Conditions for the co-existence of promoter and gene-body condensates

Arya Changiarath,[1,2,3,4] Jasper J. Michels,[5] Rosa Herrera Rodriguez,[4] Sonya M. Hanson,[6] Friederike Schmid,[1,a)] Jan Padeken,[4,b)] and Lukas S. Stelzl[1,2,3,4,c)]

[1)] *Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz, Germany*

[2)] *Institute of Molecular Physiology, Johannes Gutenberg University Mainz, 55128 Mainz, Germany*

[3)] *Institute of Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg University Mainz, 55128 Mainz, Germany*

[4)] *Institute of Molecular Biology (IMB), 55128 Mainz, Germany*

[5)] *Max Planck Insitute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

[6)] *Flatiron Institute, New York, NY, USA*

In cells, transcription is tightly regulated on multiple layers. The condensation of the transcription machinery into distinct phases is hypothesized to spatio-temporally fine tune RNA polymerase II behaviour during two key stages, transcription initiation and the elongation of the nascent RNA transcripts. However, it has remained unclear whether these phases would mix when present at the same time or remain distinct chemical environments; either as multi-phase condensates or by forming entirely separate condensates. Here we combine particle-based multi-scale simulations and experiments in the model organism *C. elegans* to characterise the biophysical properties of RNA polymerase II condensates. Both simulations and the in vivo work describe a lower critical solution temperature (LCST) behaviour of RNA Polymerase II, with condensates dissolving at lower temperatures whereas higher temperatures promote condensate stability, which highlights that these condensates are physio-chemically distinct from heterochromatin condensates. The LCST behavior of CTD correlates with gradual shifts in the transcription program but is largely uncoupled from the classical stress response. Expanding the simulations we model how the degree of phosphorylation of the disordered C-terminal domain of RNA polymerase II (CTD), which is characteristic for each step of transcription, controls the existence and morphology of multi-phasic condensates. We show that the two phases putatively underpinning the initiation of transcription and transcription elongation constitute distinct chemical environments and are in agreement with RNA polymerase II condensates observed in *C. elegans* embryos by super resolution microscopy. Our analysis shows how depending on its post transcriptional modifications and its interaction partner a single protein can form multiple partially engulfed condensates, potentially promoting the selective recruitment of additional factors to these two phases.

[a)] Electronic mail: friederike.schmid@uni-mainz.de

[b)] Electronic mail: j.padeken@imb-mainz.de

[c)] Electronic mail: lstelzl@uni-mainz.de

INTRODUCTION

Phase separation of biomolecules to form condensates provides a mechanism for cells to control processes in time and space. These condensates create local differences in protein concentrations, which is a way how phase separation can specifically promote or impede biological processes. Numerous proteins within cells have the potential to undergo phase separation under physiological conditions[1]. Should these proteins converge into a single location, condensation would neither increase local protein concentration nor enable specific biological processes. Specificity requires, therefore multiple co-existing phase-separated condensates to provide for spatio-temporal control over cellular processes[2]. As an additional layer of even finer regulation molecules may not only form distinct condensates that are entirely separated but also multi-phasic condensates[3]. In multi-phase condensates, there are distinct phases that do not mix. An essential question is what are the factors that determine the existence of multiple condensates and multiple phases in the condensates[4–6]. To understand the behaviour of condensates we have to describe three layers: the primary protein sequence, post transcriptional, reversible, chemical modifications, and how the two former alter and are altered by protein-protein interactions[7,8].

To study the roles of multiphasic condensates, their co-existence, and how they might provide specific molecular recognition and thus cellular regulation, we turned to the C-terminal domain (CTD) of RNA Polymerase II (Pol II). As the central protein which transcribes genes from genomic DNA into mRNA the molecular mechanisms controlling RNA Pol II have been extensively studied over the last decades. Transcription (Fig 1A) is a linear process that can be divided into several regulatory steps, starting with transcriptional initiation at the promoter/transcriptional start side, an elongation process across the gene body, and finally transcriptional termination[9–11]. Recent experimental advances hypothesize that transcriptional initiation and post transcriptional processes associated with elongation might occur in different biomolecular condensates[10,12–17]. The formation of biomolecular condensates is frequently driven by multivalent interactions within intrinsically disordered regions of proteins and many proteins involved in transcription feature such disordered regions[18]. The CTD of RNA polymerase II is an intrinsically disordered protein domain that can form phase-separated condensates and its ability to form different condensates has been hypothesized to be important in the transcription of genes[10,11,13] (Fig 1A). CTD forms condensates *in*

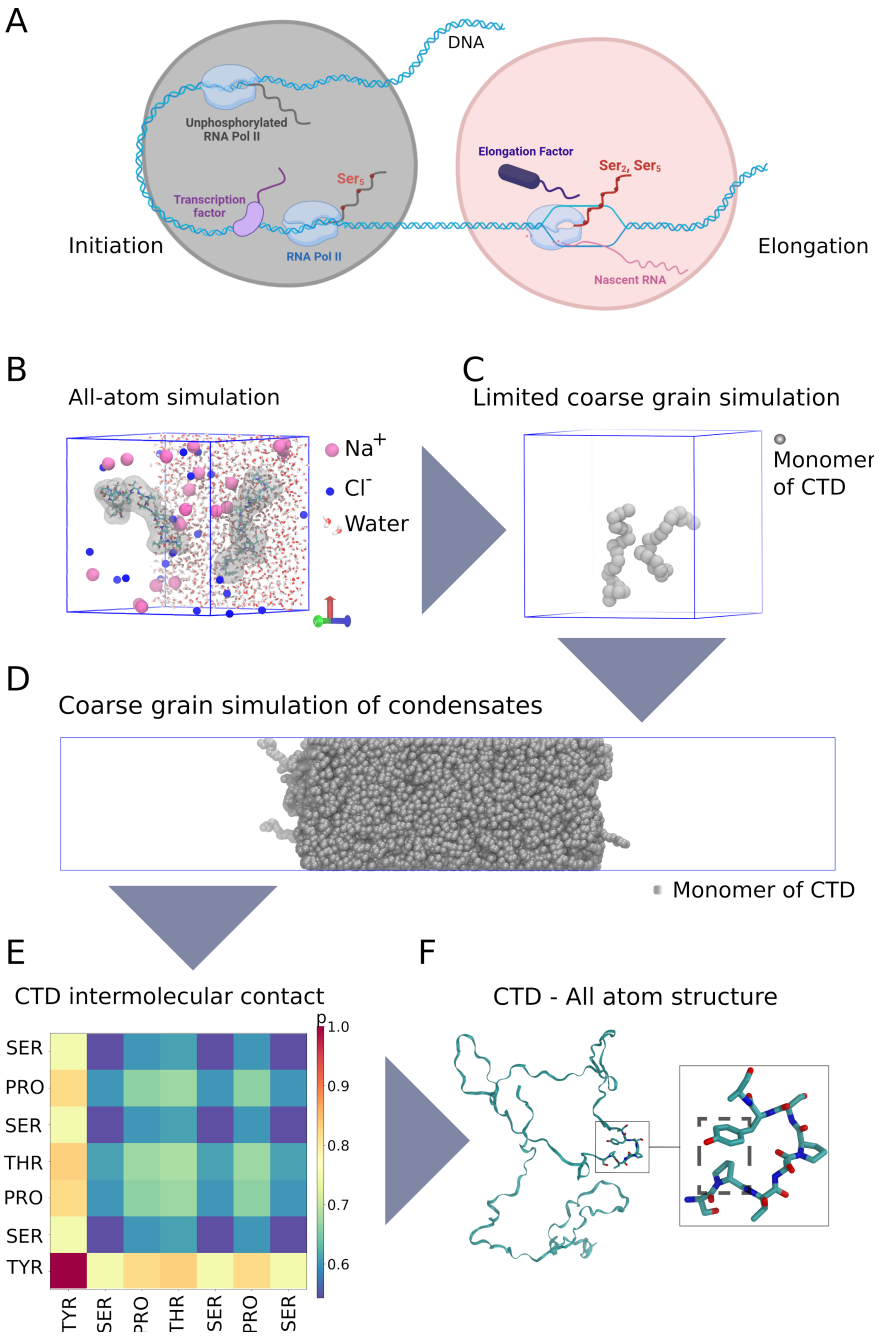

Figure 1. A) Hypothesized condensates underlying the initiation of gene transcription and the elongation phase. B) All-atom simulation capturing the behavior of two 21mer CTD peptides. C) Coarse-grained(CG) simulation featuring two 21mer CTD peptides. D) CG slab simulation depicting CTD condensate. E) Interactions among CTDs in CG simulation revealed through intermolecular contact frequency. F) All-atom model of yeast CTD from Ref. 1 highlighting Tyr-Pro contacts.

vitro and contributes to RNA Polymerase II phase separation *in vivo*[13]. These condensates are thought to underpin the recruitment of RNA Pol II and the initiation of gene transcription, which are referred to as promoter condensates. CTD is then further phosphorylated as transcription initiation switches to transcription elongation and syntheses the RNA from the DNA template. Phosphorylated CTD forms condensate with proteins involved in transcription elongation and RNA processing[14,15], which are called gene-body condensates[10]. Several proteins involved in regulating transcription elongation, such as P-TEFb directly or indirectly (splicing), have been observed to localize within these condensates. Whether such condensates will stay de-mixed or mixed when in direct contact remains unclear. Recently, a surface condensation model was proposed to account for the dynamics of RNA polymerase II as it progresses from initiation and enters elongation as revealed by super resolution experiments in zebrafish[19]. However, splicing condensates remain de-mixed from initiation condensates[15]. To disentangle how condensates impact different stages of transcription it is therefore crucial to comprehensively characterise the phase behavior of RNA Pol II and examine how phosphorylation affects condensates.

Molecular simulations can serve as a "computational microscope"[20] to elucidate drivers of cellular processes and phase behavior on the molecular-scale[3,6,21–30] and are especially powerful when combined with experiments[7,31]. Atomistic molecular dynamics (MD) simulations can resolve molecular interactions in condensates with high accuracy[7,32]. However simulating phase separation requires accessing long time and large length scales and thus the description of the system has to be simplified, by grouping atoms into particles, which is referred to as coarse-graining. Current coarse-grained models can capture important trends in how changes in sequence[33] and post-translational modifications shape the single-chain and condensed phase behavior of disordered proteins[23,34,35]. The loss of atomistic detail, begs the question whether all essential interactions are captured with sufficient accuracy. To ensure that simulations accurately describe phase behaviour requires comparison of coarse-grained simulations to atomistic molecular dynamics and experiments reporting on biophysical properties[34].

Here, we elucidate, using molecular dynamics, conditions under which the CTD, phosphorylation of the CTD (pCTD), and the binding of the HRD domain of cyclin T1 to pCTD, promote RNA Polymerase II condensate formation. Using the model organism *C. elegans*, we tested the predictions made by our simulations in a living organism under physiological

conditions. Strikingly, in the simulations and in *C. elegans* embryos, CTD forms condensate with lower critical solution temperature (LCST) behaviour. This clearly distinguishes it from other previously characterised biological condensates in the nucleus, which typically dissolve at high temperatures, and hints at discrete chemical properties intrinsic to IDRs that could contribute to the co-existence of multiple distinct condensates in the nucleus. Our simulations show that CTD and pCTD-HRD indeed do not mix but instead form distinct phases, with pCTD-HRD engulfing the CTD phase. The simulations further predict that the exact degree of engulfment would vary based on conditions such as condensate composition. To ensure that our simulations accurately capture the molecular driving forces of CTD phase separation, we compare coarse-grained simulations to atomistic molecular dynamics, nuclear magnetic resonance (NMR) based structural ensembles, and also developed a general simulation approach to predict the morphologies of condensates involving multiple phases from molecular simulations by computing interfacial tensions. In agreement with simulations, we find in super-resolution microscopy that in *C. elegans* embryos CTD is indeed fully and partially engulfed by pCTD. While the two condensates are largely de-mixed and distinct, they remain in close contact, which may underpin efficient movement of RNA Polymerase II from initiation to elongation as part of the functional cycle of RNA Polymerase II.

METHODS

Coarse-grained simulation model

We employed coarse-grained molecular dynamics simulations using the residue-level hydropathy scale (HPS) model which represents amino acids as beads and the solvent effect is considered implicitly[21]. The HPS model intricately captures the physical and chemical attributes inherent to the natural ensemble of 20 amino acids into these beads, resulting in sequence-specific interactions among them. Three kinds of interaction are considered between the beads. Firstly, beads are bonded to each other through harmonic potential energy, $V_H(r) = k(r - r_0)^2/2$ where $10\,\mathrm{kJ/\AA^2}$ is force constant and $r_0 = 3.8\,\mathrm{\AA}$ is the bond length. Secondly, electrostatic interaction between charged amino acids is modeled using the Debye–Hückel (DH) electrostatic screening term.

$$E_{ij} = \frac{q_i q_j}{4\pi D r} \exp\left(-\frac{r}{k}\right), \tag{1}$$

where $k = 1$ nm is the Debye screening length which corresponds to a salt concentration of 100 mM and $D = 80$ is the dielectric constant of the aqueous solvent. A cut-off distance of $r_{cut} = 3.5$Å is set. If r is greater than r_{cut}, then E_{ij} is set to zero. Lastly, the short-range pairwise interaction between beads is represented by Ashbaugh–Hatch potential,

$$\phi(r) = \begin{cases} \phi_{LJ}(r) + (1 - \lambda_{ij})\epsilon, & \text{if } r < 2^{1/6}\sigma_{ij}. \\ \lambda_{ij}\phi_{LJ}(r), & \text{otherwise.} \end{cases} \tag{2}$$

where ϕ_{LJ} is the Lennard-Jones potential,

$$\phi_{LJ}(r) = \begin{cases} 4\epsilon[(\sigma_{ij}/r)^{12} - (\sigma_{ij}/r)^{6}], & \text{if } r < r_{cut}. \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

and λ_{ij} and σ_{ij} are the arithmetic average of the hydrophobicity scale and size of amino acids, respectively. A cut-off distance of $r_{cut} = 2.00$Å is used and the potential is shifted accordingly. The residue specific hydrophobicity values range from 0 to 1 as the hydrophobicity increases. The HPS model also provides parameters for phosphorylated amino acids such as phospho-serine[8], which have been used to model the effect of phosphorylation on the fifth position in the CTD heptad repeat motifs, which we refer to as Ser_5. To study temperate-dependent phase behavior we used the HPS-T model, a version of the HPS model which has been extended to predict the temperature dependent properties of disordered proteins[36].

