UC Merced
UC Merced Previously Published Works

Title

Clustering of RNA Polymerase II C-Terminal Domain Models upon Phosphorylation.

Permalink

https://escholarship.org/uc/item/6tw1f3gq

Journal

Authors

Amith, Weththasinghage
Chen, Vincent
Dutagaci, Bercem

Publication Date

2024-10-24

DOI

10.1021/acs.jpcb.4c04457

Peer reviewed

Article

Clustering of RNA Polymerase II C-Terminal Domain Models upon Phosphorylation

Weththasinghage D. Amith, Vincent T. Chen, and Bercem Dutagaci*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: RNA polymerase II (Pol II) C-terminal domain (CTD) is known to have crucial roles in regulating transcription. CTD has also been highly recognized for undergoing phase separation, which is further associated with its regulatory functions. However, the molecular interactions that the CTD forms to induce clustering to drive phase separations and how the phosphorylation of the CTD affects clustering are not entirely known. In this work, we studied the concentrated solutions of two heptapeptide repeat (2CTD) models at different phosphorylation patterns and protein and ion concentrations using all-atom molecular dynamics simulations to investigate clustering behavior and molecular interactions driving the cluster formation. Our results show that salt concentration and phosphorylation patterns play an important role in determining the clustering pattern, specifically at low protein concentrations. The balance between inter- and intrapeptide interactions and counterion coordination together impact the clustering behavior upon phosphorylation.
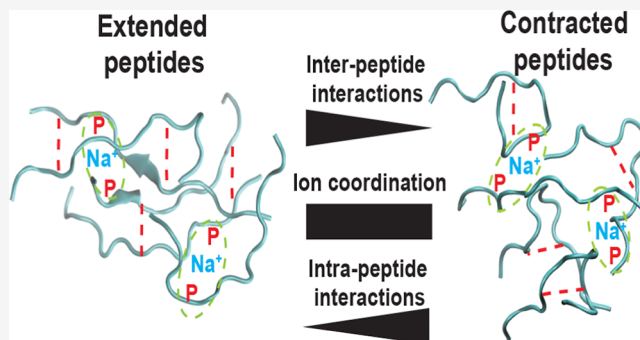
INTRODUCTION

The transcription process of genetic information on DNA into RNA is mainly regulated by the RNA polymerase II (Pol II) complex which consists of several subunits.[1,2] The C-terminal domain (CTD) of the largest subunit (RPB1) of Pol II is a disordered low-complexity region that falls into the category of intrinsically disordered proteins (IDPs).[3,4] The CTD consists of heptapeptide (YSPTSPS) repeating units and the number of repeating units differs according to the organism.[3] The CTD has critical roles in transcription at several steps including initiation,[5,6] pausing,[7,8] elongation,[9,10] splicing[11,12] and termination,[13,14] and it undergoes post-translational modifications, including phosphorylation, a key regulatory modification in these transcription steps.[3,15]

CTD has been also recognized for its fundamental role in liquid−liquid phase separation (LLPS).[4,16−20] Phosphorylation of CTD impacts the LLPS formation[4,16,21−23] and has been suggested as a regulator for CTD to switch between different condensates.[4,16,21,23] However, there is limited knowledge of the structural properties of CTD and the effects of phosphorylation on the structure and clustering behavior.[17,24−26] It has been intuitively assumed that introducing phosphorylation might extend the CTD structure due to the increased repulsion between negatively charged phosphate groups[3,27] and a recent study supported this prediction by identifying a more extended structure upon phosphorylation using the small angle X-ray scattering (SAXS) technique.[26] In addition, in our earlier work,[28] we showed that phosphorylation has complex effects on the structural properties of CTD

models, which were either compacted or extended depending on the CTD length and phosphorylation pattern.

