P-TEFb phosphorylates the NELF-A 'tentacle', which binds Pol II and is required for NELF-stabilized pausing[6,43]. Phosphorylation of the NELF-A tentacle may facilitate NELF dissociation[42]. Phosphorylation of the SPT5 linker that connects KOWx-4 and KOW5 may help to open the DSIF RNA clamp[12].

How does P-TEFb enable SPT6 recruitment? The SPT6 tSH2 domain lies adjacent to the CTD linker (**Fig. 2b**). It was recently reported that the yeast tSH2 domain binds to the phosphorylated CTD linker[44]. We therefore tested whether the human CTD linker can be phosphorylated by P-TEFb. Indeed, P-TEFb could phosphorylate six human CTD linker residues *in vitro* (**Extended Data Fig. 10a-f**, **Methods**), of which Thr1525 corresponds to the yeast site Thr1471[44]. Furthermore, a P-TEFb-treated CTD linker peptide bound the human tSH2 domain (**Fig. 5a**, **Methods**). We also found that the tSH2 domain is required for binding a linker-containing CTD variant, SPT6 incorporation into EC*, and elongation stimulation (**Extended Data Fig. 10g-j**), in accordance with prior work[45,46]. Finally, modelling shows that the phosphorylated CTD linker can meander along a positively charged crevice of the human tSH2 domain (**Fig. 5b, Extended Data Fig. 7f**). These results show that P-TEFb phosphorylates the human CTD linker, and this enables SPT6 tSH2 binding and stable docking of SPT6 to Pol II.

Binding of the CTD linker to SPT6 brings the CTD closer to the Pol II surface and the exiting RNA transcript. The exiting RNA in EC* passes through a positively charged groove formed between the S1 and RuvC-like domains of the SPT6 core (**Fig. 5c**). Consistent with this structural observation, SPT6 modestly binds single-stranded nucleic acids (**Extended Data Fig. 10k**), and yeast SPT6 crosslinks to nascent RNA in cells[47]. Factors involved in co-transcriptional RNA processing associate with the phosphorylated CTD, but also with the C-terminal region (CTR) of SPT5[11]. The CTR is also phosphorylated by P-TEFb[11,13] and extends from the KOW5 domain that lies adjacent to exiting RNA. Thus, the structural features involved in co-transcriptional RNA processing are clustered on the EC* surface (**Extended Data Fig. 8e**). The biochemical definition and structural characterization of EC* thus provides a starting point for analyzing elongation-coupled events such as co-transcriptional pre-mRNA processing.

DISCUSSION

We report here that Pol II release from the paused state and elongation activation requires P-TEFb, PAF, and SPT6 *in vitro*. We solved the structure of the activated Pol II-DSIF-PAF-SPT6 elongation complex, which we call EC*. Together with the accompanying paper[6], our work provides the molecular basis for Pol II pausing, release of Pol II from the paused state, and elongation activation. It also establishes a molecular framework for a detailed dissection of promoter-proximal transcriptional gene regulation by P-TEFb.

Comparison of the EC* structure with the PEC structure[6] provides a model for understanding how paused Pol II is released into elongation and how elongation is activated (**Fig. 6a**). PAF sterically competes with NELF for binding to the Pol II funnel, and P-TEFb phosphorylation influences NELF-PAF competition to facilitate NELF release and PAF binding. This is consistent with a requirement of PAF for pause release in cells[14]. P-TEFb phosphorylates not only the Pol II CTD, DSIF, NELF, PAF, and SPT6, it also targets the CTD linker, to promote SPT6 binding. Interactions of PAF and SPT6 induce conformational changes in the DSIF clamps on upstream DNA and exiting RNA, respectively. These changes may promote DNA rewinding at the upstream edge of the transcription bubble, to drive the polymerase forward[48], and may facilitate RNA passage through the exit tunnel, to further stimulate elongation.

Finally, PAF and SPT6 play important roles in enabling transcription of the natural template, chromatin[19,20]. Yeast PAF binds the major histone H3K4 methyltransferase, Set1[49]. PAF could reach downstream nucleosomes via its long trestle helix, and this may facilitate histone methylation by SET1 when Pol II approaches a nucleosome. SPT6 is a histone chaperone and its N-terminal region binds to histones and the nucleosome-interacting protein IWS1[45,46,50]. Although this SPT6 region is mobile in our structure, its location is restrained to the area between upstream and downstream DNA. This is consistent with the idea that SPT6 stores histones while Pol II transcribes through a nucleosome, thereby avoiding loss of histones, and retaining epigenetic information during Pol II passage (**Fig. 6b**). Thus, the EC* structure provides a starting point for analyzing the mechanisms of chromatin transcription.

Acknowledgements We thank A. Kühn and M. Raabe for identifying phosphorylation sites by mass spectrometry. We thank E. Wolf (Gene Center, LMU Munich) for *S. scrofa* thymus, F. Fischer and U. Neef for maintaining insect cell stocks, C. Oberthür and G. Kokic for assistance with protein purification, X. Liu and M. Ochmann for help with cloning and crystal refinement, and H.S. Hillen and SLS PXII for help with crystallographic data collection. We thank M. Geyer (U. of Bonn) for sharing WT P-TEFb expression plasmids. S.M.V. was supported by an EMBO Long-Term Fellowship (ALTF-725-2014). H.U. was supported by the Deutsche Forschungsgemeinschaft (DFG SFB860). PC was supported by the Advanced Grant TRANSREGULON (grant agreement No 693023) of the European Research Council, and the Volkswagen Foundation.

Author contributions S.M.V. designed and conducted all experiments unless stated otherwise. L.F. established and conducted SPT6 preparation and crystallized the SPT6 tSH2 domain. M.B. determined linker phosphorylation sites by mass spectrometry. C.W. assisted in cryo-EM data collection. A.L. performed cross-linking mass spectrometry, supervised by H.U.. P.C. supervised research. S.M.V. and P.C. wrote the manuscript with input from L.F., M.B., and H.U..

Author information Reprints and permissions information is available at www.nature.com/reprints. The author declare that they have no competing financial interest. Correspondence and request of materials should be addressed to PC (patrick.cramer@mpibpc.mpg.de).

REFERENCES

1. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13,** 720–731 (2012).
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,** 343–356 (1998).
3. Yamaguchi, Y. *et al.* NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. *Cell* **97,** 41–51 (1999).
4. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* **24,** 809–815 (2017).
5. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357,** 921–924 (2017).
6. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II-DSIF-NELF.
7. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270,** 12335–12338 (1995).
8. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92,** 451–462 (1998).
9. Marshall, N. F., Peng, J., Xie, Z. & Price, D. H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271,** 27176–27183 (1996).
10. Fujinaga, K. *et al.* Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24,** 787–795 (2004).
11. Yamada, T. *et al.* P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* **21,** 227–237 (2006).
12. Sansó, M. *et al.* P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev* **30,** 117–131 (2016).
13. Kim, J. B. & Sharp, P. A. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *J Biol Chem* (2001).
14. Yu, M. *et al.* RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350,** 1383–1386 (2015).
15. Chen, F. X. *et al.* PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162,** 1003–1015 (2015).
16. Zhu, B. *et al.* The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev* **19,** 1668–1673 (2005).
17. Mueller, C. L. & Jaehning, J. A. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol Cell Biol* **22,** 1971–1980 (2002).
18. Krogan, N. J. *et al.* RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* **22,** 6979–6992 (2002).
19. Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140,** 491–503 (2010).
20. Van Oss, S. B., Cucinotta, C. E. & Arndt, K. M. Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. *Trends Biochem Sci* **42,** 788–798 (2017).
21. Ardehali, M. B. *et al.* Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J* **28,** 1067–1077 (2009).

22. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301,** 1096–1099 (2003).
23. Endoh, M. *et al.* Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol Cell Biol* **24,** 3324–3336 (2004).
24. Wada, T. *et al.* FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. *Mol Cell* **5,** 1067–1072 (2000).
25. Kaplan, C. D., Holland, M. J. & Winston, F. Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. *J Biol Chem* **280,** 913–922 (2005).
26. Adelman, K. *et al.* Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26,** 250–260 (2006).
27. Dronamraju, R. & Strahl, B. D. A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. *Nucleic Acids Res* **42,** 870–881 (2014).
28. Xu, C. & Min, J. Structure and function of WD40 domain proteins. *Protein Cell* **2,** 202–214 (2011).
29. Close, D. *et al.* Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. *J Mol Biol* **408,** 697–713 (2011).
30. Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem* **266,** 71–75 (1991).
31. Jasiak, A. J. *et al.* Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J Biol Chem* **283,** 26423–26427 (2008).
32. Schulz, D., Pirkl, N., Lehmann, E. & Cramer, P. Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II. *J Biol Chem* **289,** 17446–17452 (2014).
33. Swanson, M. S. & Winston, F. SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics* **132,** 325–336 (1992).
34. Hartzog, G. A., Wada, T., Handa, H. & Winston, F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev* **12,** 357–369 (1998).
35. Qiu, H., Hu, C., Wong, C.-M. & Hinnebusch, A. G. The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. *Mol Cell Biol* **26,** 3135–3148 (2006).
36. Squazzo, S. L. *et al.* The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J* **21,** 1764–1774 (2002).
37. Chen, Y. *et al.* DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev* **23,** 2765–2777 (2009).
38. Liu, Y. *et al.* Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol* **29,** 4852–4863 (2009).
39. Zhou, K., Kuo, W.-H. W., Fillingham, J. & Greenblatt, J. F. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *PNAS* **106,** 6956–6961 (2009).
40. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533,** 359–365 (2016).
41. Turtola, M. & Belogurov, G. A. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* **5,** e18096 (2016).
42. Lu, X. *et al.* Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. *Nucleic Acids Res* **44,** 6853–6867 (2016).