Coarse-grained molecular dynamics simulations were run with the HOOMD-Blue package[37]. The system is initialized by placing the protein chains inside an extended rectangular box with a minimum distance between protein chains. The simulations were performed under NVT ensemble conditions, using a Langevin thermostat with a friction coefficient of $0.01ps^{-1}$. The proteins are equilibrated within the first few nanoseconds by condensing into a liquid dense and dilute phase. The density profile is extracted from the protein concentration along the z axis as described by Tesei et al.[23].

Atomistic molecular dynamics simulations

Two 21mer CTD fragments were simulated in all-atom molecular dynamics with explicit solvent. We employed the Amber99sb-star-ildn-q protein force field[38–41] and the TIP4P-D water model[42] in the simulations. Equations of motions were integrated with a 2 fs time step.

The *NPT* ensemble was maintained at 300 K and 1 bar with the Bussi-Donadio-Parrinello thermostat[43] and the Parrinello-Rahman barostat[44]. Simulations were run in GROMACS[45]. After initial equilibration for 129 ns, 100 trajectories each with different starting velocities were launched on Folding@Home[46], which resulted in a cumulative simulation duration of 12.9 µs.

Simulated systems

Employing coarse-grained molecular dynamics simulations enables us to gain an in-depth understanding of the complexities involved in the regulation of gene transcription. First, we simulated CTD proteins from *C.elegans*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Homo sapiens* (Fig S2). To understand how deviations from ideal heptad repeat sequence $Y_1S_2P_3T_4S_5P_6S_7$ affect phase behaviour we also conducted simulations of chains of the same length as these natural proteins, but with idealized sequences. Simulations of *C. elegans* CTD were also run with the HPS-T model to investigate its temperature dependent phase behaviour[36].

To study the formation of distinct phases or multiphasic condensates we focus on a simplified system, comprising three key components that play a vital role in the transcription process and may underlie key aspects of phase behaviour in the nucleus. These elements consist of CTD of RNA polymerase II (140 mer with 20 ideal heptad repeats), responsible for the promoter condensate; phosphorylated CTD (140 mer with various degrees of phosphorylation); and Histidine Rich Domains (HRD) of P-TEFb Cyclin T1, contributing to the formation of the gene body condensate (Fig. 1A).

Analysing the properties of these condensates provides insights into their role in gene regulation. Additionally, the coarse-grained nature of the simulation allows us to explore large-scale and extended-time thermodynamic and mechanical properties, crucial for understanding the stability and structure of multiphasic condensates. We initiated our study by examining the formation and regulation of the promoter condensate, which is formed through the phase separation of CTD. We explored the impact of phosphorylation on CTD phase separation and the co-phase behavior of HRD with CTD and pCTD through a series of simulations. The coacervation behavior of HRD and pCTD was explored using the flory-huggins theory. To gain a comprehensive understanding of how phosphorylation impacts the

formation of multiphase condensates, we conducted simulations involving CTD, pCTD, and HRD, systematically varying the phosphorylation levels of the pCTD chains in each simulation. Furthermore, to explore the structure and stability of these multiphase condensates, we carried out simulations of droplets of these condensates in a bigger cubic box.

Analysis of phase behaviour with a mean-field model

We analyse the simulated phase diagram of pCTD and HRD co-phase separation with a mean-field Flory-Huggins model[47],

$$\frac{\Delta f}{k_B T} = \frac{\phi_A}{N_A} \ln \phi_A + \frac{\phi_B}{N_B} \ln \phi_B + \phi_S \ln \phi_S + \chi_{AB}\phi_A\phi_B + \chi_{AS}\phi_A\phi_S + \chi_{BS}\phi_B\phi_S, \qquad (4)$$

Here, ϕ_i and N_i represent the volume fractions and relative degrees of polymerisation and χ_{ij} the binary interaction parameters. In the model, ϕ_S follows from the assumption of incompressibility: $\phi_S = 1 - \phi_A - \phi_B$. Subscripts A, B, and S respectively denote HRD, pCDT, and "solvent". The relative sizes of the polymers, which we fix at $N_A = 70$ and $N_B = 140$, have been normalized by a scaled monomeric volume ($N_S = 1$).

The mean-field phase diagram was obtained in the usual way by demanding that the chemical potential and the osmotic pressure be similar in the dilute and dense phase, using well known methodologies[48].

Calculation of interfacial tension

Interfacial tension is determined using two methods employing capillary wave theory: first, by assessing interfacial width in slab simulations, and second, by analysing shape fluctuations of droplets.

In slab geometries, capillary wave theory (CWT) with the block analysis method was employed to estimate the interfacial tension of condensates[49]. In the block analysis method, the interface region is divided into rectangular segments with dimensions of $L_x/n_B \times L_y/n_B \times L_z$, where n_B is an integer, where we choose $3 < n_B < 8$. Within each segment, intrinsic profiles are computed based on the density function:

$$\rho(z) = \frac{\rho_A + \rho_B}{2} + \frac{\rho_A - \rho_B}{2} \tanh\left(\frac{z - z_0}{W_0}\right) \tag{5}$$

Where ρ_A and ρ_B represent the bulk densities of two phases, denoted as A and B, respectively. z_0 is the position of the interface, while W_0 corresponds to the intrinsic width of the system. The apparent interfacial width (W) at a finite temperature is determined by combining the intrinsic width W_0 and the capillary contribution:

$$W^2 = W_0^2 + \frac{K_B T}{2\pi\gamma} \ln\left(\frac{L}{B}\right), \tag{6}$$

where $k_B T$ is the thermal energy, γ represents the interfacial tension, and L/B is the ratio of system size to block size. The apparent interfacial width (W) is computed as a function of the block size and interfacial tension can be calculated using equation 6. With CWT, we can determine the interfacial tension at the interface between a dense protein phase with the solvent, as well as at the interface between two distinct protein phases.

We compared the interfacial tension of the homogeneous condensate, as obtained from the above calculation, with the interfacial tension obtained using Kirkwood-Buff relation[50],

$$\gamma = \frac{L_z}{2}\left(P_{zz} - \frac{P_{xx} + P_{yy}}{2}\right), \tag{7}$$

where, $P_{\alpha\beta}$ are the components of the pressure tensor.

Furthermore, the interfacial tension of a droplet can be estimated from the thermal fluctuations of its spherical shape based on the method pioneered by Henderson and Lekner[51] which has found recent application in the analysis of biological condensates[22]. The potential energy (U) associated with these shape fluctuations is primarily determined by the interfacial tension (γ), which can be mathematically expressed as $U = \gamma\delta A$. Here, δA is the small variation in interfacial area relative to that of a perfect sphere. By analysing the droplet's shape fluctuations at the lowest order using a spherical harmonics expansion, we can derive two independent measures of the interfacial tension:

$$\gamma_{20} = \frac{5k_B T}{16\pi\langle(\delta a + \delta b)^2\rangle} \tag{8}$$

and

$$\gamma_{22} = \frac{5k_B T}{16\pi\langle(\delta a - \delta b)^2\rangle} \tag{9}$$

Here, the droplet's instantaneous shape is described by a general ellipsoid, with δa and δb representing the lengths of any pair of principal axes of the ellipsoid, and R indicating the

average radius of the droplet. By selecting the relevant protein chains, we computed surface tensions of the homogeneous protein phases as well as interfacial tensions between different protein phases.

Characterisation of disordered protein conformation

In our study, we investigate the impact of phosphorylation on the conformation and compactness of CTD chains in dilute and dense phases. A key metric for this analysis is the radius of gyration (R_G). Additionally, we assess protein anisotropy to identify structural changes resulting from phosphorylation. The gyration radius of chains is calculated using the following equation:

$$\mathbf{G} = \frac{1}{M}\sum_{i}^{N} m_i \Delta\mathbf{r}_i \Delta\mathbf{r}_i^T \tag{10}$$

Here, M represents the mass of the chain, N is the number of monomers in the chain, and$\Delta\mathbf{r}_i$ is the position vector of monomer i with respect to the chain's center-of-mass. The root mean square radius of gyration is calculated from the diagonal elements of the $\mathbf{G}$ matrix, given by $R_g = \langle Gxx + G_{yy} + G_{zz}\rangle^{1/2}$. Furthermore, we examined the anisotropy of the chain using the following equation[50]:

$$A = \left\langle 1 - 3\frac{G_1G_2 + G_1G_3 + G_2G_3}{(G_1 + G_2 + G_3)^2}\right\rangle \tag{11}$$

Here, G_1, G_2, G_3 represent the three eigenvalues of the matrix $\mathbf{G}$. Analysing the anisotropy of the chain allows us to characterise any structural modifications and deviations from a symmetric conformation that may arise due to phosphorylation.

Interactions of disordered protein chains in condensates

The excess of monomer A in the vicinity of a B monomer, n_{ab}^{exc} is calculated using the equation:

$$n_{ab}^{exc} = \rho_a \int_{r_1}^{r_2} dr 4\pi r^2 (g_{ba}(r) - 1) \tag{12}$$

Here, r_1 and r_2 represent the minimum and maximum radial distances from the B monomer, respectively, over which the excess number of monomer A is computed. The specified cutoff

values are $r_1 = 0.45, \mathrm{nm}$ and $r_2 = 7.5, \mathrm{nm}$. The correlation function $g_{ba}(r)$ is defined as:

$$g_{ab}(r) = \frac{V}{N_b N_a} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \frac{\delta(|\vec{r}_i - \vec{r}_j| - r)}{4\pi r^2} \tag{13}$$

Here, N_a and N_b represent the total number of monomers of type A and B, respectively. The excess of monomer,n_{ab}^{exc}, functions as a metric for quantifying the degree of protein-protein interactions.

C. elegans strains and maintenance

AMA-1::GFP (ama-1(ot1037[gfp::ama-1]) IV.) strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). Unless otherwise indicated, experiments were performed using early embryos isolated from animals cultured at 20 °C.

Live microscopy and Immunofluorescence (IF)

Fluorescent images of live animals or embryos were captured in M9 buffer on 2% agarose pads, embryos were incubated at the indicated temperatures for 1h prior to imaging. Images of live embryos were captured using the LAS X Life Science Microscope Software with the Stellaris 8 FALCON (Leica) equipped with an HC PL APO 100x/1.40 OIL STED WHITE objective lens and a white light laser source. Z-stacks were recorded with a z-spacing of 200 nm and an xy-pixel size of 54 nm. All images were deconvolved using the inbuild LIGHTNING deconvolution module (Leica). For IF, bleached embryos were fixed for 5 min in 1% formaldehyde before spotting onto poly-l-lysine–coated coverslips, freezing on dry ice, and storing at −80 °C. When needed, slides were freeze cracked and immediately fixed in a −20 °C 100% ethanol (EtOH) bath for 2 min and then dried. After washing 3 × 5 min with PBS +0.25% Triton X-100 (PBS-T) and blocking for 1 h with PBS-T and 2% milk, slides were incubated overnight at 4°C in a humid chamber with primary antibodies in PBS-T and 2% milk: 1:500 mouse anti-H3K9me2 MABI0317 (MBL) and 1:500 polyclonal rabbit anti-RFP (Rockland). After washing 3 × 5 min with PBS-T, slides were incubated for 1 h at RT in a humid chamber with secondary antibodies in PBS-T and 2% milk: 1:1,000 goat anti-mouse Alexa Fluor 488 (A11001; Invitrogen) and 1:1,000 donkey anti-rabbit Alexa

Fluor 555 (A31572; Invitrogen). DNA was counterstained with DAPI (1:2,000) in PBS-T for 10 min, washed 3 × 5 min with PBS-T, and mounted with ProLong Gold Antifade (Thermo Fisher Scientific). SIM images were acquired using an ELYRA 7 LS mit 2x pco.EDGE 4.2 (Zeiss) and a Plan-Apochromat 100x/1,4 Oil objective lens.

We quantified nuclear foci using the KNIME Analytics Platform[52]. In summary, nuclei were detected using a seeded watershed segmentation using the DAPI channel (SIM images) or the green channel. For foci detection, we used a Laplacian-of-Gaussian detector from TrackMate[53] (fmi-ij2-plugins-0.2.5, https://doi.org/10.5281/zenodo.1173536) on the AMA-1 (GFP) channel. Potential foci outside of a nucleus were ignored in the analysis. Results were plotted per condition using the R package ggplot2 (https://ggplot2.tidyverse.org/).

