

Minireview

Irina P. Lushpinskaia, David Flores-Solis and Markus Zweckstetter*

Structure and phase separation of the C-terminal domain of RNA polymerase II

<https://doi.org/10.1515/hsz-2023-0136>

Received February 14, 2023; accepted June 1, 2023;
published online June 20, 2023

Abstract: The repetitive heptads in the C-terminal domain (CTD) of RPB1, the largest subunit of RNA Polymerase II (Pol II), play a critical role in the regulation of Pol II-based transcription. Recent findings on the structure of the CTD in the pre-initiation complex determined by cryo-EM and the novel phase separation properties of key transcription components offers an expanded mechanistic interpretation of the spatiotemporal distribution of Pol II during transcription. Current experimental evidence further suggests an exquisite balance between CTD's local structure and an array of multivalent interactions that drive phase separation of Pol II and thus shape its transcriptional activity.

Keywords: CTD; liquid–liquid phase separation; RNA Pol II; transcription

1 Introduction

RNA Polymerase II (Pol II) executes the synthesis of DNA into RNA in eukaryotic cells. The activity of Pol II during transcription is intimately related with the post-translational modifications (PTMs) of the C-terminal domain (CTD) of its largest protein unit, the RNA polymerase subunit B1 (RPB1) (Figure 1A) (Eick and Geyer 2013; Zaborowska et al. 2016). In different organisms the length of the CTD varies from 52 repeats in humans to 26 repeats in *Saccharomyces cerevisiae*. The N-terminal half of the human CTD closely resembles the sequence from yeast CTD while the C-terminal half is less

conserved. However, while the CTD is generally well conserved, the associated phase separation is affected by CTD length and composition (Lu et al. 2019), potentially due to changes in multivalent interactions (Boehning et al. 2018). Increasing evidence further suggests that the role of the CTD in transcription is determined by the well-established patterns of PTMs, coining the “CTD code” (Eick and Geyer 2013), in combination with the ability of this intrinsically disordered region to undergo phase separation and drive the clustering of Pol II. Here we will review the findings reported for the CTD during the last four years focusing on the structural properties of the CTD inside the pre-initiation complex and the contribution of CTD phase separation to eukaryotic gene transcription.

2 CTD sequence and PTMs

The CTD is a low complexity domain comprising multiple repetitions of the consensus heptad $Y_1S_2P_3T_4S_5P_6S_7$ (Figure 1B) (Eick and Geyer 2013; Zaborowska et al. 2016). The high proline content is linked to the presence of serine/proline-motifs, which are a preferred substrate for cyclin-dependent kinases (CDKs) which phosphorylate serine or threonine residues. The CTD structure changes in response to phosphorylation or other PTMs that occur in the CTD (Figure 1C) (Zhang and Corden 1991). Pol II phosphorylated at Ser5 in the consensus CTD is more concentrated near the promoter, while phosphorylation at Ser2 is connected to elongation, and Ser7 in the consensus CTD repeat might be phosphorylated or glycosylated. Ser5 and Thr4 of the heptad repeat may also be glycosylated, while Lys7 or Arg7 in non-consensus repeats are sites for acetylation, methylation, and ubiquitination. These specific patterns of PTMs have led to the creation of the term “CTD code”.

Mutagenesis studies on CTD demonstrated the importance of its amino acid composition (West and Corden 1995). More details of particular amino acids and positions in the heptad were demonstrated, recently. For instance, Tyr1, Ser2 and Thr4 are required for efficient termination and for Pol II pausing at the 5' end of genes (Collin et al. 2019). Tyr1 is important for substrate recognition by CDK7 (Ramani

*Corresponding author: Markus Zweckstetter, German Center for Neurodegenerative Diseases (DZNE), Von-Siebold Straße 3A, D-35075 Göttingen, Germany; and Department of NMR-based Structural Biology, Max Planck Institute for Multidisciplinary Sciences, Am Faßberg 11, D-37077 Göttingen, Germany, E-mail: Markus.Zweckstetter@dzne.de. <https://orcid.org/0000-0002-2536-6581>

Irina P. Lushpinskaia and David Flores-Solis, German Center for Neurodegenerative Diseases (DZNE), Von-Siebold Straße 3A, D-35075 Göttingen, Germany