43. Narita, T. *et al.* Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol* **23,** 1863–1873 (2003).
44. Sdano, M. A. *et al.* A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. *eLife* **6,** e28723 (2017).
45. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21,** 160–174 (2007).
46. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22,** 3422–3434 (2008).
47. Battaglia, S. *et al.* RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife* **6,** e25637 (2017).
48. Kireeva, M. *et al.* RNA-DNA and DNA-DNA base-pairing at the upstream edge of the transcription bubble regulate translocation of RNA polymerase and transcription rate. *Nucleic Acids Res* **1799,** 389 (2018).
49. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Mol Cell* **11,** 709–719 (2003).
50. Bortvin, A. & Winston, F. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272,** 1473–1476 (1996).

FIGURE LEGENDS

Figure 1 | Formation of EC* requires P-TEFb kinase.
a. DSIF, PAF, and SPT6 (75 nM) were incubated with Pol II (75 nM) on the modified pause scaffold (50 nM) (**Extended Data Fig. 1b**). WT P-TEFb (left) or the inactive P-TEFb mutant D149N (right) (100 nM) and 1 mM ATP were added 15 min prior to initiating transcription by adding 10 μM GTP and CTP. Reactions were quenched at various times. Experiments were performed three times.
b. Formation of EC*. The Pol II-DSIF EC assembled on the EC* scaffold was incubated with PAF and SPT6, either in the presence (+) or in the absence (-) of P-TEFb and ATP, and the resulting complexes were separated by size exclusion chromatography (dashed lines). Dashed vertical lines mark the elution peaks of free factors. Experiments were performed three times.
c. Schematic showing conversion of the paused Pol II-DSIF-NELF EC (PEC) to the activated Pol II-DSIF-PAF-SPT6 EC (EC*).

Figure 2 | Cryo-EM structure of EC*.
a. Domain architectures of DSIF, PAF and SPT6. The colour code is used throughout. Black, dashed black, and grey lines indicate regions of the EC* structure that were included as atomic model, backbone model, or backbone model with unknown register, respectively.
b-d. The EC* structure viewed from the Pol II front (**b**), side (**c**), and top (**d**). Pol II is shown as a silver surface. DSIF, PAF and SPT6 are depicted as ribbon models. DNA template, DNA nontemplate, and RNA are in blue, cyan and red, respectively. Dotted lines represent mobile protein regions.
e. PAF and NELF binding sites overlap. The EC* and PEC structures were superimposed by aligning Pol II. PAF subunits CTR9 (orange) and WDR61 (teal) clash with NELF (red).

Figure 3 | Details of EC* structure.
a. PAF subunit CTR9 contacts Pol II and positions WDR61. Orange circles demarcate CTR9 TPR repeats.
b. SPT6 structure. Shown are the SPT6 core structure within EC* and the SPT6 tSH2 crystal structure. SPT6 domains are coloured in shades of blue (DLD: death like domain, HhH_2: double helix hairpin helix domain, HtH: helix turn helix domain).
c. Nucleic acids in the EC* active center adopt a post-translocated state that can accept an incoming nucleoside triphosphate (NTP) substrate. Cryo-EM density from map A shown as mesh. The Pol II bridge helix and metal A are indicated.

Figure 4 | Conformational changes in DSIF.
a. EC* structure viewed from the top. Regions with apparent conformational changes detected after superposition of EC* and PEC[6] structures, are demarcated as dotted ovals and shown in detail in panels b-e.
b. Movement of RPB4-RPB7 stalk upon SPT6 binding. RPB4 is red and RPB7 blue, with corresponding pale shades for these subunits in the PEC.
c. Opening of the DSIF RNA clamp. SPT5 domain KOWx-4 is rotated away from exiting RNA, whereas the position of KOW5 remains unchanged.

d. Alteration of the DSIF DNA clamp. KOW1 is repositioned and upstream DNA is tilted in EC*. KOW1 rotates by 40º compared to its position in the PEC, whereas SPT4 and the SPT5 NGN domain remain in similar positions.
e. The LEO1 C-terminal extension forms a wedge between the Pol II protrusion and the upstream DNA.

Figure 5 | SPT6 binds CTD linker and RNA.
a. Fluorescence anisotropy titration shows that the SPT6 tSH2 domain binds a CTD linker peptide that was incubated with P-TEFb and ATP ($K_{d,app}$~0.84 µM ± 0.15). Points represent the mean of 3 independent experiments and error bars are the standard deviation between the replicates. Source data, **Supplementary Table 8**.
b. Model of the humanized yeast CTD linker (PDB ID: 5VKO[44]) onto the human SPT6 tSH2 crystal structure (this work). The surface representation of the tSH2 domain is coloured according to charge (blue, positive; red, negative). The conserved RPB1 phosphorylation site Thr1525 is shown as a yellow sphere. The position of phosphorylated Thr1540 is indicated.
c. Exiting RNA traverses a positively charged groove formed between SPT6 S1 and YqgF/RuvC domains.

Figure 6 | Comparison of PEC and EC* structures.
a. The structures of the PEC[6] and EC* are depicted schematically, with proteins colored as in the accompanying paper[6] and in Fig. 2a.
b. Model of putative interactions between EC* and a downstream (incoming) nucleosome during chromatin transcription.

METHODS

Cloning and protein expression

DSIF, NELF, and TFIIS were cloned as described in the accompanying paper[6]. cDNA clones encoding full-length (FL) human PAF subunits CDC73, WDR61, PAF1, LEO1, and CTR9 were obtained from the Harvard Plasmid Repository and the MRC PPU (Dundee, Scotland). cDNAs were used as PCR templates for insertion into a modified pFASTbac vector (438-A, Addgene: 55218) via ligation independent cloning (LIC)[51]. CTR9 was cloned with a C-terminal tobacco etch virus (TEV) protease cleavable 6x His tag. All subunits were incorporated into a single plasmid by successive rounds of LIC. A cDNA clone encoding FL human SPT6 (1-1726) (Harvard Plasmid Repository) was used as a PCR template for insertion of SPT6 into the 438-C vector (Addgene: 55220), which bears an N-terminal His6-MBP-tag followed by a TEV protease site. SPT6 ΔtSH2 (1-1297) was cloned by round-the-horn site-directed mutagenesis. The SPT6 tSH2 (1323-1520) was amplified from cDNA and cloned into the 438-C vector.

A baculovirus expression plasmid encoding P-TEFb (CDK9 1-372, CYCT1 1-272) (pACEBac1, Geneva Biotech) was a kind gift of Matthias Geyer (U. Bonn, Bonn, Germany). CDK9 is tagged with an N-terminal His8 tag followed by a TEV protease cleavage site. CYCT1 is tagged N-Terminally with a GST tag followed by a TEV protease cleavage site. The CDK9 mutation Asp149Asn was introduced by site directed mutagenesis.

Bacmid, virus, and protein production for PAF, SPT6, and P-TEFb were performed as previously described[52]. Sf9 (ThermoFisher), Sf21 (Expression Systems, Davis, CA, USA), and Hi5 (Expression Systems, Davis, CA, USA) cell lines were not tested for mycoplasma contamination and were not authenticated in-house. Hi5 cells expressing PAF or SPT6 or P-TEFb were harvested by centrifugation, resuspended in Lysis buffer (300 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine), flash-frozen, and stored at –80 ºC until purification.

Regions of human RPB1 corresponding to the linker and CTD (1488-1970, 1593-1970, 1488-1592) were amplified from an RPB1 cDNA clone and inserted into a modified pET24b vector with an N-terminal His6-MBP tag followed by a TEV site (Addgene: 29654, 1C vector) by LIC. The 1593-1970 variant was also cloned into a modified pGEX vector as previously described[47]. RPB1 CTD variants were overexpressed in *E. coli* BL21(DE3) RIL cells grown in LB medium. Cells were grown at 37ºC and protein expression was induced with 0.5 mM IPTG when cells reached OD600 ~0.5. Cells were harvested after 3 h by centrifugation, resuspended in Lysis buffer, flash-frozen, and stored at –80 ºC.

Protein purification

All steps were performed at 4ºC unless otherwise noted. Pol II, DSIF, NELF, and TFIIS were purified as described in the accompanying paper[6].

PAF was purified from 2-4 L of Hi5 expression. Cell pellets were lysed by sonication and cleared by centrifugation. Clarified lysate was filtered through 0.8 µm syringe filters and applied to a 5mL HisTrap HP column (GE Healthcare Life Sciences) equilibrated in Lysis 400 buffer (400 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine). The column was washed with 10 CV of Lysis 400 buffer, followed by 3

CV Lysis 800 buffer (800 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine), and 3 CV Lysis 400 buffer. The column was then equilibrated in Low salt buffer (150 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine). A HiTrap Q column equilibrated in Low salt buffer (5 mL, GE Healthcare Life Sciences) was attached to the base of the HisTrap column. The nickel column was developed over a gradient with Nickel 150 elution buffer (150 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 500 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine). The HisTrap column was then removed, and the HiTrap Q was washed with 5 CV of Low salt buffer. The HiTrap Q column was then developed over a gradient into Lysis 800 buffer. Peak fractions were assessed by SDS-PAGE and Coomassie staining. Fractions containing PAF were pooled and mixed with TEV protease[53] and lambda protein phosphatase. The protein was placed in a Slide-A-Lyzer 10 kDa MWCO (ThermoFisher Scientific) and dialyzed overnight against Lysis buffer 400 containing 1 mM $MnCl_2$. The protein was then applied to a 5 mL HisTrap column equilibrated in Lysis buffer 400 to remove uncleaved protein, the His tag, and TEV protease. The flow through was collected and concentrated in 100 kDa MWCO Amicon Ultra Centrifugal Filters (Merck). The protein was then applied to a HiLoad S200 16/600 pg column equilibrated in SE buffer (300 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, and 1 mM DTT). Peak fractions were assessed by SDS-PAGE and Coomassie staining. Pure peak fractions containing PAF were pooled and concentrated in a 100 kDa MWCO Amicon Ultra Centrifugal Filters (Merck), aliquoted, snap frozen, and stored at -80 ºC until use. Identity of individual subunits was confirmed by MS analysis.