The knowledge of the clustering behavior of the CTD is also limited. Previous studies on crowded systems of proteins using both experimental[20,29−32] and computational[32−36] methods show that protein−protein interactions play an important role in forming clusters, while the nature of such interactions for CTD clustering has not been entirely known. The clustering of CTD and formation of liquid droplets were observed by in vitro studies,[4,16] and a recent study suggested a critical role for tyrosine interactions in cluster formation by molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) spectroscopy.[17] How the phosphorylation affects the clustering, on the other hand, is still an open question. To obtain insights into this open question, in this work, we investigated the clustering behavior of CTD models in different phosphorylation states by using all-atom MD simulations. Following our previous work,[28] we studied the 2-heptapeptide repeating units (2CTD) model in its non-phosphorylated form and at three different phosphorylation states. We performed simulations in three different protein

Table 1. 2CTD Models with Their Sequences, Abbreviations, Protein Concentrations, Number of Peptides, and Net Charge[a]

2CTD sequence	abbreviation	protein concentration (mM)	number of peptides	net charge
YSPTSPSYSPTSPS	2CTD-non-phos	1	5	0
		2.5	5	0
		8.3	5	0
		17	10	0
		25	15	0
ASPTSPSASPTSPS	2CTD-non-phos-Y_A	8.3	5	0
YSPASPSYSPASPS	2CTD-non-phos-T_A	8.3	5	0
YSATSASYSATSAS	2CTD-non-phos-P_A	8.3	5	0
YSPTSPSYSPTSPS	2CTD-2P-5P	8.3	5	−4
		17	10	−4
		25	15	−4
YSPTSPSYSPTSPS	2CTD-2P-12P	8.3	5	−4
		17	10	−4
		25	15	−4
YSPTSPSYSPTSPS	2CTD-2P-5P-9P-12P	8.3	5	−8
		17	10	−8
		25	15	−8

[a]Sequences were given with different phosphorylation states represented by the underlined serine residues on each sequence. Each protein concentration was simulated separately at three different NaCl concentrations (0, 150, and 300 mM), except for the simulations of 2CTD-non-phos at low protein concentrations (1 and 2.5 mM) and mutant 2CTD-non-phos systems, which were performed only for 150 mM NaCl concentration. The net charge is given per peptide.

concentrations to mimic a high-concentration crowded environment and three different salt concentrations to monitor the effects of ionic strength on the clustering of CTD. The initial systems were prepared using the low energy conformations of each 2CTD model that were extracted from the replica-exchange MD (REMD)[37] simulations performed in our recent study.[28] After performing μs time scale MD simulations, we analyzed the clustering behavior and the molecular interactions that could drive the cluster formation. We observed that interpeptide interactions along with ion coordination with negatively charged phosphate groups of different peptides dictate the clustering patterns of these crowded systems.

■ COMPUTATIONAL METHODS

System Preparation. For this study, we prepared multiple systems at different concentrations of 2CTD, patterns of phosphorylated serine residues, and salt concentrations. The exact models used for this study, with their respective sequences, abbreviations, protein concentrations, number of peptides, and net charge per peptide, are shown in Table 1. To model the concentrated systems at each protein concentration, we used initial structures from the low-energy conformations of the four 2CTD sequences that were obtained from our previous study.[28] All of the crowded systems were generated using the following protocol: for all 2CTD peptides, acetyl (ACE) and -NHCH3 (CT3) groups were utilized to cap the N-terminus and C-terminus, respectively. Each 2CTD crowded system was solvated in a cubic box with each side length of the simulation box set to 100 Å except for the systems with 1 and 2.5 mM protein concentrations, for which we used 200 and 150 Å cubic box lengths, respectively. The systems were neutralized by including sodium (Na+) ions without additional salt (0 mM NaCl). In addition, each protein concentration was modeled at 150 and 300 mM NaCl concentrations to investigate the salt effect on clustering. Overall, the CHARMM-GUI server[38,39] was utilized to prepare and model the 38 systems with different salt concentrations,

protein concentrations, and phosphorylated states as specified in Table 1. In addition, we performed three mutation simulations for 2CTD-non-phos. In each simulation, we mutated either all Y, all T, or all P into A (see Table 1).

MD Simulations. The CHARMM-modified TIP3P force field[40] was utilized for the explicit water. For protein, we used CHARMM 36m[41] force field with modifications in the nonbonded parameters between protein and water that the details were described in Huang et al.[41] The energy minimization for each crowded solution was performed for 5000 steps with a tolerance of 100 kJ/mol. Then, systems were equilibrated for 1 ns while increasing the temperature from 100 to 300 K. Restraints for each 2CTD system were turned on for the initial 625 ps of equilibration and turned off for the last 375 ps of the equilibration. During the initial 625 ps of the equilibration, both the backbone and side chains of the 2CTDs were constrained using the force constants 400 and 40 kJ/mol/nm^2 respectively. The Lennard-Jones interactions were switched between 1.0 and 1.2 nm where the time step was set to 1 fs for the equilibration steps.

