CellPress

The Long and the Short of the RNA Polymerase C-Terminal Domain and Phase Separation

R. Stefan Isaac[1] and L. Stirling Churchman[1,*]
[1]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: churchman@genetics.med.harvard.edu
https://doi.org/10.1016/j.molcel.2019.03.004

In this issue of *Molecular Cell*, Lu et al. (2019) analyze the role of the length and sequence complexity of the RNA polymerase II unstructured C-terminal domain in animal viability, development, and the dynamics of RNA polymerase II *in vivo*.

The largest subunit of RNA polymerase II (Pol II), Rbp1, contains a long, unstructured C-terminal domain (CTD) comprising repeated heptapeptides with the consensus sequence YSPTSPS. This tail serves as a docking site for myriad proteins involved in transcription-related functions, including splicing, 5′ capping, and 3′ processing of mRNAs (reviewed in Harlen and Churchman, 2017). Despite its conserved function and the presence of the consensus repeat sequence, the CTD varies widely across eukaryotes in both length and degeneracy of the heptapeptide sequences (Yang and Stiller, 2014). For example, the yeast *Saccharomyces cerevisiae* CTD is made up of 26 repeats that, like the first half of mammalian CTDs, consist of the heptapeptide consensus sequence. By contrast, the mammalian CTD continues for another 26 repeats that are largely degenerate in sequence.

In this issue of *Molecular Cell*, Lu et al. (2019) describe a study in which they used *Drosophila melanogaster* as a model organism to study the functional importance of both CTD length and sequence for animal viability and development. Of the 42 repeats in the *Drosophila* CTD, only 2 match the consensus sequence. Lu et al. show that shortening the length of the CTD to roughly the length of the *S. cerevisiae* tail while also altering the sequence such that every repeat consists of consensus heptad repeats has no impact on fly viability, proper development, or fertility. These findings demonstrate that the divergence of CTD motifs across phyla is not required for normal function—why, then, has the sequence degenerated over time?

Surprisingly, Lu et al. (2019) found that a fully consensus CTD with the normal *Drosophila* length of 42 repeats is lethal (Figure 1). In other words, *Drosophila* function perfectly well without their divergent sequences but cannot survive with too many repeated consensus heptads. This phenomenon is reminiscent of unstructured protein domains that undergo liquid-liquid phase separations (LLPS) under normal conditions but aggregate when mutated (Banani et al., 2017). Phase separation is driven by weak, multivalent interactions that prevail among low-complexity or intrinsically disordered domains (IDR). As the name implies, IDRs lack a rigid three-dimensional structure and often consist of repeated sequence elements enriched in particular amino acids, especially tyrosine and serine (Banani et al., 2017; Kato et al., 2012). Notably in this regard, 4 of the 7 CTD heptad residues are tyrosine or serine. These repeated sequences engage in multiple intermolecular interactions that drive phase separation of these molecules. When these interactions are too numerous, however, phase separation leads to protein aggregation. If the CTD is capable of phase separation, then an overabundance of heptad repeats would progressively lead to aggregation, loss of function, low transcription, and lethality.

On its own, the CTD undergoes phase separation *in vitro*, and this partitioning is highly dependent on length: the shorter *S. cerevisiae* CTD forms droplets at much higher concentrations than its longer mammalian counterpart (Boehning et al., 2018). In human cells, Pol II forms clusters throughout the nucleus; shortening the CTD results in fewer clusters, whereas lengthening it results in more (Boehning et al., 2018; Cho et al., 2018). Phosphorylation of serine 5 in the heptad sequence, which allows Pol II to escape from the promoter-proximal region, dissolves droplets *in vitro* (Boehning et al., 2018). These observations and others support a model in which Pol II phase-separates into transient clusters, escaping only upon CTD phosphorylation.

Lu et al. (2019) hypothesized that CTD length and sequence affect the localization of Pol II to sites of transcription. To test this idea, they analyzed the CTD directly by fusing just the tail to a nuclear-localized fluorescent protein and then monitoring the localization of this construct (Figure 1). The authors took advantage of the large, transcriptionally active loci on polytene chromosomes in *Drosophila* salivary glands (“puffs”), which are readily observable due to their size. They found that fluorescently tagged CTD localized to puffs in a highly dynamic manner, a characteristic property of phase separations. Outside of the puffs, the CTD formed static foci more consistent with protein aggregates. Increasing the length of the CTD with consensus heptads increased the number of static foci. Together, these findings demonstrate that the CTD is sufficient for localization of Pol II to clusters but aggregates when containing too many repeats of the consensus heptad sequence.