RNA sequencing

RNA was extracted from early embryos from a synchronized culture grown at 20°C and exposed to indicated temperatures for 2 h, using Trizol as previously described 6. Embryos were freeze cracked five times, then RNA was extracted with choloroform followed by isopropanol precipitation. Further purification was performed with the RNA Clean and Concentrator kit (Zymo). Libraries were produced using the Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit (Illumina). rRNA was depleted using the Ribo-Zero Plus kit supplemented with custom made C. elegans specific rDNA oligos (IDT). Libraries were profiled in a DNA 1000 Chip on a 2100 Bioanalyser (Agilent technologies) and quantified using the Qubit 1x dsDNA HS Assay Kit, in a Qubit 4.0 Fluorometer (Life technologies). Equimolar amounts of indexed libraries were pooled and sequenced on a NextSeq 2000 (Illumina). Reads were analysed as described previously[54]. Adapter were trimmed using Trimmomatic v0.39. Reads were aligned to the *C. elegans* genome (ce10) with the R package QuasR v1.42.1, (`www.bioconductor.org/packages/2.12/bioc/html/QuasR.html`). The command `"proj <-qAlign("samples.txt","BSgenome. Celegans.UCSC.ce10", splicedAlignment=TRUE)"` instructs hisat296 to align using default parameters, considering unique reads for genes and genome wide distribution. Count Tables of reads mapping within annotated exons in WormBase (WS220) were constructed using the qCount function of the QuasR package to quantify the number of reads in each window `(qCount(proj, GRange_object,orientation="same"))` and normalized by division by the total number

of reads in each library and multiplied by the average library size. Transformation into log2 space was performed after the addition of a pseudocount of 8 to minimize large changes in abundance fold change (FC) caused by low count numbers. The EdgeR package v4.0.14 was applied to select genes with differential transcript abundances between indicated genotypes (contrasts) based on false discovery rates (FDR) for genes. Replica correlations are shown in Fig S4A of GO-Terms were extracted from wormbase.org and the GO-Term and KEGG pathway enrichment analysis was performed using the gprofiler2 package (v 0.2.2) as an interface to g: Profiler[55].

RESULTS

Molecular drivers of CTD phase separation and functional implications

We first established that residue-level coarse-grained molecular dynamics simulations[21] are a suitable model of the interactions of RNA pol II CTD (Fig. 1B-F) and in turn, simulations revealed how phase separation is driven by the interactions of the heptad repeats and how deviations from the ideal heptad sequence and truncation affect the tendency of CTD to phase separate. To ensure that our simulations capture the essential elements of the behavior of CTD, we compare to experiments and atomistic reference simulations reporting on the contact statistics of CTD at the level of condensates, inter-chain interactions, and intra-chain interactions (Fig 1B-F and Fig S1), which sets the scene of investigating functional aspects of CTD. In coarse-grained simulations, we find that the CTD (140mer with 20 ideal heptad repeats) phase separates on its own. Contacts between aromatic Tyr residues within the CTD are particularly prominent, as demonstrated in contact analysis of CTD (Fig 1E). These observations are consistent with previous experimental evidence highlighting the crucial role of hydrophobic interactions in driving phase separation[13,56].

Tyr-Try contacts are prominent in the interactions of two 21mer CTD fragments in extensive atomistic reference simulations ($100 \times 0.12\,\mu s$) (Fig S1C). By comparison of these atomistic reference simulations (Fig S1A, B Movie S1) of two CTD fragments to coarse-grained simulations, we could also show that our coarse-grained molecular dynamics simulations capture CTD interactions not only qualitatively, but also to the overall magnitude of interactions as expected from atomistic molecular dynamics simulations (Fig 1B, C). The

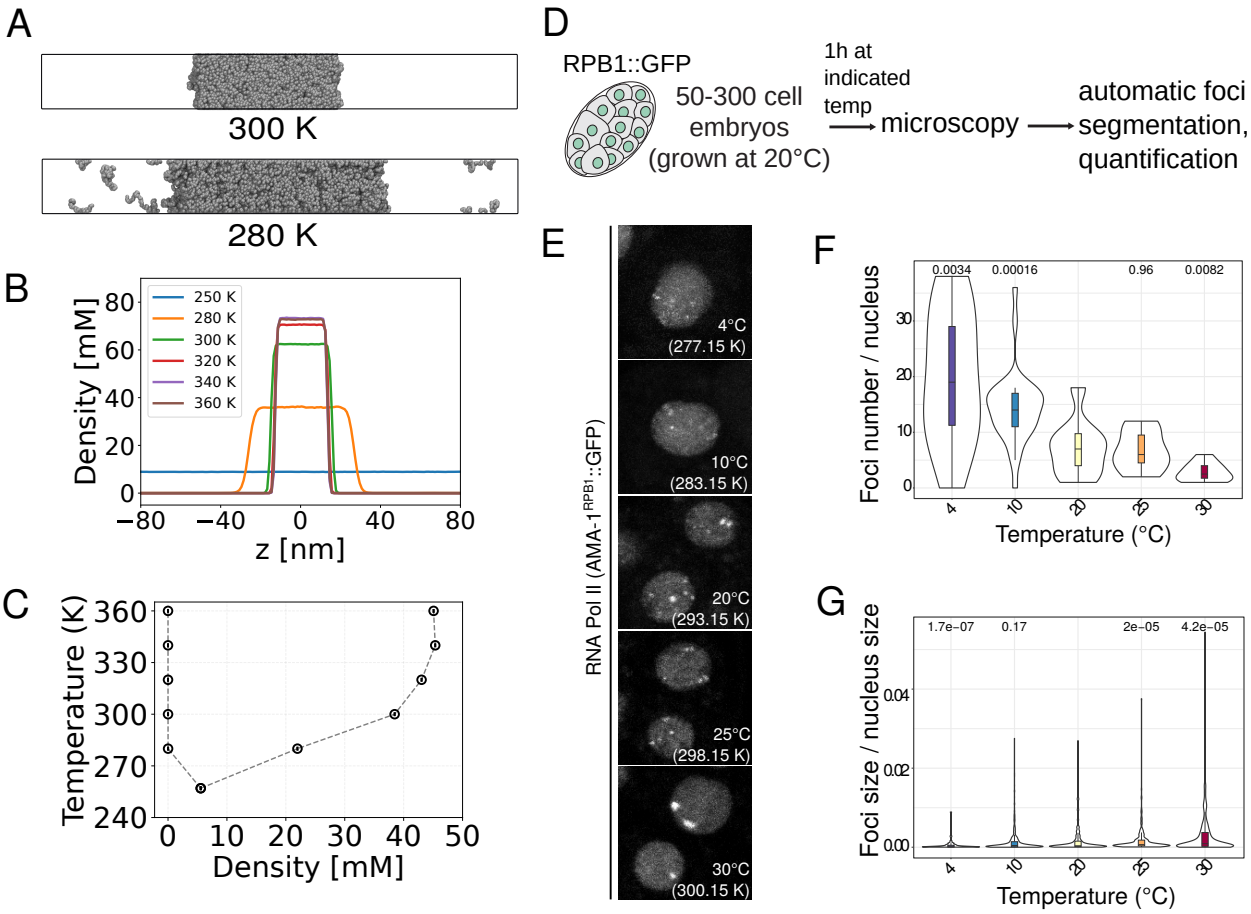

Figure 2. Temperature dependent phase behaviour of RNA Pol II CTD. A) Snapshots from coarse-grained simulations at 280 K and 300 K. The system phase separates at higher temperatures. B) The density profiles at different temperatures. C) Phase diagram illustrating LCST phase behavior. The uncertainties are indicated but are smaller than the circle markers. D) Experimental setup for investigating the effect of temperature on CTD condensate in *C. elegans* embryos at different temperatures. E) RNA Pol II foci in *C. elegans* embryos at different temperatures. As the temperature rises, the number of foci decreases, while their size increases. F) The number of foci per nucleus. G) The size of foci relative to the total nuclear volume.

affinity for CTD fragments for each other is estimated, assuming a kinetic rate model[57,58], to be $K_d = 2.61 \pm 0.01$ mmol and $K_d = 1.19 \pm .005$ mM for the atomistic and the coarse-grained simulations respectively (SI Text).

The contact analysis also highlights Tyr-Pro contacts (Fig 1C) which were recently highlighted by nuclear magnetic resonance (NMR) experiments and atomistic modeling to generate ensembles of all-atom CTD structures of human and yeast CTD, which account for

the experimental data (Fig 1F)[56]. Finally, we compare simulations to experiments with CTD truncations, which shows that our modeling captures functional relevant trends in the behavior of CTD. As anticipated, shorter CTD sequences exhibited less phase separation than longer ones[13] (Fig S2). Additionally, we observed that the actual sequences from *Drosophilia melanogaster*, *Saccharomyces cerevisiae*, and *Homo sapiens*, which all feature deviation from the heptad repeat pattern $Y_1S_2P_3T_4S_5P_6S_7$, to be less prone to phase separate than sequences of equal length which feature only copies of the ideal heptad repeat. In line with experiments, a truncated version of *Drosophilia* (20 repeats) CTD[59], which features only idealised CTD repeats phase separates about as much as full-length *Drosophilia* real CTD, whereas the full length ideal *Drosophilia* CTD showed a considerably higher tendency to phase separate.

CTD condensates dissolve at lower temperatures but are stable at room temperature in simulations and *C. elegans* unlike other nuclear condensates

Phase separated condensates are highly temperature dependent[36] and indeed measuring temperature dependent changes in condensate behaviour *in vivo* is one of the gold standards in the field[60]. To test our model, we therefore calculated the temperature dependent phase behaviour of RNA Pol II CTD (Fig 2A-C). We then compared the predicted temperature dependent changes in condensate behaviour to the behaviour of RNA Pol II condensates in *C. elegans* embryos, exposed to different temperatures. To quantify RNA Pol II condensates *in vivo* we imaged living embryos expressing the main RNA Pol II subunit RPB1 (ama-1 in *C. elegans*) tagged with GFP. Worms were cultured at 20°C (standard laboratory condition) and exposed for 1 h to a range of temperatures (4°C-30°C) (Fig 2D). Using automated image analysis we segmented RPB1::GFP condensates and quantified condensate number, size, and fluorescence intensity (Fig 2E-F). Strikingly we observed a high agreement between the simulations (Fig 2A-C) and the *in vivo* behaviour of RPB1 condensates. Interestingly, in contrast to most other described condensates which dissolve at higher temperatures and display an upper critical solution temperature (UCST), both the simulation model and the *in vivo* condensates display a lower critical solution temperature (LCST), resulting in reduced condensate formation at low temperatures and an increase in condensate formation at higher temperatures. Such LCST behaviour has recently been detected for RNA Pol II CTD in

vitro[56], which is in line with our *in vivo* experiments and simulations. LCST behavior has been thoroughly investigated for synthetic polymers and more recently for disordered proteins[61–65]. Many Pro-rich protein sequences show LCST behaviour. Water molecules are highly organised around Pro residues[66] and the formation of such water networks is associated with an entropic cost. At higher temperatures, the water network around Pro residues is less favourable and Pro is more prone to interact with other residues, favouring compact chain conformations and protein condensate formation. Other nuclear condensates, e.g., the heterochromatin condensates formed by the histone methyl transferase MET-2[54] and HP1α[67] dissolve with increasing temperatures and these differences in their response to temperature suggest that multiple distinct condensates co-exist in the nucleus.

Temperature dependent changes in embryonic gene expression are largely different from the classical stress response

The impact of the temperature shifts on the *in vivo* RNA Polymerase II condensates raises the question whether, or how these gradual changes in temperature affect the embryonic transcriptome. We therefore sequenced total RNA from embryos exposed to 10°C, 20°C, 25°C and 30°C for 2 h. When compared to embryos cultured at 20°C (standard laboratory conditions) we observed a modest downregulation of genes at 10°C, while exposing worms to 25°C resulted in a predominantly upregulation of genes, which was further pronounced at 30°C (Fig 3A, S4). Interestingly these transcriptional changes showed a linear correlation with temperature (Fig 3B). This is in contrast to what would be expected from a stress response where a dedicated signaling cascade triggers a defined transcriptional program upon reaching an activation threshold[54]. Indeed none of the temperatures result in an up- or downregulation of genes associated with known stress response pathways comparable with the effect that a 1 h heat shock at 37°C has. We note that there are two notable exceptions hsp-70 and hsp-16.41. Both of these are heatshock proteins that show a gradual increase in expression from 10°C-30°C (Fig 3C). Instead, the gene ontology terms (GO-Terms) enriched in the changed genes indicate that these small changes in temperature have a gradual impact on the expression of genes involved in metabolic processes.