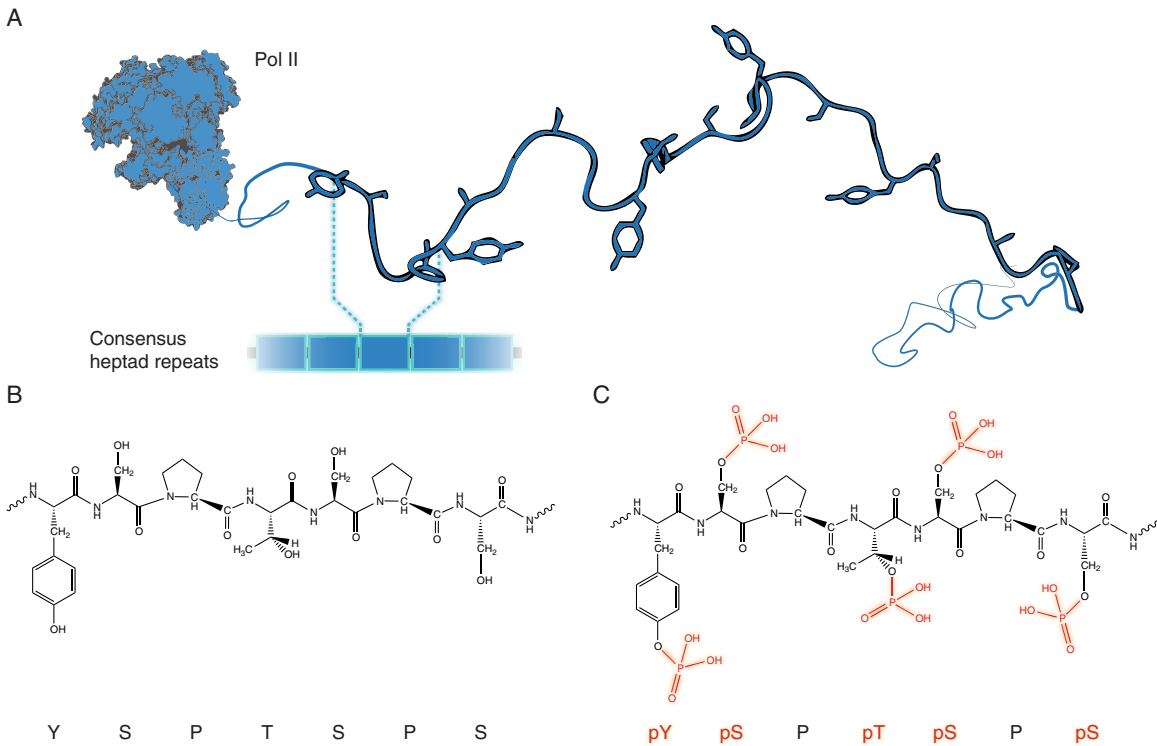


Figure 1: The CTD of RNA polymerase II. (A) Schematic representation of RNA Pol II and its low complexity C-terminal domain comprising heptad repeats. (B) Stereochemical representation of a YSPTSPS repeat of the CTD of RNA Pol II. (C) Possible phosphorylation sites in the consensus heptad repeat.

et al. 2020) and phosphorylation of Ser2 by P-TEFb (Mayfield et al. 2019). Phosphorylation of Tyr1 by c-Abl also promotes synthesis of strand-specific, damage-responsive transcripts, which force the formation of double-stranded RNA and the recruitment of p53-binding protein 1 and Mediator of DNA damage checkpoint 1 to endogenous double strand breaks (Burger et al. 2019). The high density of serine/proline-motifs in the CTD encodes yet another level of control by prolyl isomerases, such as Ess1 in *S. cerevisiae* and its human homolog Pin1. Isomerization is part of the “CTD-code”, which also regulates the recruitment of proteins required for transcription and RNA processing (Namitz et al. 2021), or to overcome energetic barriers to metabolic activities by concentrating enzymes and substrates via CTD phase separation (Palumbo et al. 2022).

3 CTD structure

The CTD lacks stable secondary and tertiary structure. However, this does not exclude the presence of local structure, i.e. structural bias, determined by the CTD amino acid sequence in particular the consensus $Y_1S_2P_3T_4S_5P_6S_7$ repeats (Jasnovidova and Stefl 2013). To gain insight into potential structural biases in the CTD, studies initially concentrated on

small CTD peptides, which were sometimes circularized and frequently treated in solvents that promote turns at low pH to maintain structure (Cagas and Corden 1995; Kumaki et al. 2001). The structural features of non-repeating sections of the *Drosophila melanogaster* CTD have also been examined (Gibbs et al. 2017; Lu et al. 2019; Portz et al. 2017). In addition, the structures of short CTD fragments in association with CTD-binding partners were characterized (Noble et al. 2005), including the mRNA anti-terminator protein hSCAF4-CID binding to S2/S5-phosphorylated CTD (Zhou et al. 2022), the prolyl isomerase Ess1 in complex with CTD peptide (Namitz et al. 2021), and a CTD peptide comprising five consensus repeats bound to CDK kinases (Ramani et al. 2020).