WT SPT6, SPT6ΔtSH2, and the SPT6 tSH2 were purified from 1.2 L of Hi5 cells. Cell pellets were lysed by sonication and cleared by centrifugation. Clarified lysate was filtered through 0.8 µm syringe filters and applied to a 5 mL HisTrap HP column (GE Healthcare Life Sciences) equilibrated in Lysis buffer. The column was washed with 5 CV of Lysis buffer followed by 2 CV of High salt 1000 buffer (1000 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine) and 2 CV of Lysis buffer. The HisTrap column was then attached to a 15 mL amylose column equilibrated in Lysis buffer (New England Biolabs), packed in an XK column (GE Healthcare Life Sciences). Protein was eluted from the HisTrap column directly onto the amylose column in Nickel elution buffer (300 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 500 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine). After 5 CV, the HisTrap column was removed and the amylose column was washed with 2 CV High salt 1000 buffer followed by 2 CV lysis buffer. The protein was eluted from the amylose column in Amylose elution buffer (300 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 117 mM maltose, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine). Peak fractions were assessed on SDS-PAGE and Coomassie staining. Peak fractions corresponding to FL SPT6 were pooled, mixed with TEV protease and lambda protein phosphatase, and dialyzed against Lysis buffer with 1 mM $MnCl_2$ overnight in a 10 kDa MWCO Slide-A-Lyzer. The protein was then applied to a HisTrap column equilibrated in lysis buffer to remove the uncleaved protein, the His6-

MBP tag, and TEV protease. The flow through was collected, concentrated in a 100 kDa MWCO Amicon Ultra Centrifugal Filter (Merck), and applied to a HiLoad S200 16/600pg column equilibrated in SE buffer. Peak fractions were assessed by SDS-PAGE and Coomassie staining. Pure peak fractions containing SPT6 were pooled and concentrated in a 100 kDa MWCO Amicon Ultra Centrifugal Filters (Merck), aliquoted, flash-frozen and stored at -80 ºC until use. The tSH2 was purified in essentially the same way with the exception that the protein was not subjected to amylose purification.

WT P-TEFb and D149N P-TEFb were purified from 4 L of Hi5 expression. Cell pellets were thawed, lysed by sonication, and cleared by centrifugation. Clarified lysate was filtered through 0.8 µm syringe filters and applied to a 5mL HisTrap HP column (GE Healthcare Life Sciences) equilibrated in Lysis buffer. The column was washed with 5CV of Lysis buffer followed by 2CV of High salt 1000 buffer and 2CV of lysis buffer. The column was then washed with 5CV of Low salt buffer and connected to a HiTrap S column (GE Healthcare Life Sciences) equilibrated in Low salt buffer. The HisTrap was developed with a gradient of Nickel 150 elution buffer. The flow through was collected and peak fractions were analyzed by SDS-PAGE followed by Coomassie staining. Peak fractions containing P-TEFb were pooled and mixed with TEV protease. The protein was dialyzed against Lysis buffer overnight in a 10 kDa MWCO Slide-A-Lyzer. The protein was removed from the Slide-A-Lyzer and applied to a HisTrap column equilibrated in Lysis buffer. The flow through was collected and concentrated in a 10 kDa MWCO Amicon Ultra Centrifugal Filters (Merck) and applied to a HiLoad S200 16/600pg column equilibrated in SE buffer. Peak fractions were assessed by SDS-PAGE and Coomassie staining. Pure peak fractions containing P-TEFb were pooled and concentrated in a 10 kDa MWCO Amicon Ultra Centrifugal Filters (Merck) to a final concentration of 5-10 µM, aliquoted, snap frozen, and stored at -80 ºC until use.

His6-MBP RPB1 constructs (1488-1592, 1488-1970, and 1593-1970) were purified using a similar scheme. Cell pellets were thawed, lysed by sonication, and cleared by centrifugation. Lysates were filtered through 0.8 µm syringe filters and applied to 5 mL HisTrap columns equilibrated in Lysis buffer. The columns were washed with 10 CV lysis buffer, 2 CV High salt 1000 buffer followed by 2 CV lysis buffer. The proteins were eluted from the HisTrap column with Nickel elution buffer over a gradient of 9 CV. For the 1488-1592 construct, peak fractions were pooled and concentrated in 10 kDa MWCO Amicon Ultra Centrifugal Filters (Merck) and applied to a HiLoad S75 16/600 column equilibrated in SE buffer. For the 1488-1970 and 1593-1970 constructs, the HisTrap column was attached to an Amylose column equilibrated in lysis buffer as the HisTrap column was eluted. The HisTrap column was then removed and the amylose column was washed with 2 CV High salt 1000 buffer followed by 2 CV lysis buffer. The protein was eluted from the amylose column with Amylose elution buffer. Peak fractions were concentrated in 30 kDa MWCO Amicon Ultra Centrifugal Filters (Merck) and applied to a HiLoad S200 16/600pg column equilibrated in SE buffer. Peak fractions eluting from the S75 and S200 columns were assessed by SDS-PAGE followed by Coomassie staining. The protein constructs were concentrated as above, flash-frozen in liquid nitrogen, and stored at -80 ºC until use.

RNA extension assays

Transcription assays were performed with complementary DNA scaffolds that were designed to disfavor ATP misincorporation[54-56]. All oligos were purchased from Integrated DNA

Technologies (IDT), resuspended in water (100 µM), flash-frozen in liquid nitrogen, and stored at -80 ºC. The sequences used for transcription assays are as follows: Modified pause scaffold: template DNA 5'- CCA CAG GAA GAA CAG AAA CAA CGG GCG GAA CTA TGC CGG ACG TAC TGA CCA-3', non-template DNA 5'-Biotin-TTT TTG GTC AGT ACG TCC GGC ATA GTT CCG CCC GTT GTT TCT GTT CTT CCT GTG G-3', RNA 5'-6-FAM-UUU UUU GGC AUA GUU-3'; EC* transcription scaffold: template DNA GTT TCC CCC AGC TCC CAG CTC CCT GCT GGC TCC GAG TGG GTT CTG CCG CTC TCA ATG G, non-template DNA CCA TTG AGA GCG GCA GAA CCC ACT CGG AGC CAG CAG GGA GCT GGG AGC TGG GGG AAA C, RNA 5' 5-6 FAM- UUA AGG AAU UAA GUC GUG CGU CUA AUA ACC GGA GAG GGA ACC CAC U-3' . The Modified pause scaffold contains 13 nts of upstream DNA, 28 nts of downstream DNA, a 9-base pair (bp) DNA•RNA hybrid, and 6 nts of exiting RNA bearing a 5'-6 FAM label (**Extended Data Fig. 1**). The modified pause scaffold was derived from the pause scaffold (bacterial) [6,56] and was altered to disfavor ATP misincorporation during incubation steps with P-TEFb[54]. The EC* transcription scaffold contains 15 nts of upstream DNA, 34 nts of downstream DNA, a 9 bp DNA•RNA hybrid, and 37 nts of exiting RNA bearing a 5' -6 FAM label. The EC* transcription scaffold has same sequence as the EC* scaffold but a matched DNA bubble and an additional ten bases of DNA at the downstream edge of the DNA. RNA extension assays performed on the EC* transcription scaffold resemble the activity observed on the modified pause scaffold (**Extended Data Fig. 1l, m**).

RNA and template DNA were mixed in equimolar ratios and were annealed by incubating the nucleic acids at 95ºC for 5 min and then decreasing the temperature by 1ºC/min steps to a final temperature of 30 ºC in a thermocycler in a buffer containing 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 3 mM $MgCl_2$, and 10% (v/v) glycerol. All concentrations refer to the final concentrations used in the assay. *S. scrofa* Pol II (75 nM) and the RNA•template hybrid (50 nM) were incubated for 10 minutes at 30 ºC, shaking at 300 rpm. The NT DNA (50 nM) was added and the reactions were incubated for another 10 minutes. The reactions were then diluted to achieve final assay conditions of 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 3 mM $MgCl_2$, 4% (v/v) glycerol, and 1 mM DTT and were again incubated for 10 min. Factors were diluted in protein dilution buffer (300 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol and 1 mM DTT) and added to Pol II ECs as serial dilutions (0-750 nM) or at a concentration of 75 or 150 nM for time course experiments. WT P-TEFb or the inactive P-TEFb mutant D149N was added (100 nM) with 1 mM ATP and incubated with Pol II and the elongation factors for 15 minutes at 30 ºC. Transcription reactions were initiated by adding GTP and CTP (10 µM) to permit elongation to position +7 (modified pause scaffold) or GTP, CTP, and UTP (10 µM) (EC* transcription scaffold). Reactions were quenched after 1-2 min (titration experiments) and after (0-5 min) for time course experiments in 2x Stop buffer (6.4 M urea, 50 mM EDTA pH 8.0, 1x TBE buffer). Samples were treated with 4 µg of proteinase K for 30 min (New England Biolabs) and were separated by denaturing gel electrophoresis (8 µL of sample applied to an 8 M urea, 1x TBE, 20% Bis-Tris acrylamide 19:1 gel run in 0.5x TBE buffer at 300V for 90 min). Products were visualized using the 6-FAM label and a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences).

Gel images were quantified using ImageJ version 1.48v[57]. The integrated density of the elongated product was measured using a box size of 0.35x0.15 cm. All integrated density values were normalized by subtracting the background integrated density from each gel. Graphs were

prepared in GraphPad Prism version 6. Each bar or point represents the mean intensity from 2-3 individual replicates. Error bars reflect the standard deviation between the replicates. Source data for gel quantification can be found in **Supplementary Table 8**.