The output configurations were extracted from the last equilibration step and utilized as the input configurations for the production runs performed for 1 μs for each system except for the 2CTD-non-phos systems at low protein concentrations. Simulations were performed for 200 and 500 ns for protein concentrations of 1 and 2.5 mM, respectively. Long-range electrostatic interactions were calculated using the particle mesh Ewald (PME)[42,43] algorithm with periodic boundary conditions. The Lennard-Jones interactions were switched between 1.0 and 1.2 nm. The temperature was maintained at 300 K for the production runs. The friction coefficient was set to 1 ps^{-1} for the Langevin thermostat. The time step was set to 2 fs, and frames were saved in the MD simulation trajectories at every 10 ps for the production runs. All the MD simulations were performed using OpenMM package[44] on GPU-enhanced environment. Overall, a total of 39.7 μs were achieved after all the production runs.
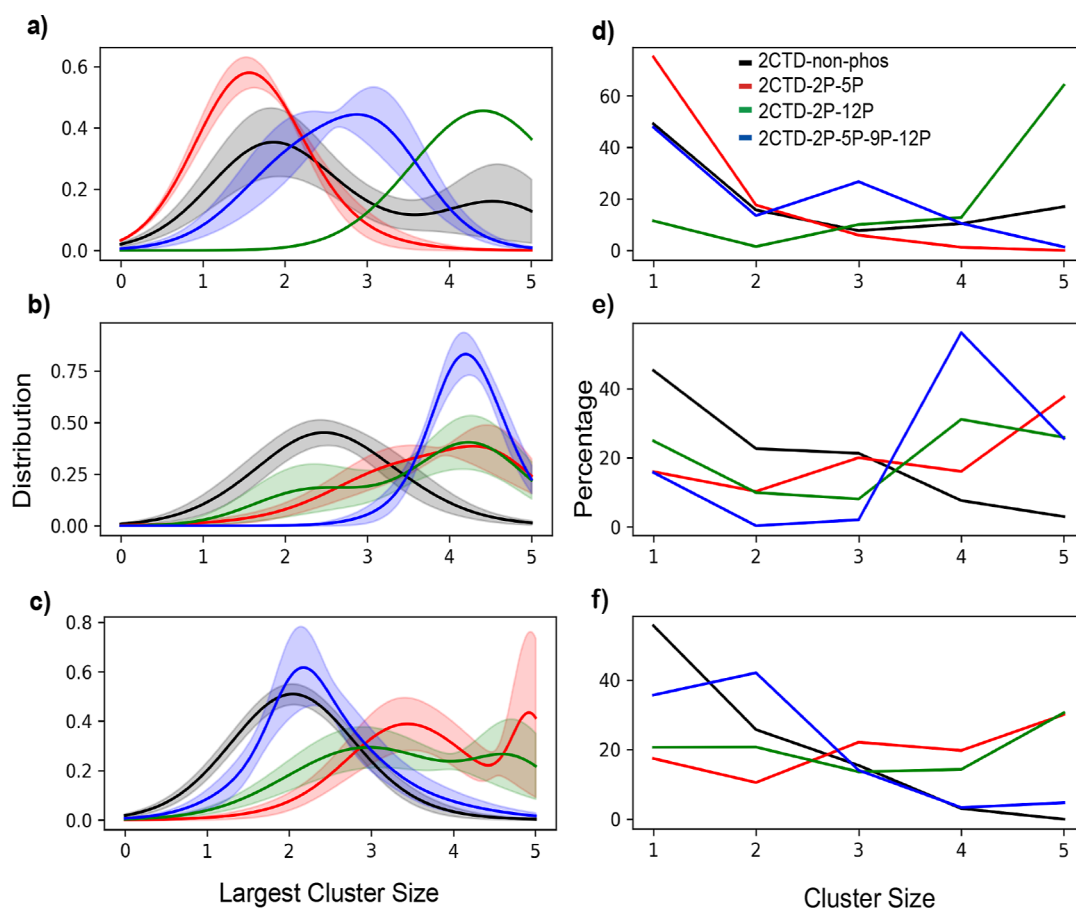
Figure 1. Cluster distribution and cluster size percentages in 2CTD crowded systems at an 8.3 mM protein concentration. The largest cluster size distributions are shown at (a) 0, (b) 150, and (c) 300 mM NaCl. Percentages of each cluster size are shown at (d) 0, (e) 150, and (f) 300 mM NaCl. Standard errors [for (a–c)] were calculated by splitting the last 800 ns of the 1 μs trajectory into 160 ns small trajectories for each system. For 2CTD-2P-12P in panel (a) the standard error is not shown because approximately for the last 320 ns the largest cluster size remained constant at 5. The colors of the curves are the same as in panel (d) for all the other panels.