Investigations of the unmodified distal portion of the human CTD (the C-terminal 26 heptad repeats) showed that lysine residues at position 7, which depart from the canonical $Y_1S_2P_3T_4S_5P_6S_7$ sequence, participate in electrostatic interactions with aspartic and glutamic acids of the three proteins fused in sarcoma (FUS), Ewing sarcoma (EWS) and TATA-box binding associated factor (TAF15) of the FET-family, and mediate the interaction with TAF15 fibril assemblies (Burke et al. 2015; Janke et al. 2018; Murthy et al. 2021). These findings proposed a heterotypic mode of electrostatic interactions between charged proteins and the CTD.

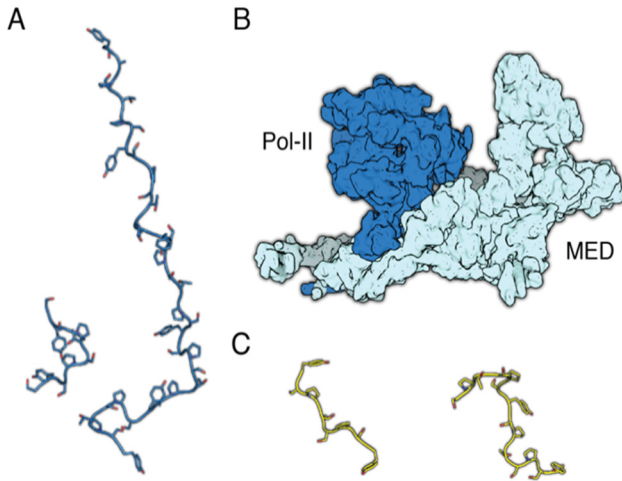


Figure 2: CTD structure bound to the PIC. (A) Two structured segments of the CTD in complex with the pre-initiation complex (PDB ID: 7ENC) determined by cryo-EM. (B) RNA Pol II (blue) and Mediator (cyan) complex, extracted from the full pre-initiation complex determined by Chen et al. (2021) (PDB ID: 7ENC). Some molecules are omitted for clarity. (C) Two segments of CTD local structure bound to the pre-initiation complex resolved by Abdella et al. (2021) (PDB ID: 7LBM).

Recently, cryo-EM studies have revealed the structure of the human Mediator at near-atomic detail (Abdella et al. 2021; Aibara et al. 2021; Chen et al. 2022, 2021; Rengachari et al. 2021), revealing that the preinitiation complex/Mediator complex forms a “Head-Middle sandwich” that holds two CTD segments close to CDK7 for phosphorylation (Abdella et al. 2021; Chen et al. 2022, 2021) (Figure 2). The two U-shaped CTD segments are located between the Middle’s Knob and the Head’s HB1. Although the Head and the Middle do not have direct interactions, they are connected by two CTD repeats. Tyr1 from one CTD repeat makes contacts with residues MED8-Pro83, Knob-Phe96, Knob-Cys93, MED6-Arg118 and MED17-Val111. A second Tyr1 residue of the CTD inserts into the hydrophobic pocket formed by the two Mediator residues MED6-Tyr106 and MED8-Pro106. These CTD-Mediator interactions are further reinforced by other contacts (Chen et al. 2021).

4 Phase separation in transcription

Proteins and nucleic acids can be organized into cellular compartments through the formation of biomolecular condensates (Rawat et al. 2021; Shao et al. 2022). While the exact mechanisms how transcription foci and other biomolecular condensates form are still a matter of debate, transient multivalent interactions between intrinsically disordered protein regions are believed to play a key role

(Chong et al. 2018; Martin et al. 2020). Such multivalent interactions may cause liquid-liquid phase separation of proteins and nucleic acids into highly dynamic cellular compartments (Guo et al. 2019, 2020). For example, the disordered regions of transcription factors can phase separate together with the Mediator coactivator into liquid-like droplets *in vitro* (Boija et al. 2018). The CTD is also able to phase separate into liquid-like droplets *in vitro* with the aid of crowding agents (Boehning et al. 2018) (Figure 3A), in agreement with the formation of dynamic Pol II clusters *in vivo* (Cho et al. 2018; Lu et al. 2019). However, the ability of the CTD to form liquid-like droplets by itself without crowding agents in near physiological conditions requires further studies.