We observe extension from a fraction of the input RNA molecules. We attribute this to inefficient EC assembly on the perfectly complementary scaffolds. It was previously shown that only 10-50 % of yeast Pol II molecules successfully assemble on perfectly complementary scaffolds[58-61] due to NT DNA displacement of the RNA primer[58]. Others have resolved the problem of displaced RNA primer by incorporating radioactive NTPs or by immobilizing NT DNA containing complexes on beads. We chose to perform RNA extension experiments in bulk with a fluorescently labelled RNA to maintain consistent Pol II concentrations across experiments and reproducibility in time course experiments.

Analytical gel filtration

ECs were formed on a bubble scaffold with the following nucleic acid sequence (EC* scaffold): template DNA 5'-GCT CCC AGC TCC CTG CTG GCT CCG AGT GGG TTC TGC CGC TCT CAA TGG-3', non-template DNA 5'- CCA TTG AGA GCG GCC CTT GTG TTC AGG AGC CAG CAG GGA GCT GGG AGC-3', and RNA 5' 5-6 FAM- UUA AGG AAU UAA GUC GUG CGU CUA AUA ACC GGA GAG GGA ACC CAC U-3'. Pol II ECs were formed as described for the transcription assays (25 pmol final Pol II, 50 pmol RNA•DNA template, 100 pmol NT DNA). Elongation factors were added in 2-3 molar excess relative to Pol II in a final buffer containing 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 3 mM $MgCl_2$. 1 mM DTT, and 4% (v/v) glycerol. Reactions that included P-TEFb (18 pmol) were supplemented with 1 mM ATP pH 7.5. Reactions were incubated for 30 minutes at 30 ºC. Samples were applied to a Superose 6 increase 3.2/300 column equilibrated in complex buffer (100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, 3 mM $MgCl_2$, and 1 mM DTT). Peak fractions were analyzed by SDS-PAGE followed by Coomassie staining.

Sample preparation for cryo-EM

Samples for cryo-EM were prepared essentially as described for analytical gel filtration runs. Final protein amounts/concentrations used for complex formation were 112 pmol Pol II, 168 pmol RNA•template DNA hybrid, 200 pmol NT DNA, 224 pmol PAF, DSIF, and SPT6, 38 pmol P-TEFb, and 313 pmol TFIIS (when included). Peak fractions corresponding to the complex were individually crosslinked with 0.1% (v/v) glutaraldehyde for 10 minutes on ice. Reactions were quenched with 8 mM aspartate and 2 mM lysine and were dialyzed against a buffer containing 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 20 mM Tris-HCl pH 7.5, 1 mM DTT, and 3 mM $MgCl_2$, in 20 kDa MWCO Slide-A-Lyzer MINI Dialysis Unit for 6 h at 4ºC. Sample from the peak (150-175 nM) was applied to R2/2 gold grids and R2/1 carbon grids (Quantifoil). The grids were glow discharged for 45 s before applying 2 µL of sample to each side of the grid (4 µL total). After incubation for 10 s and blotting for 8.5 s, the grid was vitrified by plunging it into liquid ethane with a Vitrobot Mark IV (FEI Company) operated at 4ºC and 100% humidity.

Cryo-EM data collection and data processing

Three separate data sets were collected, two of which were collected in the presence of TFIIS. TFIIS was included because of its reported role in stabilizing PAF association with Pol II[19].

Here we describe the structure lacking TFIIS. Cryo-EM data was collected on a FEI Titan Krios II transmission electron microscope operated at 300 keV. A K2 summit direct detector (Gatan) with a GIF quantum energy filter (Gatan) was operated with a slit width of 20 eV. Automated data acquisition was done with FEI EPU software at a nominal magnification of 130,000x, corresponding to a pixel size of 1.049 Å/pixel. Image stacks of 40 frames were collected over 10 s in counting mode. The dose rate was 3.4-4.7 e^- per $Å^2$ per s for a total dose of 34-47 $e^-/Å^2$. A total of 20,198 image stacks were collected.

Frames were stacked and subsequently processed with MotionCorr2[62]. CTF correction was performed with Gctf[63]. Image processing was performed with RELION 2.1[64,65]. Particles were auto-picked using projections of an initial reconstruction of PAF and DSIF bound to Pol II (data not shown) yielding 1,775,917 particle images. Particles were extracted using a box size of 360^2 pixels, normalized, and screened using iterative rounds of reference-free 2D classification resulting in 1,675,585 particles. Particles from each of the three data sets were initially processed separately. An initial reconstruction of Pol II bound to PAF and DSIF (not shown) was low pass filtered to 50 Å and used for hierarchical 3D-classification with and without image alignment and 3D-refinement. Classes showing density for TFIIS were omitted. The best resolved, non-TFIIS bound classes from each data set were selected and combined resulting in 374,964 particles (Data set 1: 101,509 particles, Data set 2 (TFIIS): 58,720 particles, Data set 3 (TFIIS): 214,745 particles). The combined particles were subjected to 3D refinement using a 50 Å low pass filtered map from a previous 3D-refinement resulting in a reconstruction with a resolution of 3.10 Å (Map A). Some domains were not well resolved in the reconstruction, so 3D classifications without image alignment with applied soft masks around the regions of interest were performed. Masks were generated in Chimera and RELION 2.1 around the DSIF NGN (Map B, 3.10 Å), KOW1 (Map C, 3.49 Å), and KOWx-4 (Map D, 3.20 Å) domains, the SPT6 core (Map E, 3.28 Å) and SPT6 tSH2 (Map F, 3.49 Å), upstream DNA (Map G, 3.10 Å), and CTR9 (Map H, 3.34 Å)[66]. Particles containing the desired density were subjected to global 3D refinement. To further improve densities, focused refinement was used for regions of CTR9, SPT6, and the RPB4-7 stalk. Focused refinements were performed by continuing global refinements after the first iteration of local searches and applying a soft mask. Masks were generated for three regions of CTR9/WDR61 comprising the N-terminal region, middle region, and C-terminal region/WDR61 (Map H). The C-terminal region/WDR61 focused refinement resulted in a final resolution of 3.59 Å with an applied *B*-factor of -145.94 $Å^2$. Focused refinement for RPB4-7 was performed using Map H resulting in a final resolution of 3.63 Å with an applied *B*-factor of -140.45 $Å^2$. Focus refinement was performed on the SPT6 core using the same mask used for classification (Map E) resulting in a resolution of 4.44 Å with an applied *B*-factor of -171.70 $Å^2$. Post-processing of refined models was performed using automatic *B*-factor determination in RELION and reported resolutions are based on the gold-standard FSC 0.143 criterion[67] (applied *B*-factors ($Å^2$): Map A: -98.65, Map B: -90.81, Map C: -90.57, Map D: -90.87, Map E: -88.56, Map F: -109.1, Map G: -94.30, Map H: -86.13). Local resolution estimates were determined using a sliding window of 30^2 voxels with an FSC cut-off of 0.3 on sharpened and non-*B*-factor sharpened maps as previously described[68].

Model building

The structure of EC* was solved by first placing the structure of a bovine EC into Map A in Chimera[66] (PDB ID: 5OIK)[4]. Adjustments were made to the protein sequence, DNA

sequence, and positioning of the upstream DNA in Coot[69]. The human RPB4-7 crystal structure (PDB ID: 2C35)[70] was placed into a focused refined version of Map H in Chimera.

Human DSIF from a previously solved cryo-EM structure (PDB ID: 5OIK)[4] was divided into 5 regions for modelling, corresponding to the SPT5 NGN and SPT4, KOW1, KOW2-3, KOWx-4 and KOW5 and placed into globally refined maps. KOW2-3 and KOW5 were placed in Map A by rigid body fitting in Chimera. The NGN domain and SPT4 were placed in Map B by rigid body fitting in Chimera. KOW1 and KOWx-4 were placed into Map C and Map D, respectively, by rigid body fitting in phenix.real_space_refine[71]. Densities for all five PAF subunits are observed. PAF was modelled using the known crystal structure for WDR61 (PDB ID: 3OW8) and homology modelling for the remaining subunits. A model for CTR9 1-798 was generated with Robetta[72] using PDB ID 4BUJ[73] as a template. Secondary structure predictions from Sable[74] and Psipred[75] were used to confirm the model. TPRs were identified and validated with TPRpred[76]. CTR9 807-892 was built *de novo* in Coot in a focused refined version of Map H. Crosslinking restraints and densities from bulky residues such as Arg and Tyr were used as sequence markers. The Robetta model of CTR9 was divided into four parts corresponding to residues 1-303, 303-677, 678-750, and 750-798 and fit into Map H or focused refined versions of Map H. WDR61 was placed into a focused refined version of Map H and is shown as an atomic model. The orientation was determined from crosslinking data (**Extended Data Fig. 6**). CTR9 and WDR61 were flexibility fit into focused refined versions of Map H using VMD and MDFF[77]. The N-terminus of CTR9 (1-300) is not well resolved, and is modelled as a backbone trace with unknown register. Clear helical densities are observed for residues 301-750 and are shown as backbone traces.

The predicted structural similarity between the triple barrel dimerization domain of TFIIF and PAF1/LEO1 was used to generate a homology model for PAF1/LEO1[78]. PAF1 (188-341) and LEO1 (367-608) were separately threaded through the RAP74 and RAP30 subunits of human TFIIF (PDB ID: 5IYC)[40], respectively, using Phyre2 (99.05/99.81% confidence of threading). These regions were chosen due to their predicted secondary structure similarity to TFIIF. The threaded model for LEO1 was truncated to residue 497. The threaded models for PAF1 and LEO1 were aligned on the TFIIF structure[40] and placed into the corresponding density in UCSF Chimera. Residues 498-529 of LEO1 were built *de novo* in Coot using crosslinking restraints and secondary structure predictions. PAF1 and LEO1 residues are modeled as backbone traces with an unknown register.