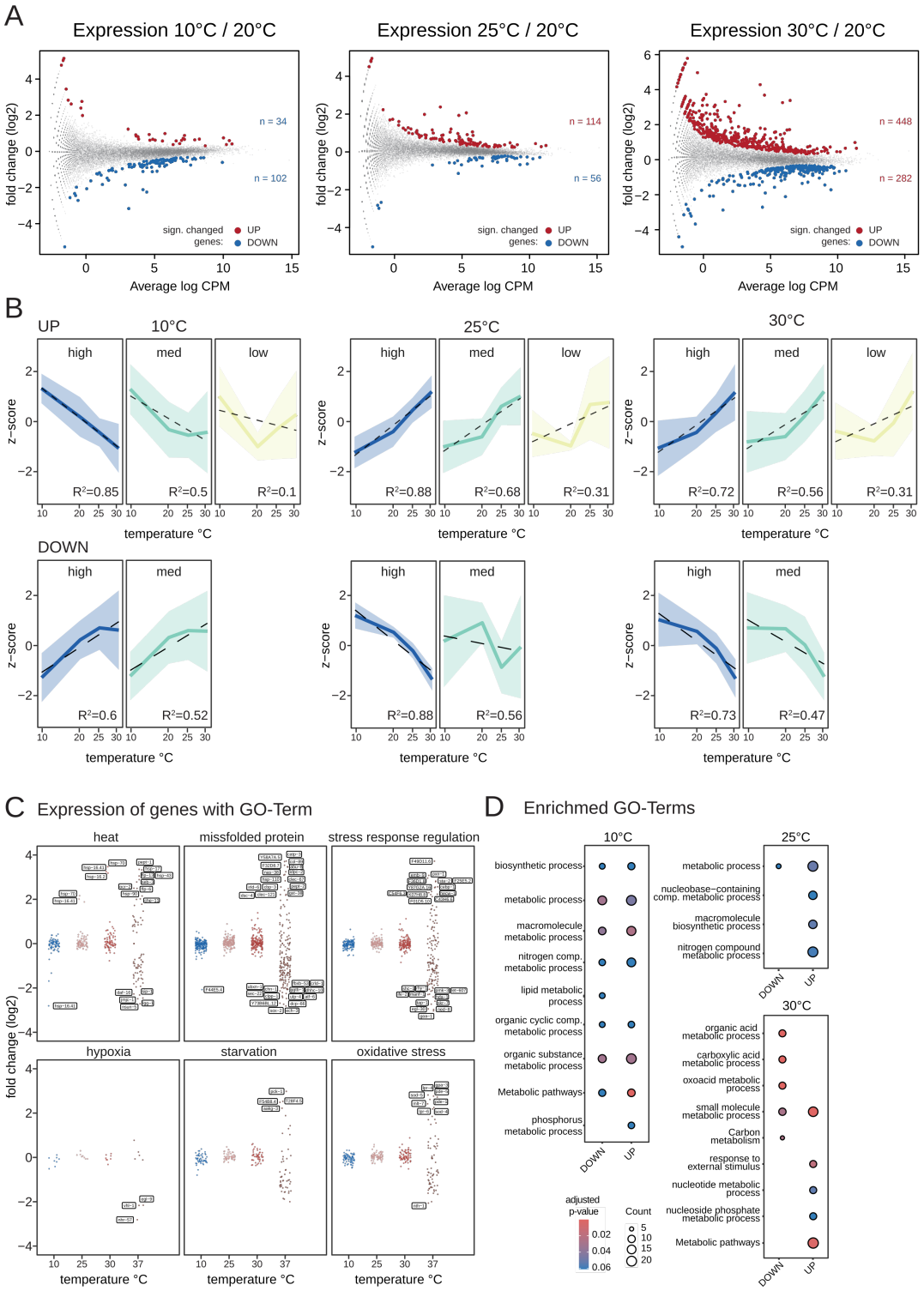

Figure 3. Exposing *C. elegans* embryos to a range of temperatures from 10°C-30°C results in a gradual transcriptional response.

Figure 3. (Previous page.) (A) Mean expression (CPM) versus expression change (log2 fold change) plots comparing transcriptome of embryos cultured at $20C$ (standard laboratory conditions) to embryos exposed to 10°C, 25°C, or 30°C for 2hrs. n indicates the number of deferentially expressed genes ($FDR < 0.05$), highlighted red significantly unregulated genes (fold change > 2), and blue significantly downregulated genes (fold change < -2) in all three plots.(Caption continues next page.) B) Expression changes of significantly up and downregulated genes identified in (a) at each temperature plotted as the mean z-score (solid line) and its standard deviation (shaded area). Dotted black line indicates a fitted linear regression. Each gene set was divided according to its mean expression level at 20°C (high$>$ 75 percentile; med$<$ 75 and $>$ 25 percentile; low$<$ 25 percentile). C) Expression changes as mean fold change (log2) over 20°C for genes with the indicated GO-Terms involved in a range of stress responses. Genes with a mean fold change > 2 or < -2 are labeled. Gene expression at 10°C, 25°C and 30°C are compared to expression changes at heat stress (37°C)[54]. D) GO-Term and KEGG pathway analysis of UP and downregulated genes.

Co-condensation of Ser_5 phosphorylated pCTD is driven by valence but not pattering

In addition to the separation of active from inactive regions in the genome, transcription itself occurs in multiple defined steps. The post-transcriptional phosphorylation state of RNA Polymerase II CTD is characteristic of each of these transcriptional stages[9,11] and has been implicated in condensate formation. Here we ask how the degree of CTD phosphorylation and the distribution of phosphosites impacts condensate formation. To specifically examine how phosphorylation influences phase separation, we conducted simulations using CTD and pCTD chains, each comprising 140mers (20 repeats of the ideal heptad repeats). We incrementally phosphorylated 50 CTDs in our simulations and in accordance with experiments, we find phosphorylation dissolves CTD condensates in a dose-dependent manner (Fig 4A). Phosphorylating up to 15 Ser_5 in the heptad motif out of a total of 20 Ser5 in the 140mer pCTD, still predominantly leads to partitioning into the dense phase of CTD. However, as the phosphorylation level of pCTD increases above the threshold of 15 Ser_5, the dilute concentration of pCTD rises gradually as the electrostatic repulsion between pCTDs in the dense phase dominates and expels the pCTD chains into the dilute phase (Fig 4B). The contact map illustrating pCTD-CTD interactions (Fig S3B-C) highlights the relatively weak

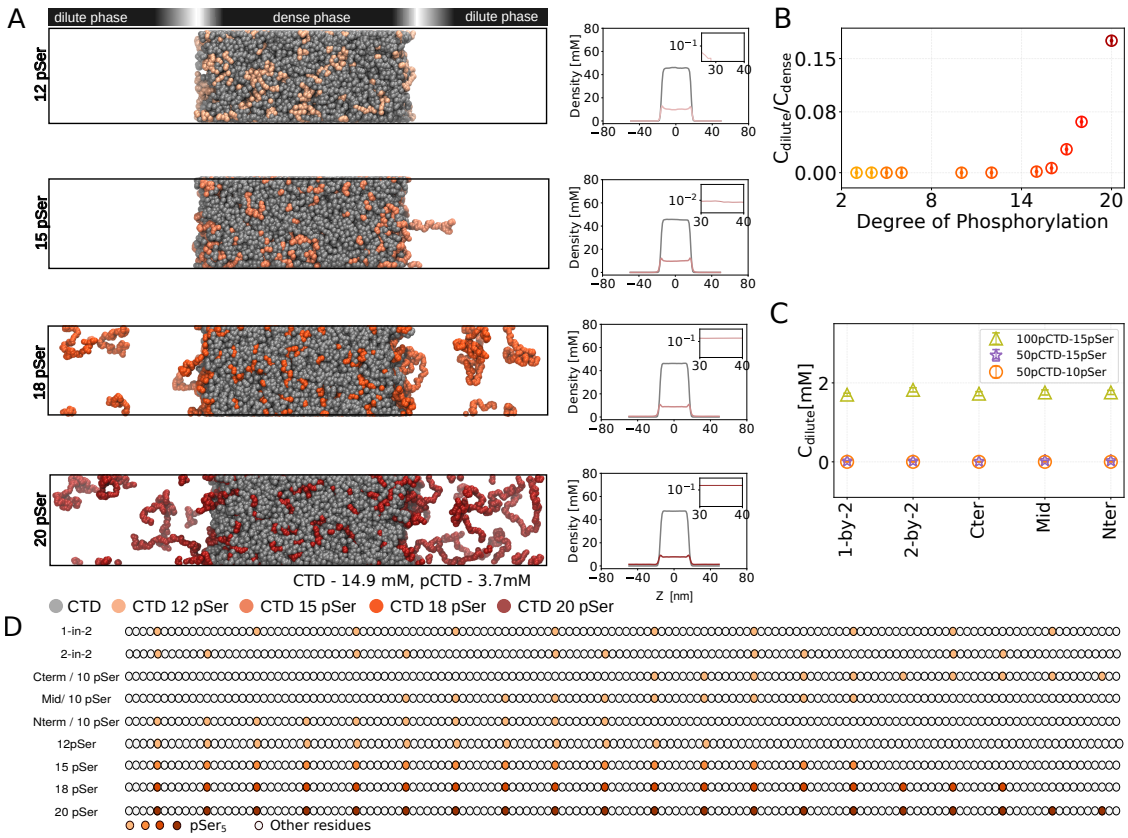

Figure 4. Phase behavior of 100 CTD and 50 pCTD chains. Visualization of CTD and pCTD in molecular dynamics simulations in slab geometry of box dimension 15 nm×15 nm×100 nm. A) The degree of phosphorylation of pCTD chains varies in each simulation. Density plots for simulation with CTD and pCTD with 10pSer, 12pSer, 18pSer, and 20pSer respectively. B) Quantifying c_{dilute}/c_{dense} of partially phosphorylated pCTD chains as the Ser_5 phosphorylation increases. C) Variation in c_{dilute} of pCTD with pSer residues distributed in five distinct patterns. D) Distribution of $pSer_5$ residues in the 140mer pCTD.

nature of the intermolecular binding between the CTD and pCTD with all the 20 Ser_5 phosphorylated, as compared to the intermolecular interactions observed between CTD-CTD. pCTD with each heptad repeat phosphorylated does not phase separate on its own.

To explore the impact of different patterns of distributing an identical number of phosphorylated Ser_5 residues on CTD sequences, we conducted simulations with partially phosphorylated pCTD chains, each carrying different distribution of pSer, in the presence of a CTD condensate and compared the variation in dilute concentration of pCTD (Fig 4C, D). Three sets of simulations were conducted: 50 pCTD chains with 10 pSer residues, 100 and,

50 pCTD chains with 15 pSer residues. We considered five distinct distributions: maximally distributed in every second heptad repeat, in adjacent heptad repeats, in the middle, and in N and C-termini, respectively. For pCTD chains with 10 phosphorylation values (below the threshold), rarely any pCTD chains were found outside the CTD condensate, regardless of their distribution. With 15 pSer and 50 or 100 pCTD chains (above the threshold), the dilute concentration remained nearly the same value for different distributions (Fig 3C). In conclusion, our simulations show a phosphorylation dependent demixing of unphosphorylated and phosphorylated CTD. The degree of demixing is thereby dependent on the number of phosphorylation sites, but not their distribution along the CTD heptad repeats.

Thus, our simulations can explain why in vitro CTD condensates are dissolved over time in the presence of the kinase CDK7 and ATP[13]. However, this also poses a clear contradiction with all observations made in cells[14,19]. Where transcribing RNA Polymerase CTD localizes into condensates even when it is highly phosphorylated[15]. A likely explanation for this discrepancy is that in cells the phosphorylation of CTD can create binding sites for multiple other interacting proteins[14]. We therefore wondered if we can simulate the impact of additional factors in our model. As a first indicator, we consider the interactions of CTD and pCTD with FUS condensates. In our simulations the unphosphorylated CTD binds more strongly to FUS low complexity domain (LCD) condensates than pCTD (Fig S5), which is in line with experiments that demonstrated that CTD phosphorylation abrogates the binding of CTD to FUS gels and condensates. Interestingly while FUS and CTD both localize to a single dense phase, they are not homogenously mixed in the condensate. We further observed that FUS results in a minor enrichment of pCTD in the dense phase. Together this shows that we can indeed expand our simulation to encompass CTD interaction partners.

Co-phase separation pCTD and HRD by complex co-acervation

FUS mainly interacted with the non-phosphorylated form of CTD. Inside the cell, numerous interaction partners of CTD might shield the effect of CTD phosphorylation. One protein with a well described role in transcriptional elongation is Cycline T1. As part of the positive transcription elongation factor b (P-TEFb) Cycline T1 stabilizes elongating RNA Polymerase II and it is essential for pCTD condensation in vitro and potentially also in vivo. In agreement with experiments, simulations show that pCTD co-phase separates with the