The number of heptad repeats in the CTD of RPB1 impacts the propensity for liquid-liquid phase separation (Boehning et al. 2018; Lu et al. 2019) and the length also correlates with gene density (Quintero-Cadena et al. 2020). Besides the number of CTD repeats, sequence variations in the heptad repeats are important. For instance, the distal part of the human CTD enhances CTD binding to the RNA-binding protein FUS, mediated by the lysine residues located in the seventh position of the distal heptads (Murthy et al. 2021). The lysine residues in the distal part of the CTD enhance the network of multivalent interactions and thus increase CTD’s capacity to phase separate and organize into biomolecular condensates (Burke et al. 2015, Janke et al. 2018, Murthy et al. 2021). This suggests that it would be valuable to examine the role of each heptad amino acid in forming the molecular CTD structure required for liquid-liquid phase separation and gather additional information on how this is influenced by PTMs in the CTD.

Post-translational modifications play a crucial role in the catalytic activity of Pol II and are dynamically affecting it. Phosphorylation of CTD is linked to the different stages of transcription, although RNA synthesis occurs without the CTD (Zehring et al. 1988). The presence of droplets formed by the hyperphosphorylated CTD suggests that liquid-liquid phase separation may have a broad impact on various stages of transcription (Lu et al. 2018) (Figure 3B–D). For instance, phosphorylation of serine 5 by the human TFIIF subcomplex containing the CDK7 kinase dissolved CTD droplets *in vitro* (Boehning et al. 2018). In addition, phosphorylation of the CTD modifies the molecular interactions between the CTD, the Mediator complex and serine/arginine-rich domains triggering a shift in CTD partitioning from transcription to splicing condensates (Guo et al. 2019). Supporting an important role of liquid-liquid phase separation in transcription, Pol II and the Mediator complex colocalize in nuclear condensates (Cho et al. 2018; Guo et al. 2019). Pol II also partitions into the phase-separated

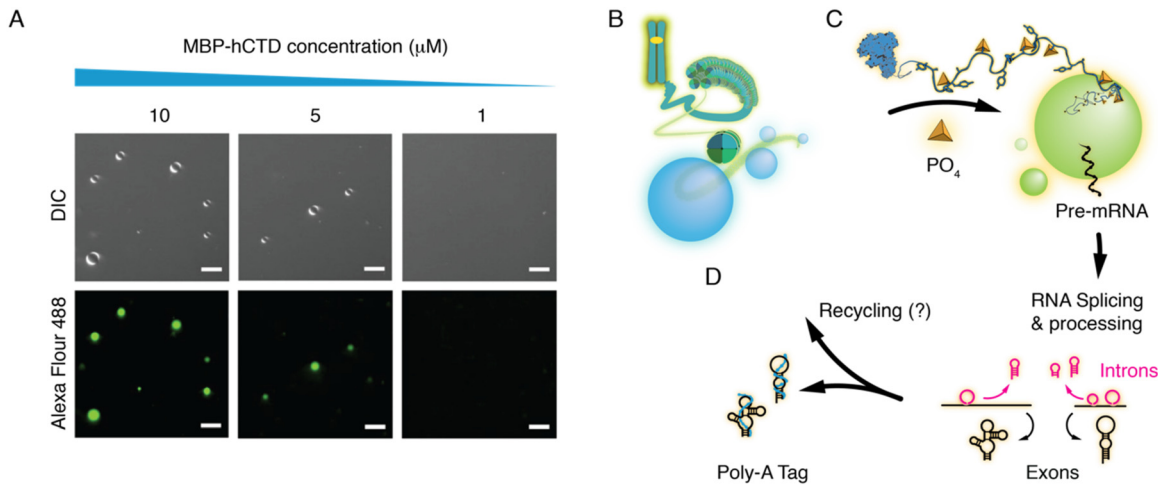


Figure 3: CTD phase separation. (A) Differential interference contrast (DIC) and fluorescence microscopy revealing concentration-dependent formation of liquid droplets of MBP-hCTD in the presence of 16 % dextran. Reproduced from Boehning et al. (2018). (B) Transcription factors, Mediator and RNA Pol II can undergo LLPS creating “transcriptional hubs”. The role of “transcriptional hubs” for transcription initiation (blue droplets) is currently being studied. (C) Site-specific phosphorylation of the CTD causes the transfer of RNA Pol II from initiation condensates to elongation condensates activating the synthesis of pre-mRNAs (green droplets). (D) The role of the CTD in RNA processing (splicing, capping, polyadenylation, etc.) is enigmatic. PTMs in the CTD are believed to regulate the transcriptional activity of RNA Pol II.

super elongation complex through heterotypic phase separation (Guo et al. 2020), and in presence of the negative elongation factor (Rawat et al. 2021).