CDC73 is the least well-resolved subunit in our structure. Extensive crosslinking between CDC73 and CTR9 and noisy density near CTR9 suggests that CDC73 is highly mobile. We observe an additional helix immediately adjacent to CTR9 TPR 17 that cannot be assigned to CTR9. Crosslinking data and secondary structure predictions assigned this ‘anchor helix’ to CDC73 residues 249-262. This assignment is consistent with biochemical experiments that have collectively shown that a region of CDC73 corresponding to residues 200-337 is required for its association with PAF[79-81].

We compared our PAF structure with the published yeast Pol II-PAF-TFIIS structure[78]. The general Pol II binding surfaces of yeast and human PAF are shared, however, there are several notable differences between the structures. First, there is a substantial difference in subunit composition between yeast and human PAF[16-18,82]. Yeast PAF stably associates with Rtf1 and does not associate with Ski8 (WDR61 in human), whereas the opposite is true for human PAF. Secondly, the yeast Ctr9 construct used for cryo-EM was severely truncated and lacked

the trestle helix. The trestle helix makes additional contacts with Pol II that may confer stability. Lastly, the yeast structure was solved as a ternary structure with Pol II and TFIIS, which may have rendered the complex more flexible. DSIF, SPT6 and P-TEFb phosphorylation greatly stabilize human PAF association with Pol II (data not shown). Together, these differences may have contributed to the higher flexibility of the yeast structure and differences in subunit assignment.

To generate a model for SPT6, the human sequence for SPT6 was threaded through a crystal structure of the *S. cerevisiae* Spt6 core region (PDB ID: 3PSI)[29] with Phyre2[83]. The model generated by Phyre2 was flexibility fit into a focused refined version of Map E using VMD and MDFF[77]. The model was manually adjusted in Coot. Most domains of the central region are easily resolved with the exception of the death like domain (DLD), which is more flexible than the rest of the complex (**Extended Data Fig. 4, 5**). We modelled 3 of a total of 9 human-specific short insertions within the SPT6 core. The core was modelled as a backbone trace. The crystal structure of the human tSH2 was placed into Map F using rigid body fitting in phenix.real_space_refine. A loop corresponding to residues 1385-1395 was removed. The CTD linker was modelled using a previously solved crystal structure of yeast SPT6 tSH2 with bound CTD linker (PDB ID: 5VKO[44]). The CTD linker was mutated to the corresponding human sequence with Phyre2 and fit onto the human tSH2 by matching the cores of the yeast and human tSH2 crystal structures in Pymol (Schrödinger LLC, version 1.8.6.0).

The model was manually adjusted in Coot[69] and refined with phenix.real_space_refine against a locally filtered, non-sharpened version of Map A. The final model has 95.07% of residues in most-favored regions of the Ramachandran plot according to Molprobity[84]. The structure has a Molprobity score of 1.64. Figures were generated in Pymol (Schrödinger LLC, version 1.8.6.0) and UCSF Chimera (version 1.10.2). Surface charge was calculated with PDB2PQR[85] and visualized with APBS[86] in Pymol (Schrödinger LLC version 1.8.2.3).

Crosslinking-mass spectrometry

Samples for crosslinking and mass spectrometry analysis were essentially prepared as those used for cryo-EM. Fractions containing EC* were pooled and mixed with 2 mM of BS3 dissolved in complex buffer (No Weigh Format, ThermoFisher Scientific). The protein was incubated for 30 min at 30 ºC. The crosslinking reaction was quenched by adding 100 mM Tris-HCl pH 7.5 and 20 mM ammonium bicarbonate (final concentrations). The reaction was incubated for 15 min further at 30 ºC. The protein was precipitated with 300 mM Na•Acetate pH 5.2 and 4 volumes of acetone and incubated overnight at -20 ºC. The protein was pelleted by centrifugation, briefly dried, and resuspended in 4 M urea and 50 mM ammonium bicarbonate.

Crosslinked proteins were reduced with 10 mM DTT for one hour at room temperature (RT). Alkylation was performed by adding iodoacetamide to a final concentration of 40 mM, incubated 30 min in the dark at RT. After dilution to 1 M urea with 50 mM ammonium bicarbonate (pH 8.0), the cross-linked protein complex was digested with trypsin in a 1:50 enzyme-to-protein ratio at 37°C overnight. Peptides were acidified with trifluoroacetic acid (TFA) to a final concentration of 0.5% (v/v), desalted on MicroSpin columns (Harvard Apparatus) following manufacturer's instructions and vacuum-dried. Dried peptides were dissolved in 50 µL 30% acetonitrile (ACN)/0.1% TFA and peptide size exclusion (pSEC, Superdex Peptide 3.2/300 column on an ÄKTAmicro system, GE Healthcare) was performed

to enrich for cross-linked peptides at a flow rate of 50 µL/min. Fractions of 50 µL were collected. Fractions containing the cross-linked peptides (1–1.7 mL) were vacuum-dried and dissolved in 2% ACN/0.05% TFA (v/v) for LC-MS/MS analysis.

Cross-linked peptides derived from pSEC were analyzed as technical duplicates on an Orbitrap Fusion and Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific), respectively, coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Scientific) equipped with an in house-packed C_{18} column (ReproSil-Pur 120 C18-AQ, 1.9 µm pore size, 75 µm inner diameter, 30 cm length, Dr. Maisch GmbH). Samples were separated applying the following 58 min gradient: mobile phase A consisted of 0.1% formic acid (FA, v/v), mobile phase B of 80% ACN/0.08% FA (v/v). The gradient started at 5% B, increasing to 8% B on Fusion and 15% on Fusion Lumos, respectively, within 3 min, followed by 8–42% B and 15–46% B within 43 min accordingly, then keeping B constant at 90% for 6 min. After each gradient the column was again equilibrated to 5% B for 6 min. The flow rate was set to 300 nL/min. MS1 spectra were acquired with a resolution of 120,000 in the orbitrap (OT) covering a mass range of 380–1580 *m/z*. Injection time was set to 60 ms and automatic gain control (AGC) target to 5×10^5. Dynamic exclusion covered 10 s. Only precursors with a charge state of 3–8 were included. MS2 spectra were recorded with a resolution of 30,000 in OT, injection time was set to 128 ms, AGC target to 5×10^4 and the isolation window to 1.6 *m/z*. Fragmentation was enforced by higher-energy collisional dissociation (HCD) at 30%.

Raw files were converted to mgf format using ProteomeDiscoverer 1.4 (Thermo Scientific, signal-to-noise ratio 1.5, 1000–10000 Da precursor mass). For identification of cross-linked peptides, files were analyzed by pLink (v. 1.23), pFind group[87] using BS3 as cross-linker and trypsin as digestion enzyme with maximal two missed cleavage sites. Carbamidomethylation of cysteines was set as a fixed modification, oxidation of methionines as a variable modification. Searches were conducted in combinatorial mode with a precursor mass tolerance of 5 Da and a fragment ion mass tolerance of 20 ppm. The used database contained all proteins within the complex. FDR was set to 0.01. Results were filtered by applying a precursor mass accuracy of ±10 ppm. Spectra of both technical duplicates were combined and evaluated manually. Crosslinking figures were made with XiNet[88] and the Xlink Analyzer plugin in Chimera[89]. Distances between structured regions were calculated with Xlink Analyzer version 1.1.

Kinase assays

A modified ATP/NADH coupled ATPase assay was used to measure relative rates of ATP hydrolysis of WT P-TEFb and D149N P-TEFb[90]. P-TEFb was titrated from 0-1 µM in a final solution containing 3 mM $MgCl_2$, 0.1 mM NADH, 0.4% (w/v) pyruvate kinase/lactate dehydrogenase, 1 mM phosphoenolpyruvate, 4 µM GST- RPB1 1593-1970, 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, and 1 mM DTT. Samples (50 µL, final volume) were incubated for 2 min at 30 ºC prior to adding ATP (1 mM final concentration, pH 7.0). The decrease in absorption at 340 nm, corresponding to NADH oxidation, was measured in 384-well plates (Greiner Bio-one 384 well, clear flat bottom product number 781101) over 60 min at 30 ºC in a Tecan Infinite Pro M1000 plate reader. The rate of change in absorbance 340 nm over time was determined from the linear region of the resulting absorbance curves. The experiment was performed three times and error bars represent the standard deviation between the three measurements.

Immunoblotting experiments were performed with GST-RPB1 1593-1970 and *S. scrofa* Pol II treated with WT P-TEFb or D149N P-TEFb. GST-RPB1 1593-1970 (4 µM) was incubated with 100 nM WT P-TEFb or D149N P-TEFb in a final buffer containing 3 mM $MgCl_2$, 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, 1 mM ATP pH 7.5, and 1 mM DTT. The reactions were incubated at 30 ºC and aliquots were taken at 0, 1, 5, 10, 15, and 30 minutes after ATP addition. Reactions were quenched by mixing 3 µL of sample with 12 µL of 4x LDS loading buffer (Invitrogen). Samples (2 µL) were run on 4-12% Bis-Tris gels in MES buffer (ThermoFisher Scientific). Proteins were transferred to nitrocellulose membranes (GE Healthcare Life Sciences) and were blocked with 5% (w/v) milk powder in PBS and 0.1% TWEEN 20 for 1-3 h at room temperature. The membranes were incubated with antibodies against Ser2 (3E8), Ser5 (3E10) (1:14 dilution, gift from Dirk Eick), and the CTD (MABI0601, MBL International Corporation, 1:1000) overnight at 4ºC. Antibodies were diluted in 2.5% (w/v) milk in PBS with 0.1% TWEEN 20. Membranes were washed three times with PBS with 0.1% TWEEN 20. HRP conjugated anti Rat secondary antibody (1:5000) (Sigma-Aldrich A9037) was incubated with the membranes treated with Ser2 and Ser5 antibodies whereas the CTD membrane was treated with HRP conjugated anti mouse antibody (1:3000) (Abcam, ab5870) in PBS with 0.1% TWEEN 20 for 1 h at room temperature. SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) was used for detection.