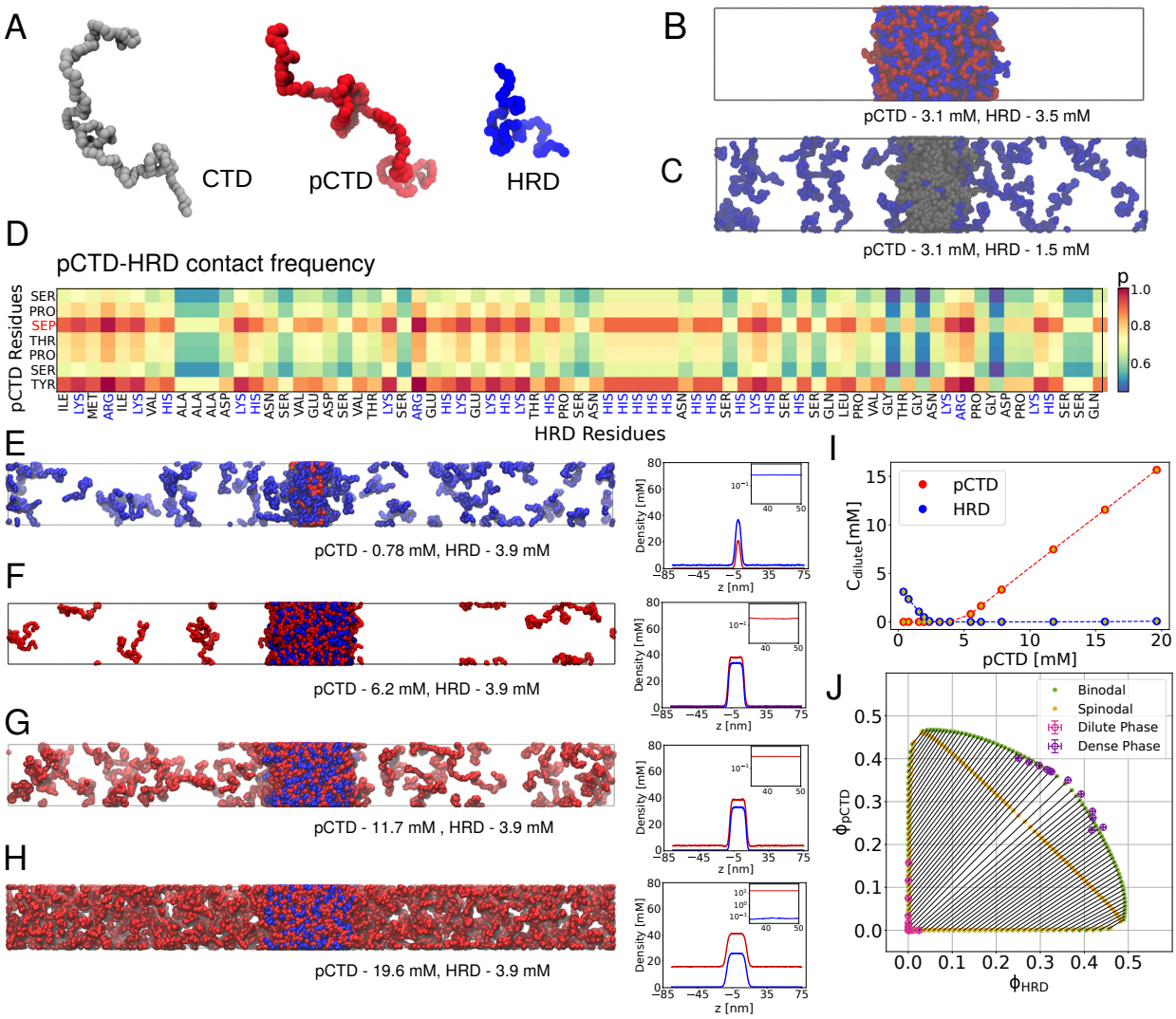

Figure 5. Co-phase separation HRD with CTD and pCTD in slab simulation. A) Single chains of CTD, pCTD, and HRD. B) Simulation snapshot showing pCTD and HRD forming condensate. C) Simulation snapshot of CTD with HRD. While CTD forms a condensate, HRD exhibits no favorable interaction with CTD. D) Contact frequency analysis between pCTD and HRD showing the important electrostatic interactions between positive (blue) and negative (red) charged residues. E-H) Simulation snapshots depicting the effect of increasing pCTD while maintaining a constant concentration of HRD in a slab simulation. Density profiles for pCTD and HRD correspond to the simulations shown on the right side. I) Variation of dilute concentration of pCTD and HRD as the pCTD concentration increases. J) Biphasic diagram of the pCTD-HRD system. Positive tie lines indicate associative phase separation. The volume fractions from molecular dynamic simulations are in magenta and purple. Binodal and spinodal from mean field fit are shown in green and yellow dots.

histidine rich domain (HRD), which is a disordered domain of P-TEFb Cyclin T1[14].

HRD, which is positively charged on account of its histidine side chains, on its own does not interact strongly with CTD (Fig 5C). CTD chains in our simulation are uncharged since we are simulating an idealized version of the CTD with no deviation from the canonical heptad $Y_1S_2P_3T_4S_5P_6S_7$ repeat motifs. In the pCTD and HRD condensates (Fig 5B), pSer residues of pCTD interact strongly with positively charged residues Arg, Lys, His in HRD. pSer forms few contacts with Ala, Ser, Thr, and Gly residues of HRD (Fig 5D). Tyr in pCTD also engages in strong contacts with Lys and Arg in HRD and to a somewhat lesser degree with many other residues, e.g., Ile, Met, and Val, but interacts only weakly with Ser and Gly. The contact analysis clearly shows that context matters a lot[68], a negatively charged Asp next to a positively charged Lys residue is more frequently in contact with pSer than an uncharged Ala residues in a stretch of Ala residues.

To understand the interaction between pCTD and HRD, we conducted simulations to construct a phase diagram, which shows that the pCTD and HRD form complex coacervates through electrostatic interactions. Throughout these simulations, we kept the concentration of HRD constant at 3.9 mM while systematically increasing concentrations of pCTD (Fig 5E-H). Considering that pCTD and HRD are oppositely charged we expect them to interact electrostatically and form a complex coacervate. This type of phase separation, also known as cross-interaction-driven phase separation (CIP), happens when two species are mutually attracted to each other and cannot phase separate independently[69]. CIP typically gives rise to a looped phase diagram, since the binodal lines do not intersect with the composition axes. The relative concentrations of the proteins are critical for determining the extent of phase separation and the location of the coexistence region. Notably, a rise in the concentration of either protein can initiate phase separation, leading to the formation of condensates. This occurrence is depicted in Fig 5I, where it is shown that, at an approximate concentration of around 5 mM of pCTD, all the HRDs in the dilute phase undergo phase separation. Further increasing the pCTD concentration has no effect on the dense phase, rather, the dilute concentration of pCTD increases (Fig 5F-H). The saturation concentration of HRD gradually decreases with increasing pCTD concentration, as HRDs initially present in the dilute phase transition into the dense phase by associating with added pCTD chains. This observation becomes even clearer when examining the ternary phase diagram depicted in Fig 5J. Here, the pink and purple dots represent the simulated coexisting (binodal) dilute

and dense phases respectively. Phase separation is associative, evidenced by the positive tilt in the (imaginary) lines that connect the coexisting phases in composition space (tie-lines). This is further supported by fitting the simulated binodal compositions in Fig 5J using Flory-Huggins theory (Eq. 4) (SI Text), which is shown by green dots. The fit also strengthens our interpretation that self interactions of hyper-phosphorylated CTD and HRD are negligible due to electrostatic repulsion. It is consistent with the observation that the dilute phase is devoid of polymers of one type in the simulations, but enriched by polymers of the other type if their average volume fraction increases (see also Fig 5E-H), which is also expressed by the tie-lines which develop an off-set at either axis as soon as the mean composition ratio starts deviating from unity (Fig 5J, grey).

We conclude from the simulations and from fitting the Flory-Huggins model to the simulation data that, unlike CTD condensates where Tyr-Tyr and Tyr-Pro contacts drive phase separation, pCTD, and HRD form co-acervates by electrostatic interactions, which suggests that on account of differences in molecular driving forces CTD and pCTD-HRD phases could in principle constitute distinct environments in the nucleus. The excellent fit with the simulated data, despite the fact that Flory-Huggins theory does not include a description for (long range) electrostatic interactions, supports considerable screening by the implicit background salt in the simulations, making the attractive forces between the polymers effectively short-ranged. Furthermore, owing to the predominance of a single type of interaction, a more complex mean-field model for fitting the data[70] would be redundant. The overall attraction between the polymers is captured by a negative polymer-polymer interaction parameter: $\chi_{AB} = -2.36$. The polymer-solvent interaction parameters (see Eq. 4), as obtained from the fitting, are $\chi_{AS} = 0.55$ and $\chi_{BS} = 0.36$, indicating relatively poor solubilization (near theta conditions) of, in particular, the HRD.

Phosphorylation can act as a switch and trigger the formation of a second dense phase in the co-phase separation of CTD, pCTD, and HRD

In separate simulations, increased phosphorylation disrupts the phase separation of pCTD. However, the binding of HRD to pCTD facilitates its phase separation. Yet, within cells, CTD, pCTD, and HRD coexist concurrently, raising the question if they form a single mixed phase or segregate into two distinct dense phases. Thus we have simulated fixed

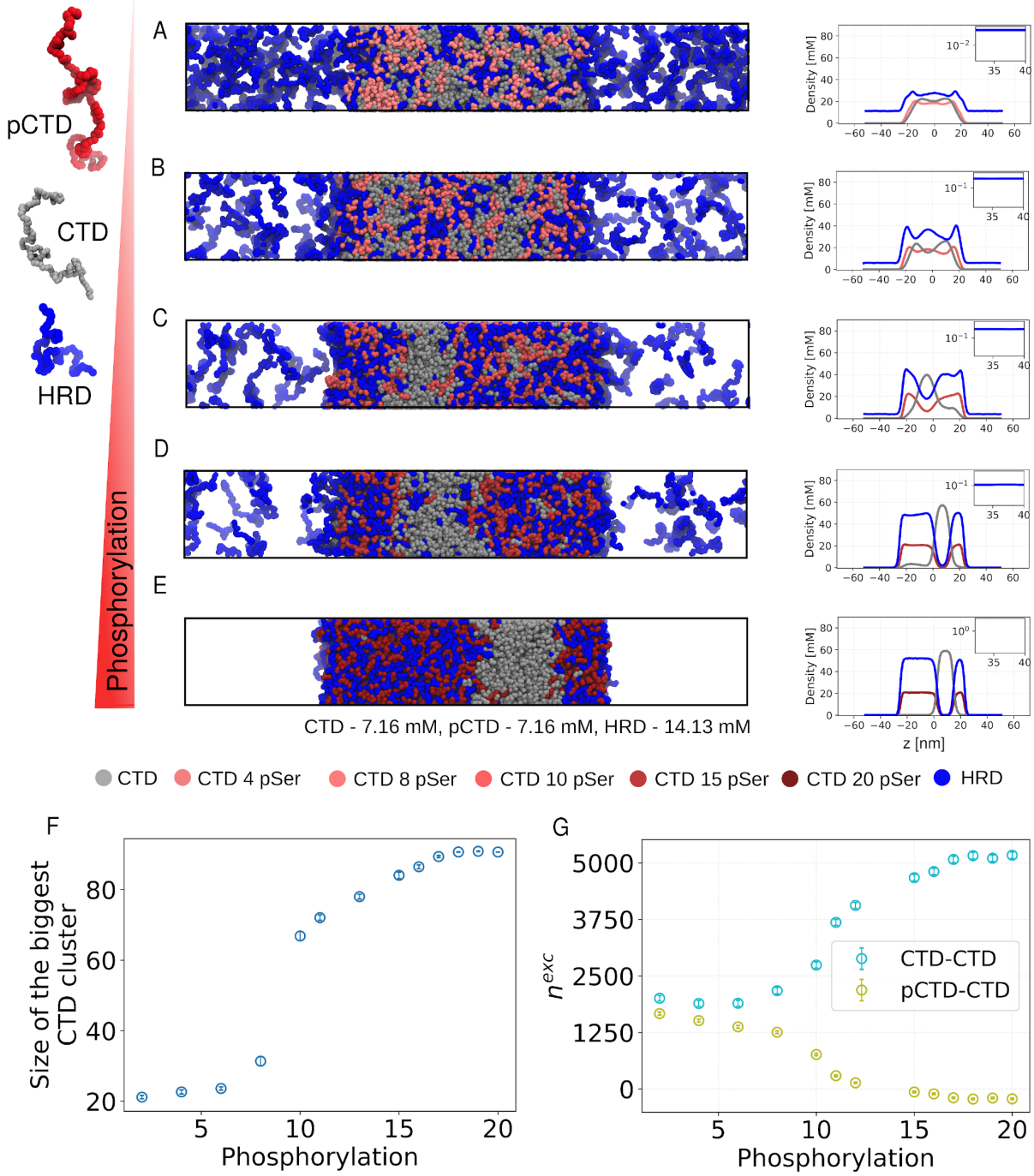

Figure 6. Mixing and de-mixing of the dense phase of 100 CTD, 100 pCTD, and 250 HRD chains, with increasing the degree of pCTD phosphorylation. A) 4pSer B) 8pSer C) 10pSer D) 15pSer E) 20 pSer of pCTD are phosphorylated. F) The size of the biggest CTD cluster as the phosphorylation of pCTD varies in the simulation. G) Comparison of the excess of CTD monomers surrounding a phosphorylated pCTD (CTD-pCTD) and surrounding another CTD (CTD-CTD), with increasing phosphorylation levels, denoted as n^{exc}. This metric quantifies the number of protein-protein contacts and is calculated using Equation 12.

numbers of CTD, pCTD, and HRD and increased the degree of pCTD phosphorylation in a series of simulations. Partially phosphorylated pCTD containing 4 $pSer_5$ co-phase separates with CTD, as it is apparent from the density profile (Fig 6A). The density profile illustrates that both pCTD, CTD, and HRD chains are present within the mixed condensate, while a significant portion of HRD is in the dilute phase. However, when the degree of phosphorylation of pCTD increases and 8 Ser_5 are phosphorylated (Fig 6B) the amount of HRD in the dilute phase decreases as it is now recruited to the dense phase by the pCTDs. There is a tendency for pCTD and HRD to co-localize as apparent in the density profiles. With 10 $pSer_5$ phosphorylated, the HRD concentration in the dilute phase is markedly reduced further and the co-localization of pCTD and HRD is enhanced (Fig 6C). The co-localization of pCTD-HRD is observed to favor the periphery of the condensate border (Fig 6B-E). This preference for condensate periphery localization could be attributed to the presence of negative charges on pCTD and positive charges on HRD, potentially facilitating interactions with the solvent compared to the uncharged CTD chains. Earlier simulations employing the same coarse-grained approach have also shown microstructuring in condensates due to electrostatic interactions[50]. With further increasing phosphorylation of pCTD, there is an enhanced tendency for both pCTD and HRD to localize towards the boundaries of the CTD-pCTD-HRD condensate (Fig 6C-D). Unphosphorylated CTD is clearly at the center of the condensate. Adding more phosphates to pCTD results in a core-shell morphology, with the CTD density mostly at the center of the condensate, with steep decreases in CTD density towards the boundaries. De-mixing is not complete, since some CTD chains co-localize with pCTD-HRD as is apparent from both visualisation of the simulation and the density profile. Phosphorylating all Ser_5 residues in the 140mer pCTD chains (Fig 6E), further enhances the de-mixing.