Besides transcription factors, the co-activator Mediator and RNA-binding proteins, increasing evidence shows that RNA can facilitate the formation of nuclear condensates containing Pol II and other components of the transcriptional machinery. With its high negative charge RNA binds to positively charged disordered protein regions and changes the physicochemical properties of its surroundings, which can influence CTD phase separation (Shao et al. 2022). The CTD of Pol II also forms coacervates in presence of intracisternal A-particle RNA (Asimi et al. 2022). RNA thus can play an important role in the condensation of Pol II, contributing to its regulation and function during gene transcription.

5 Concluding remarks

Our understanding of how transcription is regulated is constantly evolving with new evidence of the presence of transcriptional hubs *in vitro* and in cells. Liquid-liquid phase separation provides a mechanism for the spatio-temporal organization during transcription, with the CTD playing a crucial role in regulating the activity of RNA Pol II through post-translational modifications and migration between coacervates. In addition, growing interest focuses on exploring the underlying physical and chemical forces

that determine the selectivity of Pol II localization to different condensates, and to uncover how these forces can be targeted to modulate specific transcriptional hubs at different stages of transcription. This may open up new avenues for developing strategies to regulate transcription with greater precision and specificity.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: M.Z. was supported by the Deutsche Forschungsgemeinschaft (DFG; SFB860/3, Project B2).

Conflict of interest statement: The authors declare no competing interest.

References

- Abdella, R., Talyzina, A., Chen, S., Inouye, C.J., Tjian, R., and He, Y. (2021). Structure of the human Mediator-bound transcription preinitiation complex. *Science* 372: 52–56.
- Aibara, S., Schillbach, S., and Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature* 594: 124–128.
- Asimi, V., Sampath Kumar, A., Niskanen, H., Riemenschneider, C., Hetzel, S., Naderi, J., Fasching, N., Popitsch, N., Du, M., Kretzmer, H., et al. (2022). Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* 54: 1238–1247.
- Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kocic, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* 25: 833–840.

- Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.v., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175: 1842–1855.e16.
- Burger, K., Schlackow, M., and Gullerova, M. (2019). Tyrosine kinase c-Abl couples RNA polymerase II transcription to DNA double-strand breaks. *Nucleic Acids Res.* 47: 3467–3484.
- Burke, K.A., Janke, A.M., Rhine, C.L., and Fawzi, N.L. (2015). Residue-by-Residue view of *in vitro* FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell.* 60: 231–241.
- Cagas, P.M. and Corden, J.L. (1995). Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins Struct. Funct. Genet.* 21: 149–160.
- Chen, X., Wang, X., Liu, W., Ren, Y., Qu, X., Li, J., Yin, X., and Xu, Y. (2022). Structures of +1 nucleosome-bound PIC-Mediator complex. *Science* 378: 62–68.
- Chen, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., and Xu, Y. (2021). Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* 372: eabg0635.
- Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361: 412–415.
- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361: 378.
- Collin, P., Jeronimo, C., Poitras, C., and Robert, F. (2019). RNA polymerase II CTD tyrosine 1 is required for efficient termination by the *nrd1-nab3-sen1* pathway. *Mol. Cell* 73: 655–669.e7.
- Eick, D. and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113: 8456–8490.
- Gibbs, E.B., Lu, F., Portz, B., Fisher, M.J., Medellin, B.P., Laremore, T.N., Zhang, Y.J., Gilmour, D.S., and Showalter, S.A. (2017). Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* 8: 15233.
- Guo, C., Che, Z., Yue, J., Xie, P., Hao, S., Xie, W., Luo, Z., and Lin, C. (2020). ENL initiates multivalent phase separation of the super elongation complex (SEC) in controlling rapid transcriptional activation. *Sci. Adv.* 6: eaay4858.
- Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.v., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572: 543–548.
- Janke, A.M., Seo, D.H., Rahmanian, V., Conicella, A.E., Mathews, K.L., Burke, K.A., Mittal, J., and Fawzi, N.L. (2018). Lysines in the RNA polymerase II C-terminal domain contribute to TAF15 fibril recruitment. *Biochemistry* 57: 2549–2563.
- Jasnovidova, O. and Stefl, R. (2013). *The CTD code of RNA polymerase II: a structural view*. Wiley Interdiscip. Rev. 4: 1–16.
- Kumaki, Y., Matsushima, N., Yoshida, H., Nitta, K., and Hikichi, K. (2001). Structure of the YSPSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1548: 81–93.
- Lu, F., Portz, B., and Gilmour, D.S. (2019). The C-terminal domain of RNA Polymerase II is a multivalent targeting sequence that supports *Drosophila* development with only consensus heptads. *Mol. Cell.* 73: 1232–1242 e4.
- Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 558: 318–323.
- Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.v., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* 367: 694–699.
- Mayfield, J.E., Irani, S., Escobar, E.E., Zhang, Z., Burkholder, N.T., Robinson, M.R., Rachel Mehaffey, M., Sipe, S.N., Yang, W., Prescott, N.A., et al. (2019). Tyr1 phosphorylation promotes phosphorylation of Ser2 on the C-terminal domain of eukaryotic RNA polymerase II by P-TEFb. *eLife* 8: e48725.
- Murthy, A.C., Tang, W.S., Jovic, N., Janke, A.M., Seo, D.H., Perdikari, T.M., Mittal, J., and Fawzi, N.L. (2021). Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nat. Struct. Mol. Biol.* 28: 923–935.
- Namitz, K.E.W., Zheng, T., Canning, A.J., Alicea-Velazquez, N.L., Castañeda, C.A., Cosgrove, M.S., and Hanes, S.D. (2021). Structure analysis suggests Ess1 isomerizes the carboxy-terminal domain of RNA polymerase II via a bivalent anchoring mechanism. *Commun. Biol.* 4: 1–14.
- Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat. Struct. Mol. Biol.* 12: 144–151.
- Palumbo, R.J., McKean, N., Leatherman, E., Namitz, K.E.W., Connell, L., Wolfe, A., Moody, K., Gostinčar, C., Gunde-Cimerman, N., Bah, A., et al (2022). Coevolution of the Ess1-CTD axis in polar fungi suggests a role for phase separation in cold tolerance. *Sci. Adv.* 8: eabq3235.
- Portz, B., Lu, F., Gibbs, E.B., Mayfield, J.E., Rachel Mehaffey, M., Zhang, Y.J., Brodbelt, J.S., Showalter, S.A., and Gilmour, D.S. (2017). Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* 8: 15231.
- Quintero-Cadena, P., Lenstra, T.L., and Sternberg, P.W. (2020). RNA Pol II length and disorder enable cooperative scaling of transcriptional bursting. *Mol. Cell* 79: 207–220.e8.
- Ramani, M.K.V., Escobar, E.E., Irani, S., Mayfield, J.E., Moreno, R.Y., Butalewicz, J.P., Cotham, V.C., Wu, H., Tadros, M., Brodbelt, J.S., et al. (2020). Structural motifs for CTD kinase specificity on RNA polymerase II during eukaryotic transcription. *ACS Chem. Biol.* 15: 2259–2272.
- Rawat, P., Boehning, M., Hummel, B., Aprile-Garcia, F., Pandit, A.S., Eisenhardt, N., Khavaran, A., Niskanen, E., Vos, S.M., Palvimo, J.J., et al. (2021). Stress-induced nuclear condensation of NELF drives transcriptional downregulation. *Mol. Cell.* 81: 1013–1026.e11.
- Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., and Cramer, P. (2021). Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature* 594: 129–133.
- Shao, W., Bi, X., Pan, Y., Gao, B., Wu, J., Yin, Y., Liu, Z., Peng, M., Zhang, W., Jiang, X., et al. (2022). Phase separation of RNA-binding protein promotes polymerase binding and transcription. *Nat. Chem. Biol.* 18: 70–80.
- West, M.L. and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140: 1223–1233.
- Zaborowska, J., Eglhoff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* 23: 771–777.
- Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro*. *Proc. Natl. Acad. Sci. USA* 85: 3698–3702.

Zhang, J. and Corden, J.L. (1991). Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J. Biol. Chem.* 266: 2297–2302.

Zhou, M., Ehsan, F., Gan, L., Dong, A., Li, Y., Liu, K., and Min, J. (2022). Structural basis for the recognition of the S2, S5-phosphorylated RNA polymerase II CTD by the mRNA anti-terminator protein hSCAF4. *FEBS Lett.* 596: 249–259.