S. scrofa Pol II (2.4 µM) was treated with 0.4 µM WT P-TEFb for 1 h at 30 ºC in a buffer containing 6 mM $MgCl_2$, 3 mM ATP pH 7.0, and 1 Roche PhosStop tablet (Sigma-Aldrich) prior to size exclusion chromatography. An aliquot of the peak fraction was used for immunoblotting. 0.5 µg of the protein was loaded on 4-12% Bis-Tris SDS-PAGE. Western blotting procedures are identical to those described for the GST-RPB1 CTD. Blots were performed using additional antibodies raised against phosphorylated Ser7 (4E12) and Tyr1 (3D12) (gift from Dirk Eick) as described above.

Phosphorylation site mapping

NELF, PAF, and SPT6 were treated with lambda protein phosphatase during purification to remove phosphorylations that were added by insect cells during protein expression. P-TEFb was incubated with individual factors or with ECs containing PAF, SPT6, and DSIF prior to size exclusion chromatography (SEC). Protein or fractions from gel filtration chromatography were applied to NuPAGE 4-12% Bis-Tris SDS-PAGE gels (ThermoFisher Scientific) and stained with InstantBlue (Expedeon). Appropriate bands were selected for MS analysis.

Phosphopeptides derived after in-gel digest of the sample were enriched as described previously[91]. Enriched phosphopeptides were analyzed on a LC-coupled Q-Exactive HF mass spectrometer (ThermoFisherScientific) under standard chromatography conditions as described[91]. The MS raw files were processed by MaxQuant[92] (version 1.5.2.8) and MS/MS spectra were searched against Uniprot human database with Andromeda[93] search engine. Allowed variable modifications included phosphorylation of serine, threonine, and tyrosine, methionine oxidation, and carbamidomethlyation of cysteine. Sites reported here were present in at least two biological replicates. 80% of the reported sites are found in Phosida[94] and PhosphoSitePlus[95] (with the following exceptions: SPT5 S148, S149, T153; SPT6 S1525; NELF-A S244 and all sites for CDC73 and PAF1). Five sites we detected were previously shown to be P-TEFb phosphorylation sites (NELF-E S181, NELF-A T277, S363, SPT5 666, 806)[10,12,42].

In vitro kinase assay and mapping of phosphorylation sites in the CTD linker

We were unable to detect P-TEFb specific phosphorylation sites in the linker from our initial MS experiments because the region is devoid of basic residues that are required for trypsination. The CTD linker has a high frequency of hydrophobic residues, which makes it amenable for chymotrypsin digestion. MBP RPB1 1488-1592 was incubated with WT P-TEFb or P-TEFb mutant D149N for 30 min at 30 ºC under the following conditions: P-TEFb 1 µM, RPB1 1488-1592 36.5µM, 3 mM ATP pH 7.0, 6mM $MgCl_2$, 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, and 1 mM DTT. The assay was independently conducted two times.

Immediately after each *in vitro* phosphorylation reaction, proteins were precipitated using chloroform-methanol extraction as described[96]. Protein precipitates were resuspended in 50 mM ammonium bicarbonate containing 1% Rapigest surfactant (Waters), reduced with DTT and alkylated with iodoacetamide. Residual iodoacetamide was quenched with DTT. The Rapigest concentration was adjusted to 0.1% with 50 mM ammonium bicarbonate and $CaCl_2$ was added to a final concentration of 2 mM. Proteins were digested at a weight ratio of 75:1 with chymotrypsin (Roche) for 12 h at 25 °C[97]. The digest was then acidified and insoluble material was removed by centrifugation. The peptide mixture was desalted using self-made StageTips[98] containing Empore C18 solid phase extraction material (3M). Each sample was analyzed in duplicate using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex UltiMate 3000 nano liquid-chromatography system (Thermo Fisher Scientific). Peptides were initially loaded onto a C18-trap column (0.3 x 5 mm, Dionex) in loading buffer (2 % ACN/0.05 % TFA) and then separated on an analytical column (self-packed with 1.9 µm ReproSil-Pur C18-AQ material, 30 cm x 75 µm, Dr. Maisch) at flow rate of 300 nL/min using a 90 min multi-step gradient (2 % ACN/0.1 % FA to 48 % ACN/0.1 % FA).

The mass spectrometer was operated in a data-dependent mode to select from a MS survey scan (range: 350-1550 m/z) the up to 20 most intense peptide precursors with charge states 2-7 for higher-energy collisional dissociation (HCD). Spectra were acquired in the Orbitrap at a resolution of 120,000 (MS1) and 30,000 (MS2) with AGC target values of $6x10^5$ (MS1) and $1.5x10^4$ (MS2) respectively. A dynamic precursor exclusion of 10 s was used. MaxQuant[92] (version 1.5.2.8) equipped with the Andromeda search engine[93] was used to analyze the raw files against a database containing the recombinant protein sequences. Chymotrypsin was selected as protease with cleavage specificity for W, Y, F, L and M. A maximum of two missed cleavage sites was allowed. Precursor and fragment ion tolerances during database search were 4.5 ppm (after internal recalibration) and 20 ppm, respectively. Cysteine carbamidomethylation was set as static modification; serine, threonine and tyrosine phosphorylation, methionine oxidation and N-terminal protein acetylation were variable modifications. Label-free quantification was enabled. False discovery rates for peptide-spectrum matches and protein identifications were set to 1 %. Phosphorylation sites were filtered for high confidence ($p > 0.75$), further examined manually and only considered relevant, when phosphorylated precursors were identified in both injection and assay replicates. Selected annotated MS2 spectra were exported as vector graphic using the MaxQuant Viewer and for better legibility labels were further modified in Adobe Illustrator CS6 (version 16.0.0).

Full sequence coverage for the RBP1-linker region and near-complete coverage (>88 %) for MBP could be obtained for all replicates. Only non-phosphorylated MBP peptides

were detected, confirming the specificity of the phosphorylation reaction. Phosphorylated CTD linker peptides were observed after incubation with WT P-TEFb (but not with D149N P-TEFb) and phosphorylation sites could be assigned with high confidence (PTM score > 0.98) to six different residues within the CTD linker (S1514, T1518, T1525, T1540, S1584, and S1590). We did not obtain direct evidence for phosphorylation of S1547, which was previously described to co-mediate SPT6 recruitment in yeast[44], although we noted a considerable intensity decrease for the unphosphorylated counterpart peptide upon incubation with WT P-TEFb.

Fluorescence anisotropy binding assays with CTD linker peptide

A 5,6-FAM labeled peptide corresponding to human RPB1 residues 1521-1552 was purchased from Caslo ApS (Lyngby, Denmark). The peptide was dissolved in 22% dimethylformamide to a concentration of 1 mM. The peptide (30 µM) was incubated with 1 µM WT P-TEFb or D149N P-TEFb in buffer containing 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, 6 mM $MgCl_2$, 3 mM ATP pH 7.0, and 1 mM DTT for 30 minutes at 30 ºC. The reaction was quenched by addition of 10 mM EDTA pH 8.0.

The tSH2 was diluted in half log steps in protein dilution buffer and mixed with diluted peptide on ice for 5 min. The reaction was diluted to achieve a final volume of 20 µL, final conditions: 10 nM peptide, 100 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, 1 mM DTT, 1 mM EDTA pH 8.0. The reactions were incubated at room temperature for 20 minutes in the dark. 17 µL aliquots were removed and transferred to a Greiner 384 well Black Flat Bottom Small volume plate. Fluorescence anisotropy was measured at room temperature with an Infinite M1000 Pro plate reader (Tecan) with an excitation wavelength of 470 nm ±5 nm, an emission wavelength of 518 nm ± 5 nm, and gain of 75. The experiment was performed in triplicate and analyzed with GraphPad Prism Version 6. Binding curves were fit using a single site quadratic binding equation as described[52].

Fluorescence anisotropy binding assays with SPT6 and nucleic acids

SPT6 was dialyzed overnight against SPT6 FA buffer (150 mM NaCl, 20 mM Na•HEPES pH 7.4, 10% (v/v) glycerol, and 1 mM DTT) in a 20 kDa MWCO Slide-A-Lyzer MINI Dialysis Unit. The protein was then used directly for anisotropy measurements. 25-mer 5'-/6-FAM labeled DNA and RNA oligonucleotides bearing the sequence AAG GGG AGC GGG GGA GGA TAA TAG G (T substituted with U in RNA sequence) were obtained from IDT and dissolved in water. SPT6 was serially diluted in half log steps in SPT6 FA buffer. Nucleic acids (2.2 µL, 10 nM final concentration) and SPT6 (4.4 µL, 0-9.5 µM final concentration) were mixed on ice and incubated for 5 minutes. The assay was brought up to a final volume of 22 µL and incubated at RT in the dark for 20 minutes (final conditions: 30 mM NaCl, 3 mM $MgCl_2$, 20 mM Na•HEPES pH 7.4, 50 µg/mL BSA, 5 µg/mL yeast tRNA, and 1 mM DTT). 18 µL of each solution was transferred to a Greiner 384 Flat Bottom Black Small volume plate. Fluorescence anisotropy was measured and analyzed as above but with a gain of 70 and an emission wavelength of 518 nm ± 20 nm.

Pull-down experiments

MBP-RPB1 1488-1970 and MBP RPB1 1593-1970 (5µM) were incubated with 0.4 µM WT or P-TEFb mutant D149N in pulldown buffer (100 mM NaCl, 20 mM Na•HEPES pH 7.4, 4% (v/v) glycerol, 1 mM DTT, and 3 mM $MgCl_2$) for 30 min at 30 ºC. The CTD constructs were

incubated with amylose beads for 10 minutes further. The beads were washed three times with pulldown buffer to remove P-TEFb and ATP. FL SPT6 or SPT6ΔtSH2 were then added at a final concentration of 7.5 µM. The reactions were incubated at 30 ºC for 15 min and washed three times with pulldown buffer. The MBP tag was eluted from the beads by applying pulldown buffer with 116 mM maltose to the beads. 20 µL of the eluted sample was applied to a 4-12% SDS-PAGE and stained with Coomassie blue.