The size of the largest CTD cluster in simulation shows a sharp transition when about half of the Ser_5 of pCTD chains are phosphorylated, indicating the onset of the separation process into two distinct phases: CTD and pCTD-HRD (Fig 6F). The microphase separated domains of the homopolymer CTD phase and the heteropolymer pCTD-HRD phase begin to evolve as the degree of phosphorylation increases. This change becomes more pronounced with increasing phosphorylation, leading to an expansion in the size of these domains. The excess of CTD in the vicinity of CTD and pCTD (n^{exc}) also shows a similar transition as phosphorylation increases (Fig 6G). The parameter, n^{exc}, reveals how the number of

nearby pCTD and CTD chains within a specific distance from a CTD chain changes with increasing phosphorylation of pCTD. It shows changes in the local environment, providing insights into the frequency and strength of interactions between CTD and pCTD chains as the phosphorylation of pCTD varies. At lower phosphorylation levels, a CTD chain is surrounded by both pCTD and CTD, and the numbers of CTD and pCTD are consistent. However, a noticeable deviation becomes apparent as the degree of phosphorylation surpasses 10 $pSer_5$. Additionally, the overall structure and shape of pCTD chains are affected by phosphorylation (Fig S6). In dilute solution, phosphorylation expands the protein in line with experiments[71] as the chains become less globular and more rod-like due to electrostatic repulsion. Similarly, in the dense phase, pCTD becomes more rod-like. The radius of gyration (R_G) linearly increases with phosphorylation below a threshold of 10 $pSer_5$. Once this threshold is crossed, R_G stabilises and the anisotropy of pCTD chains decreases as interactions with HRD compensate for electrostatic repulsion. This shift once again highlights the impact of phosphorylation in modulating specific interactions between the protein pairs in forming two distinct dense phases.

Morphology of CTD and pCTD condensates in simulations and *C. elegans* embryos

The simulation reveals a core-shell arrangement. By calculating interfacial tensions through our molecular dynamics simulations[22,72], we examined the potential morphologies these condensates could take on and explained the reasons why such a core-shell structure is possible. When two immiscible droplets of CTD and pCTD-HRD are brought into contact, four distinct morphological configurations become possible[6,73]. We visually represent these four distinct morphologies using a stability diagram generated from the contact angle equation, as depicted in Fig 7D. The contact angle equation is given by:

$$\theta_{[\text{pCTD-HRD}]} = \arccos\left(\frac{\gamma^2_{[\text{CTD}]} - \gamma^2_{[\text{pCTD-HRD}]} - \gamma^2_{[\text{pCTD-HRD}]\text{CTD}}}{2\gamma_{[\text{pCTD-HRD}]}\gamma_{[\text{pCTD-HRD}]\text{CTD}}}\right) \tag{14}$$

. In Figure 7D, a schematic depicts partially engulfing CTD and pCTD-HRD phases, illustrating the contact angles between these phases. The stability diagram describes the four different morphologies attainable by adjusting relative interfacial tensions, and these variations can be characterised through the corresponding contact angles. In Region I, the

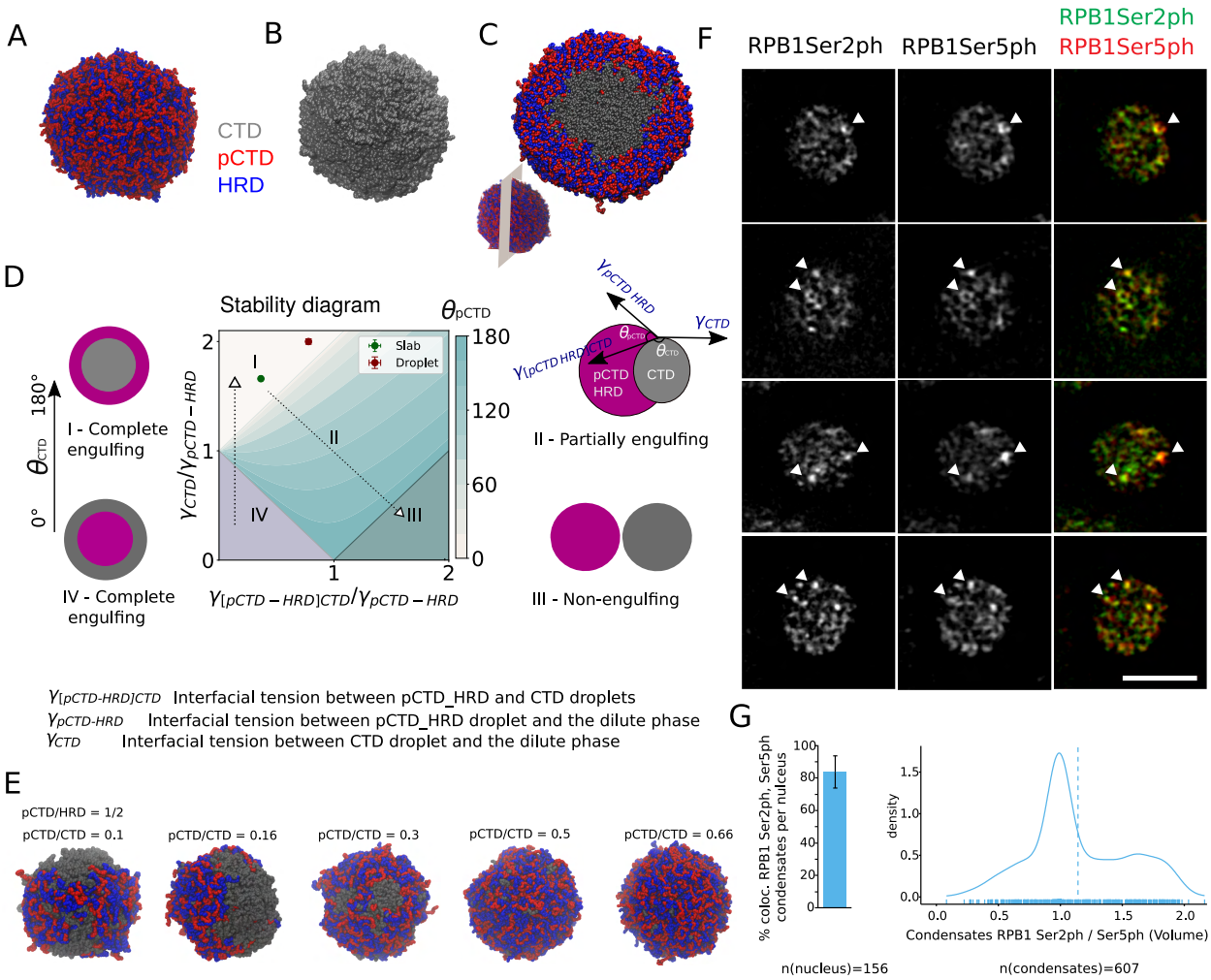

Figure 7. Morphology of the multiphasic condensate A) Formation of a droplet composed of 900 pCTD and 1800 HRD chains. B) Droplet comprising 600 CTD chains. C) Multiphasic droplet with a core-shell morphology, constituted by 900 pCTD, 1800 HRD, and 600 CTD chains. D) Stability diagram illustrating four distinct morphologies arising from the interaction between CTD and pCTD-HRD droplets. Interfacial tension analysis from slab and droplet simulations indicates the core-shell morphology. Sketch illustrating a partially engulfing CTD and pCTD-HRD phases indicating the contact angles. E) CTD condensate partially enveloped by pCTD and HRD condensate. Sequential images depict the progression from partially engulfed to fully engulfed morphology as the ratio of pCTD/CTD chains increases. F) Super-resolution microscopy image of *C. elegans* embryos stained with antibodies targeting phospho-Serine 5 (pSer5, indicating transcriptional initiation) and phospho-Ser2 (pSer2, indicating transcriptional elongation) phosphorylated RPB1. RPB1 labeled with either PTM is localized in prominent subnuclear foci. G)The segmentation and quantification of these foci demonstrate that both types of foci colocalize, exhibiting only partial overlap.

CTD phase is entirely enveloped by the pCTD-HRD phase, with the condition $\gamma_{[\text{CTD}]} > \gamma_{[\text{pCTD-HRD}]} + \gamma_{[\text{pCTD-HRD}]\text{CTD}}$. In Region II, the CTD and pCTD-HRD phases are partially engulfed, sharing an interface while also being exposed to the dilute phase. The transition from complete engulfment to non-engulfing phases is gradual as $\theta_{[\text{pCTD-HRD}]}$ increases. Here, the three interfacial tensions are comparable: $\gamma_{[\text{CTD}]} \sim \gamma_{[\text{pCTD-HRD}]} \sim \gamma_{[\text{pCTD-HRD}]\text{CTD}}$. In Region III, the significant interfacial tension between the CTD and pCTD-HRD droplets leads to a non-engulfing morphology, defined by $\gamma_{[\text{pCTD-HRD}]\text{CTD}} > \gamma_{[\text{CTD}]} + \gamma_{[\text{pCTD-HRD}]}$. In Region IV, the CTD phase engulfs the pCTD-HRD phase. The transition from Region 4 to Region I is marked by a black arrow, signifying an increase in $\theta_{[\text{CTD}]}$.

Full engulfment of CTD but pCTD-HRD is the most favourable morphology, but small changes in the system would enable partial engulfment. We used capillary wave theory (CWT)[72] to compute the interfacial tensions of CTD and pCTD-HRD with slab geometry with each Ser_5 in the heptad repeats of pCTD phosphorylated. Three sets of slab simulations were performed to determine interfacial tension at different interfaces. The first two simulations involved CTD and pCTD-HRD being simulated separately, with each forming its condensate, from which the interfacial tension at the interface with the dilute phase can be calculated. In the third simulation, CTD and pCTD-HRD were simulated together, forming a multiphasic condensate, and we computed the interfacial tension at the interface between the CTD and pCTD-HRD condensate (SI Text and Fig S7). Evaluating the ratio of interfacial tension for different interfaces, specifically $\gamma_{[\text{CTD}]}/\gamma_{[\text{pCTD-HRD}]}$ and $\gamma_{[\text{pCTD-HRD}]\text{CTD}}/\gamma_{[\text{pCTD-HRD}]}$, we confirmed that the observed core-shell morphologies are consistent with the theory of liquid interfaces (depicted by the green dot in Fig 7E). This supports our conclusion that the core shell morphology in our simulations is stable state and our observation is not an artifact of insufficiently long simulations. To rule out effects from the box geometry, we repeated the simulations in large cubic boxes $150\,\text{nm}\times150\,\text{nm}\times150\,\text{nm}$ with up to 3300 proteins chains, where the proteins form spherical droplets (Fig 7A-C, Movie S2). The ratios of interfacial tension between different interfaces, $\gamma_{[\text{CTD}]}/\gamma_{[\text{pCTD-HRD}]}$ and $\gamma_{[\text{pCTD-HRD}]\text{CTD}}/\gamma_{[\text{pCTD-HRD}]}$, were computed from the shape fluctuations method[22] (indicated by the red dot in Fig 7D). These results further confirm that the core-shell morphology is the most favorable arrangement for the CTD, pCTD-HRD system in our simulations. The comparison of interfacial tension at different interfaces using various methods is illustrated in Fig S8, demonstrating consistent values across the different approaches. We note that

the computed ratios of surface tensions are closer to the border between full and partial engulfment. We can also observe partial engulfment of CTD in simulations where there are not enough pCTD and HRD chains to fully cover the CTD surface (Fig 7E, Movie S3). Simulations clearly show that the condensate of pCTD-HRD either fully or partially surrounds the CTD condensate (Fig 7C, E). At lower concentrations of pCTD and HRD, they undergo coacervation and form a small patch on the surface of the CTD condensate rather than spreading evenly across the surface, resulting in a partially engulfed structure. With an increase in the concentration of pCTD and HRD, the degree of engulfment also increases, indicating that higher concentrations of pCTD and HRD lead to a decrease in interfacial tension between the pCTD-HRD complex and CTD[74].

The simulations thus suggest that RNA Polymerase II condensates in cells might be segregated into partially, or fully engulfed condensates, which we test in in vivo by super-resolution microscopy of *C. elegans* embryos. We stained the *C. elegans* embryos with antibodies against phospho-Serine 5 ($pSer_5$, as a marker for transcriptional initiation) and phospho-Ser2 ($pSer_2$, as a marker for transcriptional elongation) phosphorylated RPB1. RPB1 marked with either PTM can be found in large subnuclear foci (Fig 7F). Segmentation and quantification of these foci revealed that both foci types colocalize, but only partially overlap (Fig 7G). This is consistent with the partial or full engulfment of two distinct condensates, as seen in the simulations. Indeed a fraction of the $pSer_2$ and $pSer_5$ RPB1 foci formed hollow spheres, or half-moon like structures (Fig 7F). Given that total RPB1::GFP formed solid round foci (Fig 2E and Ref[75]), we conclude that these hollow spheres are part of a bigger RNA Pol II condensate. This highlights partial and full engulfment of CTD by pCTD in remarkable agreement with our simulations (Fig 7F) and hints at the prevalence of patterns of RNA polymerase II in different phosphorylation states.

Consistent with our simulations CTD condensates are stable *in vivo*, even when Cyclin T1 is knocked down by RNA interference (RNAi) and Ser_2 phosphorylation is consequently abolished[76] (Fig S9). This shows that CTD condensates can form in the absence of Ser_2 phosphorylation as in our simulations. The condensates are generally more round than in the presence of Cyclin T1. A more regular round appearance is consistent with the more spherical shapes of CTD condensates in contrast to more irregularly shaped condensates of CTD, pCTD, and HRD.