Analysis of MD Simulations. For the data analysis, we used the last 800 ns of the 1 μs MD trajectories from the production runs for each system except for the systems with protein concentrations of 1 and 2.5 mM, for which the first 100 ns of the simulations were discarded during the analysis. All the analyses were performed using MDAnalysis package[45] and custom Python and C++ scripts.

We calculated the cluster size distributions using a distance cutoff value following the previous works.[33,46] A cluster between two peptides is counted when the distance between two alpha-carbon (CA) atoms is below the cutoff of 7 Å. For example, if all 5 peptides are closer than the cutoff at an 8.3 mM protein concentration at a certain frame, the cluster size outputs as 5 for that frame. The output data were used to generate the distributions of the largest cluster sizes along the simulations and percentages of each cluster size as shown in the Results section.

The self-diffusion coefficients (D) were calculated from the mean squared displacements (MSD) using the Einstein formula in MDAnalysis[45] by considering all the peptides for each 2CTD crowded system with unwrapped MD trajectories. Figure S1 in the Supporting Information shows how we calculated D for each crowded system by plotting MSD values against the lag times. To determine the slope of the plot, we performed a linear regression between 20 and 100 ps lag times. MSD values were calculated using the last 800 ns of the

simulations and with a time window of 10 ps between steps, which was the frequency of saving the frames during the production run and the fast-Fourier transform (FFT) approach[47] implemented in MDAnalysis with the tidynamics package[48] to improve the speed of the calculation. After the linear regression, D was calculated by dividing the slope of the fitted linear line by a 2 x dimension factor. Here, the dimension factor is equal to 3 because we are considering all three coordinates of the peptides during the MSD calculation.

The radius of gyration (R_g) for each peptide was calculated in a frame and then averaged at each frame over the number of peptides for each protein concentration. To obtain the conformational landscapes, we performed a principal component analysis (PCA) on aligned Cartesian coordinates. We first aligned all of the peptides in each frame along the trajectory against the initial conformation. Then, we calculated the average structure and performed a second alignment against the average structure. The Cartesian coordinates of the backbone atoms of the aligned peptides were considered to perform the PCA. Only the first and second principal components (PC1 and PC2) were considered. The free energy landscapes of the PCA were obtained by applying the weighted histogram analysis method (WHAM) using the WHAM package developed by Grossfield lab.[49] For the PCA, MATLAB[50] (along with PC1 and PC2, which served as the reaction coordinates of the 2CTD systems) was utilized to plot

Table 2. Self-Diffusion Coefficients of 8.3 mM 2CTDs at Different Salt Concentrations

NaCl concentration (mM)	diffusion coefficients (10^{-2} Å^2 ps^{-1})			
	2CTD-non-phos	2CTD-2P-5P	2CTD-2P-12P	2CTD-2P-5P-9P-12P
0	3.81	3.39	2.01	2.48
150	3.83	2.34	2.30	1.67
300	3.92	2.23	2.17	2.16

the free energy landscapes. To extract the frames with the minimum energy conformations, first, for each peptide within each system, we identified the lowest energy conformation located at the energy minimum of the corresponding PCA plot. Then, we selected the frame corresponding to the minimum energy conformations of each peptide. Therefore, we selected five frames, each with the minimum energy conformation of one of the five peptides in the 8.3 mM concentrated CTD systems. Then, the frame with the highest number of low-energy conformations among the five selected frames was extracted for each system. We used the last 800 ns of simulations for PCA and the last 100 ns for the extraction of the low-energy conformations.

Contact maps were generated by considering the interpeptide contacts below or equal to 5 Å for 2CTD systems at 8.3 mM protein and 150 mM NaCl concentrations. Contacts were averaged for each residue pair over the number of peptide pairs and number of frames of the trajectories.