Along with other published reports, these data support the hypothesis that the CTD of Pol II partitions this enzyme, along with other key factors that bind the

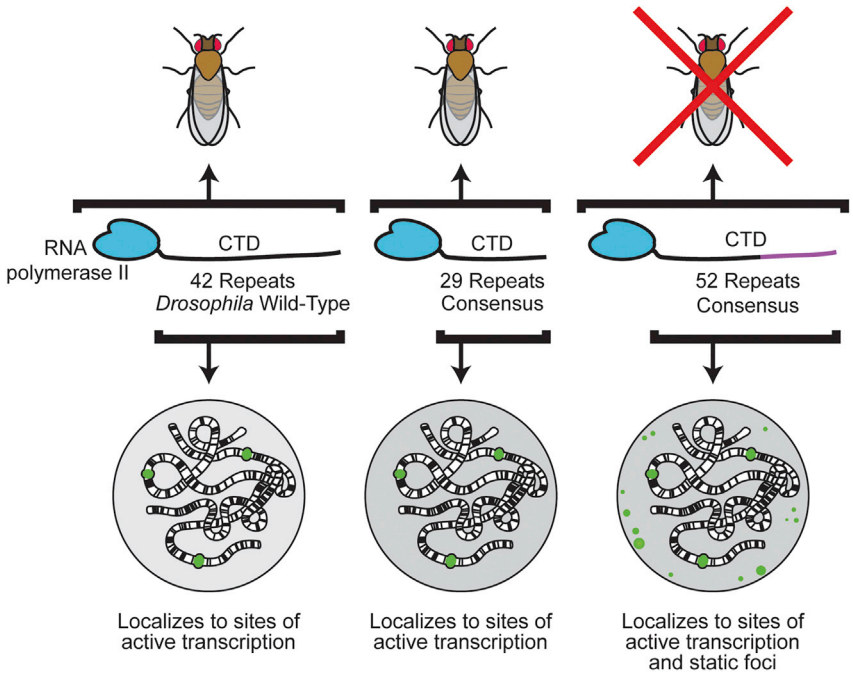

Figure 1. Effects of CTD Length and Composition on *Drosophila* Viability and CTD Localization

Depicted in the middle of the figure are three constructs used in the Lu et al. study: RNA polymerase II with the wild-type *Drosophila* CTD, with 29 consensus repeats, and with 42 consensus repeats. The fly viability when each construct is expressed in *Drosophila* is depicted above each construct. The localization pattern of the CTD alone fused to a fluorescent protein in *Drosophila* salivary gland nuclei is illustrated below.

CTD, to facilitate transcription and transcription-related processes (Boehning et al., 2018; Cho et al., 2018; Harlen and Churchman, 2017; Kato et al., 2012; Kwon et al., 2013). The number of consensus CTD repeats needs to be high enough to drive partitioning given the nuclear concentration of Pol II. On the other end of the spectrum, CTDs that are too long and consist solely of the consensus heptapeptides drive the system beyond phase separation into aggregation. Consequently, throughout evolutionary history, changes in Pol II have led to differences in length and sequence. The CTD sequence of *S. cerevisiae* is almost completely consensus—but short, promoting phase separation rather than aggregation. Longer CTDs deviate from the consensus sequence, and these differences may have arisen because they lead to weaker interactions, thereby preventing aggregation.

Phase separation of Pol II at super-enhancers and transcriptional puffs on polytene chromosomes involves hundreds of Pol II enzymes (Cho et al., 2018). However, Pol II clusters in most cells (and at most loci) are likely to be far smaller, consisting of tens of polymerases (Cho et al., 2016). At lowly expressed genes, the number of Pol II enzymes could be even smaller. The physics of phase separation assumes many degrees of freedom, which in many systems is provided by a large molecular ensemble. By contrast, when only a few degrees of freedom are available, the notion of phase separation is not well defined. For example, it is nonsensical to ask whether two H_2O molecules can undergo a phase transition from vapor to water. With this in mind, can we assess the possibility of phase separation at loci where just a few Pol II enzymes are bound? In answering this question, it is critical to consider the heptad repeats, which provide additional degrees of freedom. Even 26 repeats on a small number of Pol II enzymes could adopt a large number of possible conformations, potentially providing a large enough ensemble to enable phase separation. Moreover, other biomolecules, such as proteins with low-complexity domains, as well as RNA itself, are likely to contribute to forming Pol II clusters (Banani et al., 2017). From this perspective, even two Pol II proteins could undergo phase separation. Future work should seek to determine whether phase separation provides important advantages in gene expression dynamics and regulation. If so, existing models of transcription regulation will need to be updated to incorporate this behavior.

REFERENCES

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, 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.

Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., and Cisse, I.I. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. eLife *5*, e13617.

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.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. *18*, 263–273.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049–1060.

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, this issue, 1232–1242.

Yang, C., and Stiller, J.W. (2014). Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. Proc. Natl. Acad. Sci. USA *111*, 5920–5925.