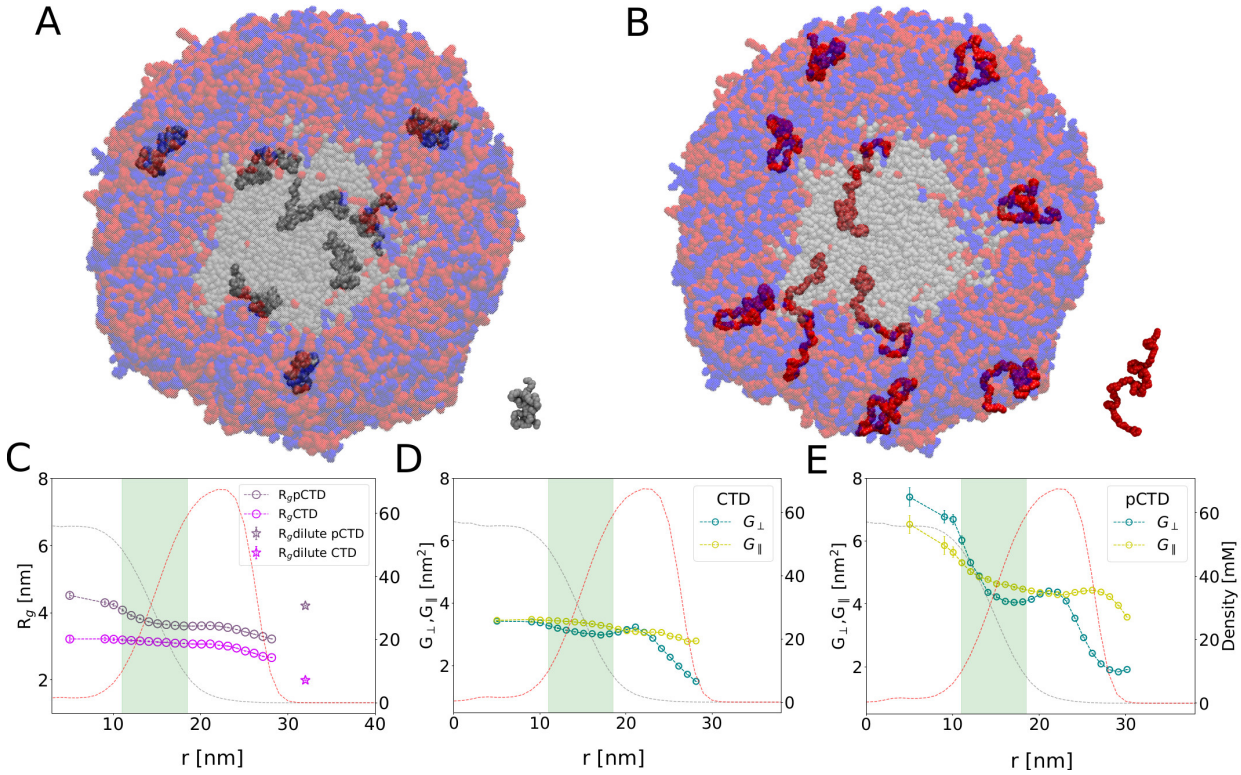

Figure 8. Conformation of CTD and pCTD within the droplet. A) Snapshot capturing CTD conformations, where the R_G of CTD chains reduces slightly in the pCTD-HRD phase. In the dilute phase, the chains become more compact. B) Snapshot showing pCTD conformations. The chains are more elongated in the CTD phase and compact in the pCTD-HRD phase. The chains are elongated in the dilute phase. C) Comparison of the Variation in R_g (radius of gyration) between CTD and pCTD Conformations, Analysed from the Center of the Condensate. D) Changes in R_G and diagonal components of the gyration tensor for CTD chains, measured from the droplet's center. E) Variation in R_G and diagonal components of the gyration tensor for pCTD chains, measured from the droplet's center.

The conformations of the CTD and pCTD chains are modulated by the two condensates phases

The conformations of CTD and pCTD are modulated by the different phases and interfaces within the condensate (Fig 8A, B) as the CTD and pCTD-HRD phases indeed have different properties. These differences could underpin distinct regulatory processes mediated by the two phases within a biphasic condensate. Recent theoretical and experimental

advances show the profound impact of condensates on the global conformations of proteins within them[77,78], including their orientations and extensions. The conformation of a protein chain within a phase can be understood with the concept of solvent quality from polymer science[50,79]. In a "good" solvent, which is compatible with the chain, the chain is extended, whether the solvent is a dilute solution or a dense phase. By contrast, a chain tends to collapse in a "poor" solvent which is not compatible with a chain. CTD is more expanded in the condensate than in a dilute aqueous solution, as tracked by the radius of gyration (R_g) (Fig 8C). The aqueous phase is a poor solvent for CTD chains since the ideal YSPTSPS heptads are devoid of charged groups and feature aromatic and non-polar residues.

To understand the conformation of chains in the condensate, we computed parallel and perpendicular components of the gyration tensor. The components $\langle G_{\parallel} \rangle$ are aligned parallel to the interface, while $\langle G_{\perp} \rangle$ represents the component that is oriented perpendicular to the interface. In the center of the CTD phase, the extended CTD chains (Fig 8A) are neither aligned parallel nor perpendicular to the condensate interface, with the components of the radius of gyration tensor,[77] $\langle G_{\perp} \rangle$, $\langle G_{\parallel} \rangle$ adopting similar values (Fig 8D).

At the interface of the CTD phase, on average, CTD chains align parallel to the interface, with $\langle G_{\perp} \rangle$ smaller than $\langle G_{\parallel} \rangle$ (Fig 8A, S11A)[80]. Occasionally, some CTD chains visit the pCTD-HRD phase but their R_G is barely affected, even though the phase is enriched in positive and negative charges and thus CTD experiences a reduction in solvent quality. Close to the interface of pCTD-HRD with the dilute aqueous phase the CTD chains collapse as they preferentially interact with the pCTD-HRD phase rather than with the dilute phase.

The conformation of pCTD is modulated by the CTD and pCTD-HRD phases in a different way. Hyperphosphorylated pCTD is more extended in the dilute aqueous phase than in the condensate (Fig 8B, E, and SI Text). The degree of phosphorylation greatly affects the global conformation of pCTD in solution, whereas the conformation of pCTD in a dense phase changes much less with phosphorylation due to its interactions with positively charged HRD (Fig S6). While CTD experiences minimal conformational changes upon transitioning from its parent phase to the pCTD-HRD phase, pCTD undergoes further expansion upon leaving its parent phase and entering the CTD phase. When pCTD is within the CTD phase, it extends even further, with $\langle G_{\parallel} \rangle$ being reduced compared to $\langle G_{\perp} \rangle$, indicating a slight preference for alignment perpendicular to the interface. This preferential alignment perpendicular to the interface enables pCTD to maximize its interactions with

positively charged HRD residues in the core. pCTD chains are squeezed at the interface with the dilute phase and $\langle G_{\parallel} \rangle$ are larger than $\langle G_{\perp} \rangle$ as the chains align parallel to the interface. We conclude that the two phases can significantly affect chain conformation due to their different chemical environments and this suggests that the two condensate phases could recruit molecules differentially and underpin specific regulation.

Recruitment of FUS and partially phosphorylated CTD to multiphasic condensates

Having established that the two phases of condensates of CTD, pCTD, and HRD are distinct chemical environments, we investigated whether they could indeed differentially interact with other molecules, which would enable each phase to underpin different stages of transcription. We observed contrasting interaction patterns between CTD and pCTD-HRD with phase-separated LCD FUS. CTD interacts with phase-separated FUS as apparent from snapshots of simulation systems of CTD, pCTD-HRD, and FUS (Fig S12). Regions of the simulation system enriched in pCTD-HRD are relatively depleted in FUS. This preference for the CTD phase over the pCTD-HRD phase is consistent with the proposed roles of CTD chains and CTD condensates in supporting transcription initiation by interacting with transcription factors to form promoter condensates and the experimentally established preference for FUS LCD to interact with CTD over pCTD[81,82]. We checked the partitioning of partially phosphorylated pCTD to multiphasic condensates and found that partially phosphorylated CTD can interact with both the CTD and pCTD-HRD phases (Fig S13). The partially phosphorylated CTD is relatively depleted at the peaks of the pCTD and HRD concentration profiles and shows how a partially phosphorylated CTD can reside in the condensate phase associated with initiation and the RNA polymerase II would only switch to the other phase upon further phosphorylation.

DISCUSSION

With molecular dynamics simulations and experiments in *C. elegans*, we show how phosphorylation of CTD can give rise to distinct compartments which could underpin transcription regulation. *In vitro*, perturbation to transcription condensates have led to a reduction

of transcription[83] and *in vivo* IDRs which can form liquid condensates have been shown to be instrumental for transcription[59]. Experiments have suggested that transcription initiation and elongation may be regulated in time and space by phase-separated condensates[15,75]. How this may be achieved by proteins enriched in seemingly structured and feature-less IDRs[84–86], has remained elusive. We show how the interactions of CTD and pCTD with additional proteins such as HRD of Cyclin T1 which can shield the negative charges introduced by hyper-phosphorylation can give rise to distinct phases, one defined by hydrophobic interactions and the other phase by electrostatically driven cross-interaction driven phase separation. For the formation of distinct phases, pCTD phosphorylation needs to cross a threshold upon which the interactions of CTD and pCTD change drastically. This rationalises how highly phosphorylated pCTD of elongating RNA Polymerase II in zebrafish is excluded from CTD clusters and condensates in zebrafish embryos[19]. CTD is fully and partially engulfed by pCTD-HRD in our simulations. For a given composition of the simulation system, we determine its most favourable morphology by computing interfacial tensions.

Our approach is completely general and can be extended to any kind of mixture of proteins. Thus it becomes possible to predict the morphologies of multi-component mixtures from molecular simulations. The coarse-grained model we employed to simulate large condensates captures the contact statistics of CTD as reported by NMR and atomistic molecular dynamics, is in line with interactions propensities for CTD chains from atomistic molecular dynamics and also reproduces experimentally determined effects of truncations and sequence changes.

In accordance with our simulations, super-resolution microscopy of *C. elegans* embryos reveal full and partial engulfment of CTD by pCTD. RNAi knockdown of Cyclin T1 preserves CTD condensates, which is in line with the stability of CTD condensates in the absence of Ser phosphorylation suggested by our simulations (Fig S9). While the two condensates are distinct and modulate the global conformations of proteins in different ways, the two dense phases are in contact (Fig 7F). The two phases also differentially recruit FUS as expected, with FUS binding preferentially to the CTD rather than pCTD-HRD. Thus our findings provide a molecular basis for the observations by Wei et al., who report that in cells Ser_5 phosphorylated CTD concentrically organises around a TAF-15 promoter condensate[75]. While the study by Wei et al., did not probe the makeup of the pCTD shell, whether pCTD accumulated on the exterior of the TAF-15 condensate on its own or in a dense phase,

this core-shell-like accumulation of pCTD around a promoter condensate is reminiscent of the morphologies we have resolved in simulations and in cells and suggest that a promoter condensate may be engulfed by a gene-body condensate.

Our study highlights an additional and unexpected layer of specificity which CTD condensates may impart[87]. Experiments showed that CTD forms condensate in a temperature dependent manner in *C. elegans* embryos. Many protein condensates dissolve at high-temperature. Yet transcription rates do not decrease with moderate temperature increases. On the contrary, transcription is typically enhanced. This apparent conundrum is resolved by our *in vivo* experiments, simulations, which highlight that the LCST behavior of CTD would be consistent with the involvement of condensates in transcription and an increase of transcription at elevated temperatures. The LCST behavior of CTD, which has previously been characterised in biophysical experiments[56], correlates with moderate shifts in gene transcription programs with temperature as shown by our RNA sequencing data. The LCST behavior is largely uncoupled from a classical heat shock response, which results dramatic changes in the pool of RNA transcripts going from 30 °C to 37 °C. At temperatures below 37 °C changes in the phase separation propensity of CTD in *C. elegans* correlate with a fine-tuning of gene transcription as temperature changes, with most changes seen for transcripts coding for proteins in metabolic pathways, which may be an important facet of homeostasis in non-stress conditions. This LCST behavior of CTD contrasts with the temperature behavior of other proteins in the nucleus, such as the heterochromatin proteins MET-2[54] and HP1α[67], which dissolves at higher temperatures. These differences further highlight that CTD condensates that support transcription are physio-chemically distinct from heterochromatin condensates, which is likely essential for transcription regulation.

ACKNOWLEDGEMENTS

A.C.S is supported by M^{3}ODEL. A.C.S, F.S. and L.S.S. thank M^{3}ODEL for support. L.S.S. acknowledges support by ReALity (Resilience, Adaptation and Longevity) and Forschungsinitiative des Landes Rheinland-Pfalz. This project was funded by SFB 1551 Project No. 464588647 of the DFG (Deutsche Forschungsgemeinschaft). Funding from the German Science Foundation is acknowledged: F.S. and L.S.S. are members of the RTG 2516 (Grant No. 405552959). Further, we gratefully acknowledge the advisory services offered

and the computing time granted on the supercomputers Mogon II at Johannes Gutenberg University Mainz, which is a member of the AHRP (Alliance for High Performance Computing in Rhineland Palatinate) and the Gauss Alliance e.V. We thank Dr. J. Mittal, Dr. P. Virnau, Dr. M. Andrade, Dr. R. V. Pappu, and Dr. M. Zweckstetter for inspiring discussions and Dr. G. Krawezik for help with the Folding@Home simulations. STELLARIS 8 Falcon confocal was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 497669232. All imaging and RNA sequencing was supported by the IMB Microscopy Core Facility and IMB Genomics Core Facility.

REFERENCES

[1]W. M. Jacobs and D. Frenkel, Biophys. J. **112**, 683 (2017).

[2]D. Zwicker and L. Laan, Proc. Natl. Acad. Sci. U. S. A. **119**, e2201250119 (2022).

[3]S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V. Pappu, *et al.*, Proc. Natl. Acad. Sci. U. S. A. **116**, 7889 (2019).

[4]A. P. Latham and B. Zhang, Aggregate **n/a**, e306, https://onlinelibrary.wiley.com/doi/pdf/10.1002/agt2.306.

[5]K. Mazarakos and H.-X. Zhou, J. Chem. Physics **156**, 191104 (2022), https://doi.org/10.1063/5.0088004.

[6]T. Kaur, M. Raju, I. Alshareedah, R. B. Davis, D. A. Potoyan, and P. R. Banerjee, Nat. Commun. **12**, 1 (2021).