Crystal structure determination of SPT6 tSH2 domain

Frozen tSH2 protein was thawed and applied to a Superdex S200 increase 10/300 column equilibrated in SE buffer. Peak fractions were pooled, concentrated, and dialyzed into a buffer containing 100 mM NaCl, 20 mM HEPES pH 7.4, and 1 mM TCEP pH 7.0 for 16 h at 4°C. Initial crystals of tSH2 were obtained by hanging drop vapor diffusion crystallization at 293 K by mixing 1 µL of protein solution and 1 µL of reservoir solution containing 100 mM Bis-Tris pH 5.5, 200 mM $MgCl_2$, and 21-25 % (v/v) PEG 3350. Larger crystals were obtained by micro-seeding. A drop with initial crystal hits was transferred to a micro centrifuge tube containing one glass bead ('Bead for seeds', Jena Biosciences) and 50 µL of solution (100 mM Bis-Tris pH 5.5, 200 mM MgCl2, and 25 % (v/v) PEG 3350). The seed stock was vortexed extensively and diluted 1:1000. Crystals used for data collection were grown using hanging drop vapor diffusion technique with a 0.5:0.5:1 ratio of protein solution, seed solution, and reservoir solution (100 mM Bis-Tris pH 5.5, 200 mM MgCl2, and 25 % (v/v) PEG 3350) at 293 K. For harvesting, crystals were exchanged into cryo-protectant solution (100 mM Bis-Tris pH 5.5, 200 mM MgCl2, 25 % (v/v) glycerol and 25 % (v/v) PEG 3350), and flash-frozen in liquid nitrogen.

Diffraction data were collected at beamline PXII of the Swiss Light Source at the Paul Scherrer Institute[99]. The native dataset was collected at a wavelength of 0.999 Å. Diffraction images were processed with XDS[100]. The structure was solved with molecular replacement in Phaser[101] using a poly-alanine model of the *S. cerevisiae* Spt6 tSH2 domain (PDB ID 3PSJ). Refinement was performed using Phenix.Refine[102] applying riding hydrogens. The human tSH2 structure is nearly identical to the yeast structure except for an N-terminal loop that is involved in crystal packing and adopts an alternative conformation (**Extended Data Fig. 8b**). The final model was refined to an R_{work}/R_{free} of 19.0%/22.3%. Molprobity[84] analysis showed that 98.61% of the residues reside in the most-favored regions of the Ramachandran plot, and 1.39% fell in allowed regions. None of the residues fell in disallowed regions.

Data availability statement

The electron density reconstructions and final EC* model were deposited with the Electron Microscopy Data Base (EMDB) under accession codes EMD-0030 to EMD-0037, and with the Protein Data Bank (PDB) accession code 6GMH. The tSH2 domain model was deposited with the PDB accession code 6GME. Source data for Figures 1a and 5a, Extended Data Figs. 1a, e-k, m, 2a-j, i-e are found in Supplementary Figures 1, 2 and Supplementary Table 8.

Extended Data Figure Legends

Extended Data Figure 1 | Protein preparation and phosphorylation activity of P-TEFb and RNA extension assays.

a. Quality of purified proteins used in this study (0.9 µg protein per lane). All proteins were purified at least two times. Representative gel was run two times. Asterisk denotes SPT5 N-terminal degradation product.

b. Nucleic acid scaffold used for RNA extension assays, 'modified pause scaffold'.

c. Nucleic acid scaffold used for analytical gel filtration and for cryo-EM analysis, 'EC* scaffold'.

d. P-TEFb kinase activity using a coupled ATP/NADH assay. Bars correspond to the absolute change in 340 nm absorbance as a function of time. Error bars represent the standard deviation between 3 individual experiments. Each bar corresponds to the mean of 3 individual experiments.

e. P-TEFb (100 nM) was incubated with GST-RPB1 CTD for different amounts of time. Membranes were incubated with antibodies that recognize phospho-Ser2 (3E10), phospho-Ser5 (3E8), or the CTD (MABI0601). Similar experiments were performed at least 3 times for the WT enzyme. The Western blot for the D149N mutant was performed once.

f. Pol II (75 nM) was incubated with WT P-TEFb or D149N P-TEFb (100 nM) and DSIF and NELF (150 nM). Reactions were quenched at various time points after GTP and CTP addition (10 µM). The experiment was performed 3 times.

g. Quantification of extended RNA products in panel f. Points are the mean of 3 individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

h. Pol II (75 nM) was incubated with the modified pause scaffold (50 nM) (**Extended Data Fig. 1b**), wild-type (WT) P-TEFb or inactive P-TEFb (D149N) (100 nM) and ATP (1 mM) (all lanes), and DSIF and NELF (150 nM). PAF was titrated into the reactions. The reactions were quenched 2 min after CTP and GTP addition (10 µM). Positions for a consensus pausing site (+2) and extended RNA (+7) are marked. RNA extension is incomplete because only a fraction of Pol II molecules assemble on the scaffold. The experiment was performed 2 times.

i. Quantification of extended RNA products in panel h. Points are the mean of 2 individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

j. Pol II (75 nM) was incubated with DSIF and NELF (150 nM) and WT P-TEFb or D149N P-TEFb (100 nM). PAF, and SPT6 were titrated into the reactions. Reactions were quenched 1 min after GTP and CTP addition (10 µM). The experiment was performed 3 times.

k. Quantification of extended RNA products in panel l. Points are the mean of 3 individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

l. Nucleic acid scaffold used for RNA extension assays, 'EC* transcription scaffold'.

m. RNA extension assays performed on the EC* transcription scaffold (50 nM). Pol II (75 nM) was incubated with elongation factors (7.5-750 nM) (DSIF, PAF, SPT6), active or inactive (D149N) P-TEFb (100 nM), and 1 mM ATP for 15 min. Reactions were quenched 1 min after GTP, CTP, and UTP addition. Experiments were performed 3 times. A large fraction of RNA primer remains due to incomplete EC assembly (see Methods for more details).

Extended Data Figure 2 | P-TEFb activity enables PAF to displace NELF and EC* formation.

a. Quantification of extended RNA products in **Figure 1a**. Points are the mean of three individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

b. PAF, DSIF, and SPT6 (23.7-750 nM) were titrated against Pol II (75 nM) and WT or D149N P-TEFb. Reactions were quenched 1 min after GTP and CTP addition (10 μM). The experiment was performed 3 times.

c. Quantification of extended RNA products in panel b. Points are the mean of 3 individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

d. Elongation factors (75 nM) were incubated with P-TEFb (100 nM) and ATP (1 mM). Reactions were quenched after 0.6 min after GTP and CTP addition (10 μM). The experiment was performed 3 times.

e. Quantification of extended RNA products in panel d. Points are the mean of 3 individual experiments and error bars represent the standard deviation between replicates. Source data: Supplementary Table 8.

f-j. SDS-PAGE analysis of size exclusion chromatography (SEC) fractions. The Pol II-EC was formed on the EC* scaffold. All experiments were performed at least 2 times. **f.** DSIF, **g.** PAF, **h.** SPT6 **i.** Pol II EC, DSIF, PAF, SPT6 **j**. Pol II EC, DSIF, PAF, SPT6, P-TEFb and ATP. Fractions used for cryo-EM are indicated.

k. NELF is released from Pol II when PAF, WT P-TEFb and ATP are present as assessed by SEC. Curves from the PEC and the PEC plus PAF are shown as a reference. The Pol II-EC was formed on the EC* scaffold. Each experiment was performed at least 2 times.

l. SDS-PAGE analysis of SEC fractions from PEC formation with PAF, P-TEFb, and ATP. The experiment was performed 2 times.

m. SDS-PAGE analysis of SEC fractions from PEC formation with PAF. The experiment was performed 2 times.

Extended Data Figure 3 | Cryo-EM data collection and processing.

a. Representative micrograph of EC* shown at a defocus of -2.5 μm. Representative of 20198 replicates.

b. Representative 2D classes of EC* particles.

c. Classification tree for data processing.

Extended Data Figure 4 | Quality and resolution of cryo-EM data.

a. Estimate of average resolution. Lines indicate the Fourier shell correlation (FSC) between the half maps of the reconstruction.

b. Angular distribution of particles from overall refinement. Red dots indicate the presence of at least one particle image within ±1°.

c. EC* reconstructions as colored by local resolution. The overall reconstruction is shown with *B*-factor sharpened and non-sharpened maps. The globally refined maps E and I are shown as non-*B* factor sharpened maps.

Extended Data Figure 5 | Fits of EC* model in representative densities.

a. EC* fit in electron density (map A) contoured to 12 Å. Black ovals indicate regions where electron density was weak. Map F and map H are shown to indicate the improvement after focused classification and refinement.

b.-f. Electron density for various elements of EC* shown as grey mesh. **a.** CTR9 vertex and TPRs 18-19, map H **c.** CTR9 trestle helix, map H **d.** WDR61, map H **e.** C-terminus of LEO1 and upstream DNA, map G **f.** tSH2 crystal structure, map F **g.** core of SPT6, map E.

Extended Data Figure 6 | Crosslinking-mass spectrometry analysis.

a. Overview of crosslinks obtained with BS3 in EC*. Connecting line thickness signifies the number of crosslinks obtained between subunits.

b. Histogram of unique crosslinks and distances between Cα pairs that were mapped onto our structure. A dotted black line marks the 30 Å distance cutoff for BS3. Venn diagram compares unique crosslinks between two biological replicates.

c.-g. Crosslinks mapped onto final model. Residues involved in crosslinks are shown as spheres. Coloured rods connecting residues signify permitted (blue) or non-permitted (red) crosslinking distances. **c.** WDR61 CTR9. **d.** DSIF KOW1 and KOWx-4 domains and SPT6. **e.** A C-terminal extension of LEO1, NGN and KOW1 domain of SPT5 and RPB2. **f.** SPT6 and Pol II. **g.** CTR9 and Pol II.