[7]L. A. Gruijs da Silva, F. Simonetti, S. Hutten, H. Riemenschneider, E. L. Sternburg, L. M. Pietrek, J. Gebel, V. Dötsch, D. Edbauer, G. Hummer, L. S. Stelzl, and D. Dormann, EMBO J. **41**, e108443.

[8]T. M. Perdikari, N. Jovic, G. L. Dignon, Y. C. Kim, N. L. Fawzi, and J. Mittal, Biophys. J. **120**, 1187 (2021).

[9]H. P. Phatnani and A. L. Greenleaf, Genes Dev. **20**, 2922 (2006), http://genesdev.cshlp.org/content/20/21/2922.full.pdf+html.

[10]P. Cramer, Nature **573**, 45 (2019).

[11]K. M. Harlen and L. S. Churchman, Nat. Rev. Mol. Cell Biol. **18**, 263 (2017).

[12]I. I. Cisse, I. Izeddin, S. Z. Causse, L. Boudarene, A. Senecal, L. Muresan, C. Dugast-Darzacq, B. Hajj, M. Dahan, and X. Darzacq, Science **341**, 664 (2013), https://www.science.org/doi/pdf/10.1126/science.1239053.

[13]M. Boehning, C. Dugast-Darzacq, M. Rankovic, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, and M. Zweckstetter, Nat. Struct. Mol. Biol. **25**, 833 (2018).

[14]H. Lu, D. Yu, A. S. Hansen, S. Ganguly, R. Liu, A. Heckert, X. Darzacq, and Q. Zhou, Nature **558**, 318 (2018).

[15]Y. E. Guo, J. C. Manteiga, J. E. Henninger, B. R. Sabari, A. Dall'Agnese, N. M. Hannett, J.-H. Spille, L. K. Afeyan, A. V. Zamudio, K. Shrinivas, B. J. Abraham, A. Boija, T.-M. Decker, J. K. Rimel, C. B. Fant, T. I. Lee, I. I. Cisse, P. A. Sharp, D. J. Taatjes, and R. A. Young, Nature **572**, 543 (2019).

[16]J. A. Morin, S. Wittmann, S. Choubey, A. Klosin, S. Golfier, A. A. Hyman, F. Jülicher, and S. W. Grill, Nat. Phys. **18**, 271 (2022).

[17]B. Portz and J. Shorter, Trends Biochem. Sci. **45**, 1 (2020).

[18]P. Quintero-Cadena, T. L. Lenstra, and P. W. Sternberg, Molecular Cell **79**, 207 (2020).

[19]A. Pancholi, T. Klingberg, W. Zhang, R. Prizak, I. Mamontova, A. Noa, M. Sobucki, A. Y. Kobitski, G. U. Nienhaus, V. Zaburdaev, and L. Hilbert, Mol. Syst. Biol. **17**, e10272 (2021), https://www.embopress.org/doi/pdf/10.15252/msb.202110272.

[20]R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw, Annu. Rev. Biophys. **41**, 429 (2012), pMID: 22577825, https://doi.org/10.1146/annurev-biophys-042910-155245.

[21]G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLOS Comput. Biol. **14**, 1 (2018).

[22]Z. Benayad, S. von Bülow, L. S. Stelzl, and G. Hummer, J. Chem. Theory Comput. **17**, 525 (2021).

[23]G. Tesei, T. K. Schulze, R. Crehuet, and K. Lindorff-Larsen, Proc. Natl. Acad. Sci. U. S. A. **118** (2021), 10.1073/pnas.2111696118, https://www.pnas.org/content/118/44/e2111696118.full.pdf.

[24]R. M. Regy, G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, Nucleic Acids Research **48**, 12593 (2020), https://academic.oup.com/nar/article-pdf/48/22/12593/34906162/gkaa1099.pdf.

[25]M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne, Cell **165**, 1686 (2016).

[26]J. Fei, M. Jadaliha, T. S. Harmon, I. T. Li, B. Hua, Q. Hao, A. S. Holehouse, M. Reyer, Q. Sun, S. M. Freier, *et al.*, J. Cell Sci. **130**, 4180 (2017).

[27]I. Alshareedah, M. M. Moosa, M. Raju, D. A. Potoyan, and P. R. Banerjee, Proc. Natl. Acad. Sci. U. S. A. **117**, 15650 (2020).

[28]T. S. Harmon, A. S. Holehouse, and R. V. Pappu, New J. Phys. **20**, 045002 (2018).

[29]A. Garaizar, J. R. Espinosa, J. A. Joseph, G. Krainer, Y. Shen, T. P. Knowles, and R. Collepardo-Guevara, Proc. Natl. Acad. Sci. U. S. A. **119**, e2119800119 (2022).

[30]S. Mukherjee and L. V. Schäfer, Nat. Commun. **14**, 5892 (2023).

[31]E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, Science **367**, 694 (2020).

[32]S. Rekhi, C. G. Garcia, M. Barai, A. Rizuan, B. S. Schuster, K. L. Kiick, and J. Mittal, bioRxiv , 2023 (2023).

[33]G. Tesei, A. I. Trolle, N. Jonsson, J. Betz, F. E. Knudsen, F. Pesce, K. E. Johansson, and K. Lindorff-Larsen, Nature **626**, 897 (2024).

[34]J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, Nat Comput Sci. **1**, 732 (2021).

[35]G. Tesei and K. Lindorff-Larsen, Open Research Europe **2** (2023), 10.12688/openreseurope.14967.2.

[36]G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, ACS Central Science **5**, 821 (2019).

[37]J. A. Anderson, J. Glaser, and S. C. Glotzer, Comput. Mater. Sci. **173**, 109363 (2020).

[38]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).

[39]R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009).

[40]K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, Proteins: Struct., Funct., Bioinf. **78**, 1950 (2010).

[41]R. B. Best, D. De Sancho, and J. Mittal, Biophys. J. **102**, 1462 (2012).

[42]S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, J. Phys. Chem. B **119**, 5113 (2015), NIHMS150003.

[43]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007), https://doi.org/10.1063/1.2408420.

[44]M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).

[45]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1-2**, 19 (2015).

[46]M. Shirts and V. S. Pande, Science **290**, 1903 (2000), https://www.science.org/doi/pdf/10.1126/science.290.5498.1903.

[47]P. J. Flory, *Principles of polymer chemistry* (Cornell university press, 1953).

[48]R. Horst, Macromol. Theory Sim. **4**, 449 (1995), https://onlinelibrary.wiley.com/doi/pdf/10.1002/mats.1995.040040304.

[49]A. Werner, F. Schmid, M. Müller, and K. Binder, Phys. Rev. E **59**, 728 (1999).

[50]D. J. Bauer, L. S. Stelzl, and A. Nikoubashman, J. Chem. Phys. **157**, 154903 (2022), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0105540/16554276/154903_1_online.pdf.

[51]J. Henderson and J. Lekner, Mol. Phys. **36**, 781 (1978).

[52]C. Dietz and M. R. Berthold, "Knime for open-source bioimage analysis: A tutorial," in *Focus on Bio-Image Informatics*, edited by W. H. De Vos, S. Munck, and J.-P. Timmermans (Springer International Publishing, Cham, 2016) pp. 179–197.

[53]J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri, Methods **115**, 80 (2017).

[54]C. E. Delaney, S. P. Methot, V. Kalck, J. Seebacher, D. Hess, S. M. Gasser, and J. Padeken, Nat. Stru. Mol. Biol. **29**, 85 (2022).

[55]L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, and H. Peterson, Nucleic Acids Research **51**, W207 (2023), https://academic.oup.com/nar/article-pdf/51/W1/W207/50736882/gkad347.pdf.

[56]D. Flores-Solis, I. P. Lushpinskaia, A. A. Polyansky, A. Changiarath, M. Boehning, M. Mirkovic, J. Walshe, L. M. Pietrek, P. Cramer, L. S. Stelzl, B. Zagrovic, and M. Zweckstetter, Nat. Commun. **14**, 5979 (2023).

[57]L. S. Stelzl and G. Hummer, J. Chem. Theory Comput. **13**, 3927 (2017).

[58]L. S. Stelzl, A. Kells, E. Rosta, and G. Hummer, J. Chem.Theory Comput. **13**, 6328 (2017), pMID: 29059525, https://doi.org/10.1021/acs.jctc.7b00373.

[59]F. Lu, B. Portz, and D. S. Gilmour, Mol. Cell **73**, 1232 (2019), doi: 10.1016/j.molcel.2019.01.008.

[60]A. W. Fritsch, A. F. Diaz-Delgadillo, O. Adame-Arana, C. Hoege, M. Mittasch, M. Kreysing, M. Leaver, A. A. Hyman, F. Jülicher, and C. A. Weber, Proc. Natl. Acad. Sci. U. S. A. **118**, e2102772118 (2021), https://www.pnas.org/doi/pdf/10.1073/pnas.2102772118.

[61]X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, and R. V. Pappu, Biophys. J. **119**, 402 (2020).

[62]K. M. Ruff, S. Roberts, A. Chilkoti, and R. V. Pappu, J. Mol. Biol. **430**, 4619 (2018).

[63]S. Roberts, T. S. Harmon, J. L. Schaal, V. Miao, K. Li, A. Hunt, Y. Wen, T. G. Oas, J. H. Collier, R. V. Pappu, *et al.*, Nature Mat. **17**, 1154 (2018).

[64]M. J. Fossat, X. Zeng, and R. V. Pappu, J. Phys. Chem. B **125**, 4148 (2021).

[65]X. Zeng, C. Liu, M. J. Fossat, P. Ren, A. Chilkoti, and R. V. Pappu, APL Mater. **9** (2021).

[66]K. M. Ruff, S. Roberts, A. Chilkoti, and R. V. Pappu, J. Mol. Biol. **430**, 4619 (2018).

[67]A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, Nature **547**, 236 (2017).

[68]A. Bremer, M. Farag, W. M. Borcherds, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, Nat. Chem. **14**, 196 (2022).

[69]G. L. Dignon, R. B. Best, and J. Mittal, Ann. Rev. Phys. Chem. **71**, 53 (2020), pMID: 32312191, https://doi.org/10.1146/annurev-physchem-071819-113553.

[70]J. J. Michels, M. Brzezinski, T. Scheidt, E. A. Lemke, and S. H. Parekh, Biomacromolecules **23**, 349 (2022).

[71]B. Portz, F. Lu, E. B. Gibbs, J. E. Mayfield, M. Rachel Mehaffey, Y. J. Zhang, J. S. Brodbelt, S. A. Showalter, and D. S. Gilmour, Nat. Commun. **8**, 15231 (2017).

[72]A. Werner, F. Schmid, M. Müller, and K. Binder, Phys. Rev. E **59**, 728 (1999).

[73]S. Torza and S. Mason, J. Colloid Interface Sci. **33**, 67 (1970).

[74]D. Liu, Y. Lin, K. Gong, H. Bo, D. Li, Z. Zhang, and W. Chen, RSC. Adv. **11**, 38316 (2021).

[75]M.-T. Wei, Y.-C. Chang, S. F. Shimobayashi, Y. Shin, A. R. Strom, and C. P. Brangwynne, Nat. Cell Biol. **22**, 1187 (2020).

[76]E. Y. Shim, A. K. Walker, Y. Shi, and T. K. Blackwell, Genes Dev. **16**, 2135 (2002), http://genesdev.cshlp.org/content/16/16/2135.full.pdf+html.

[77]M. Farag, S. R. Cohen, W. M. Borcherds, A. Bremer, T. Mittag, and R. V. Pappu, Nature Commun. **13**, 7722 (2022).

[78]L. Emmanouilidis, L. Esteban-Hofer, F. F. Damberger, T. de Vries, C. K. X. Nguyen, L. F. Ibáñez, S. Mergenthal, E. Klotzsch, M. Yulikov, G. Jeschke, and F. H. T. Allain, Nat. Chem. Biol. **17**, 608 (2021).

[79]G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proc. Natl. Acad. Sci. U. S. A. **115**, 9929 (2018).

[80]J. Wang, D. S. Devarajan, A. Nikoubashman, and J. Mittal, ACS Macro Letters **12**, 1472 (2023), pMID: 37856873, https://doi.org/10.1021/acsmacrolett.3c00456.

[81]I. Kwon, M. Kato, S. Xiang, L. Wu, P. Theodoropoulos, H. Mirzaei, T. Han, S. Xie, J. L. Corden, and S. L. McKnight, Cell **155**, 1049 (2013).

[82]A. C. Murthy, W. S. Tang, N. Jovic, A. M. Janke, D. H. Seo, T. M. Perdikari, J. Mittal, and N. L. Fawzi, Nat. Struct. Mol. Biol. **28**, 923 (2021).

[83]B. A. Lewis, S. K. Das, R. K. Jha, and D. Levens, Sci. Adv. **9**, eadi4565 (2023), https://www.science.org/doi/pdf/10.1126/sciadv.adi4565.

[84]G. Forte, A. Buckle, S. Boyle, D. Marenduzzo, N. Gilbert, and C. A. Brackley, Nat. Struc. Mole. Biol. **30**, 1275 (2023).

[85]S. Zhang, N. Übelmesser, N. Josipovic, G. Forte, J. A. Slotman, M. Chiang, H. J. Gothe, E. G. Gusmao, C. Becker, J. Altmüller, A. B. Houtsmuller, V. Roukos, K. S. Wendt, D. Marenduzzo, and A. Papantonis, Sci. Adv. **7**, eabg8205 (2021), https://www.science.org/doi/pdf/10.1126/sciadv.abg8205.

[86]D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, Cell **169**, 13 (2017).

[87]Y. H. Ling, J. Ye, C. Yu, C. Liang, G. Park, J. Corden, and C. Wu, bioRxiv (2023), 10.1101/2023.07.31.551302, https://www.biorxiv.org/content/early/2023/08/02/2023.07.31.551302.1.full.pdf.