Extended Data Figure 7 | SPT6 tSH2 crystal structure and associated EM densities.

a. Cartoon model of human tSH2 crystal structure shown in two different views.

b. Human SPT6 tSH2 is structurally similar to previously obtained SPT6 tSH2 structures from *S. cerevisiae*[29] (PDB ID: 3PSJ) (hot pink), *C. glabrata*[102] (PDB ID: 2XPI) (grey), and *A locustae*[103] (3PJP) (peach).

c. Surface charge representation of the human SPT6 tSH2.

d. Representative electron density from tSH2 crystal structure. $2F_o$-F_c maps contoured at 2σ are shown for several regions of the tSH2 crystal structure.

e. 15 Å low pass filtered map E. C-terminal density of SPT6 extends to CTR9.

f. Alternative view to **Fig. 5b.** Two P-TEFb phosphorylation sites are demarcated (T1525, T1540). The T1540 site was not observed in the yeast linker that was used for crystallization. The CTD linker is modelled.

Extended Data Figure 8 | Features of EC* and comparisons to other structures.

a. WDR61 is anchored by the vertex and TPRs 13, 18, and 19.

b-c. SPT6 binds to the C1-C3 sheets of RPB7 **a.** Surface representation of SPT6 association with RPB4-7 stalk (RPB4 red, RPB7 cyan). **b.** Book view of **a**. RPB4-7 and SPT6 are coloured according to surface charge (blue, positive; red, negative).

d. Comparison of initiation and elongation factor binding sites. The yeast preinitiation complex bound to core mediator (PIC-cMed)[104] (PDB ID: 5OQM) was aligned with the EC* Pol II core.

e. Model for RNA, CTD, and CTR paths extending from EC*.

Extended Data Figure 9 | EC* is highly phosphorylated.

a. *S. scrofa* Pol II CTD P-TEFb phosphorylations assessed by Western blot using antibodies raised against phospho-Tyr1 (3D12), phospho-Ser2 (3E10), phospho-Ser5 (3E8), and phospho-

Ser7 (4E12) or the RPB1 body (F12) or CTD (MABI0601). Experiments with the phospho-antibodies were performed 2 times. The RPB1 body and CTD antibody experiments were performed 1 time.
b. Phosphorylation sites determined by mass spectrometry. The experiment was performed 2 or more times with each protein. Reported sites are found in at least 2 independent replicates.
c-e. Representative MS spectra.
f. Phosphorylations map to flexible regions of EC*. Spheres and dotted lines represent 2 phosphorylations and flexible regions, respectively.

Extended Data Figure 10 | P-TEFb phosphorylates CTD linker and SPT6 tSH2 required for association with EC*.
a. Sequence alignment of the CTD linker from various species generated in Mafft[105] and visualized in Jalview[106] (*S. cerevisiae*, *S. pombe*, *C. elegans*, *D. melanogaster*, *D. rerio*, *X. laevis*, *M. musculus*, and *H. sapiens*). Blue columns represent regions sharing sequence identity. Orange boxes represent phosphorylation sites reported here or those obtained previously in yeast[44].
b.-f. Representative MS2 spectra of P-TEFb phosphorylated CTD linker peptides. Spectra are representative of 2 biological replicates. RPB1 residues serine 1514 (**b**; precursor m/z 759.804, z = +2, corresponding RPB1 residues 1503-1517), threonine 1518 (**c**; precursor m/z 548.730, z = +2, RPB1 residues 1511-1520), threonine 1525 (**d**; precursor m/z 608.245, z = +2, RPB1 residues 1521-1531), threonine 1540 (**e**; precursor m/z 701.789, z = +2, RPB1 residues 1532-1546) as well as serine 1584 and serine 1590 (**f**; precursor m/z 580.708, z = +2, RPB1 residues 1582-1592) are phosphorylated by P-TEFb *in vitro*. The sequence of the corresponding phosphorylated chymotryptic precursor peptide is shown with all identified b-ions (blue) and y-ions (red). Asterisks indicate neutral loss of phosphoric acid (H_3PO_4, Δ97.98 Da), which is commonly observed for phosphoserine- and phosphothreonine-containing peptides upon HCD fragmentation. Additionally, peaks corresponding to neutral loss of ammonia (NH_3, Δ17.03 Da) or water (H_2O, Δ18.01 Da) are labeled in orange.
g. Pulldowns performed with FL SPT6 and SPT6 ΔtSH2 and MBP-RPB1 CTD constructs in the presence of WT or D149N P-TEFb. The gel is representative of 2 independent experiments.
h. Quality of purified SPT6 ΔtSH2 (1-1297)(0.9 μg).
i. Time course transcription assay with SPT6 ΔtSH2, PAF, DSIF (75 nM) and WT or D149N P-TEFb. The gel is representative of 3 independent experiments.
j. Size exclusion chromatography experiment as performed in Extended Data Fig. 1. SPT6 ΔtSH2 does not stably associate with EC*. The experiment was performed twice.
k. Nucleic acid association with full-length SPT6. Binding to single stranded DNA (cyan), double stranded DNA (blue), or RNA (red) was assessed by fluorescence anisotropy. Error bars reflect the standard deviation between 3 experimental replicates. Points represent the mean of 3 experimental replicates.

EXTENDED DATA TABLES

Extended Data Table 1 | Components of EC*.

List of all protein and nucleic acid components of EC*.

Extended Data Table 2 | Cryo-EM data collection, refinement, and validation statistics.

Extended Data Table 3 | X-ray data collection and refinement statistics SPT6 tSH2.

METHODS AND EXTENDED DATA REFERENCES

51. Gradia, S. D. *et al.* MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein Complexes. *Methods Enzymol* **592,** 1–26 (2017).
52. Vos, S. M. *et al.* Architecture and RNA binding of the human negative elongation factor. *eLife* **5,** e14981 (2016).
53. Kapust, R. B. & Waugh, D. S. Escherichia colimaltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8,** 1668–1674 (1999).
54. Sydow, J. F. *et al.* Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Mol Cell* **34,** 710–721 (2009).
55. Sidorenkov, I., Komissarova, N. & Kashlev, M. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell* **2,** 55–64 (1998).
56. Larson, M. H. *et al.* A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344,** 1042–1047 (2014).
57. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9,** 671–675 (2012).
58. Komissarova, N., Kireeva, M. L., Becker, J., Sidorenkov, I. & Kashlev, M. Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol* **371,** 233–251 (2003).
59. Xu, J. *et al.* Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **13,** 5361 (2017).
60. Kireeva, M. L., Komissarova, N. & Kashlev, M. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J Mol Biol* **299,** 325–335 (2000).
61. Kireeva, M. L., Komissarova, N., Waugh, D. S. & Kashlev, M. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* **275,** 6530–6536 (2000).
62. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14,** 331–332 (2017).
63. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193,** 1–12 (2016).
64. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5,** e18722 (2016).
65. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180,** 519–530 (2012).
66. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).
67. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333,** 721–745 (2003).
68. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518,** 376–380 (2015).
69. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
70. 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,** 6435–6444 (2005).

71. Afonine, P. V., Headd, J. J., Terwilliger, T. & Adams, P. D. New tool: phenix.real_space_refine. *Computational Crystallography Newsletter* **4,** 43–44 (2013).
72. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32,** W526–W531 (2004).
73. Halbach, F., Reichelt, P., Rode, M. & Conti, E. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154,** 814–826 (2013).
74. Adamczak, R., Porollo, A. & Meller, J. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* **59,** 467–475 (2005).
75. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41,** W349–57 (2013).
76. Karpenahalli, M. R., Lupas, A. N. & Söding, J. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8,** 2 (2007).
77. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16,** 673–683 (2008).
78. Xu, Y. *et al.* Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat Comms* **8,** 15741 (2017).
79. Rozenblatt-Rosen, O. *et al.* The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Mol Cell Biol* **25,** 612–620 (2005).
80. Yart, A. *et al.* The HRPT2 tumor suppressor gene product parafibromin associates with human PAF1 and RNA polymerase II. *Mol Cell Biol* **25,** 5052–5060 (2005).
81. Amrich, C. G. *et al.* Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin. *J Biol Chem* **287,** 10863–10875 (2012).
82. Cao, Q.-F. *et al.* Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. *Mol Cell Biol* **35,** 3459–3470 (2015).
83. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10,** 845–858 (2015).
84. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 (2010).
85. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **35,** W522–5 (2007).
86. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98,** 10037–10041 (2001).
87. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat Methods* **9,** 904–906 (2012).
88. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics* **14,** 1137–1147 (2015).
89. Kosinski, J. *et al.* Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol* **189,** 177–183 (2015).
90. Kiianitsa, K., Solinger, J. A. & Heyer, W.-D. NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal Biochem* **321,** 266–271 (2003).

91. Oellerich, T. *et al.* SLP-65 phosphorylation dynamics reveals a functional basis for signal integration by receptor-proximal adaptor proteins. *Mol Cell Proteomics* **8,** 1738–1750 (2009).
92. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26,** 1367–1372 (2008).
93. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805 (2011).
94. Gnad, F. *et al.* PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8,** R250 (2007).
95. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43,** D512–20 (2015).
96. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138,** 141–143 (1984).
97. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **11,** 993–1006 (2016).
98. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2,** 1896–1906 (2007).
99. Fuchs, M. R. *et al.* D3, the new diffractometer for the macromolecular crystallography beamlines of the Swiss Light Source. *J Synchrotron Rad* **21,** 340–351 (2014).
100. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).
101. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).
102. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
103. Sun, M., Larivière, L., Dengl, S., Mayer, A. & Cramer, P. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J Biol Chem* **285,** 41597–41603 (2010).
104. Diebold, M.-L. *et al.* Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. *J Biol Chem* **285,** 38389–38398 (2010).
105. Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551,** 204–209 (2017).
106. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30,** 772–780 (2013).
107. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).