RESULTS

In this study, we investigated the clustering patterns of the nonphosphorylated 2CTD model and 2CTD models at different phosphorylation states. We obtained insights into the intermolecular interactions and effects of phosphorylation within heptapeptide repeats toward developing an understanding of the LLPS occurring in full human CTD. In the main text, we will only show the results related to the lowest protein concentration of 2CTD crowded systems, which is 8.3 mM. All the results related to the other two high protein concentrations (17 and 25 mM) are shown in the Supporting Information.

2CTD Systems Showed Varied Degrees of Clustering Depending on the Phosphorylation Pattern. Figure 1 shows the largest cluster size distributions (1a−c) and the cluster size percentages (1d−f) at an 8.3 mM protein concentration and three different NaCl concentrations. For 2CTD-non-phos, the largest cluster size distributions did not show any significant changes (average stays around 2) across different ion concentrations (Figure 1a−c), and higher percentages were observed for monomers and dimers (Figure 1d−f).

Phosphorylated 2CTD systems show more diverse trends in the largest cluster size distributions at each ion concentration. Generally, phosphorylated systems display higher cluster sizes compared to the nonphosphorylated state except for the case of 2CTD-2P−5P at 0 mM NaCl concentration in Figure 1a and d. Both 2CTD-2P-5P-9P-12P and 2CTD-2P-5P have cluster sizes increased by increased salt concentration from 0 to 150 mM, then reduced by increasing the salt concentration even more to 300 mM for 2CTD-2P−5P−9P−12P and stayed around the same for 2CTD-2P−5P. This suggested that Na$^+$ ions may form bridging electrostatic interactions with phosphate groups, which increase the clustering at a 150 mM concentration, while in a higher concentration of NaCl, electrostatic interactions are reduced due to the charge

screening. For 2CTD-2P-12P, on the other hand, we observed the largest clusters at the lowest ion concentration, suggesting that electrostatic interactions at higher ion concentrations are hindered potentially because of a conformational change. In addition, diffusion coefficients of peptides correlate with the observed clustering showing that increased clustering reduced translational diffusion of peptides (Table 2). The clustering analysis supported by diffusion coefficient calculations overall suggests that different phosphorylated systems have complex patterns of clustering at different salt concentrations.

At higher protein concentrations (17 and 25 mM), we observed a larger degree of clustering, especially at higher NaCl concentrations (at 150 and 300 mM). Figures S2 and S3 show the largest cluster size distributions and cluster size percentages for the 2CTD systems at 17 and 25 mM protein concentrations, respectively. For both concentrations, we observed higher values for the largest cluster sizes for both nonphosphorylated and phosphorylated crowded systems compared to an 8.3 mM protein concentration. We also observed an around 2−3 times decrease in diffusion coefficients especially for a 25 mM concentration (Tables S1 and S2). The diffusion coefficients at high CTD concentrations suggest that the high number of clusters reduced the dynamics of the CTDs. On the other hand, the order of reduction in diffusion coefficient suggests that the systems are in the liquid character since studies suggest more than a 10 times reduction in diffusion coefficients for transitioning from liquid to more ordered structures.[51−53]

In addition, we performed simulations of 2CTD-non-phos at lower protein concentrations (at 1 and 2.5 mM). At 1 mM, CTD tended to stay as a monomer, and no clustering was observed, while at 2.5 mM, we started to observe dimer and trimer formations (Figure S4). Diffusion coefficients for the low protein concentration systems are higher than the values obtained for the higher concentration simulations, as expected due to the decreased number of cluster formations (Table S3). Figure S5 shows the cluster size distributions of 2CTD-non-phos at four concentrations (2.5, 8.3, 17, and 25 mM) and suggests that only at the 25 mM concentration, phase separation may take place in which the dilute phase has small oligomers while the condensed phase is formed by larger clusters, while at 1 mM, no clusters formed. Overall, our results suggest that 2CTD-non-phos may have distinct critical concentrations to start forming clusters and to form phase separation.

Crowding and Ion Concentration Alter the Conformations and Impact Clustering. We observed that phosphorylation and ion concentration have complex effects in clustering as one phosphorylated system has the largest cluster sizes at 0 mM salt concentration, while the other two show an opposite trend from 0 to 150 mM salt concentrations. To understand these clustering patterns, we needed to investigate the structural properties of these 2CTD crowded systems. For that purpose, we analyzed the conformational changes for the peptides in each 2CTD system.
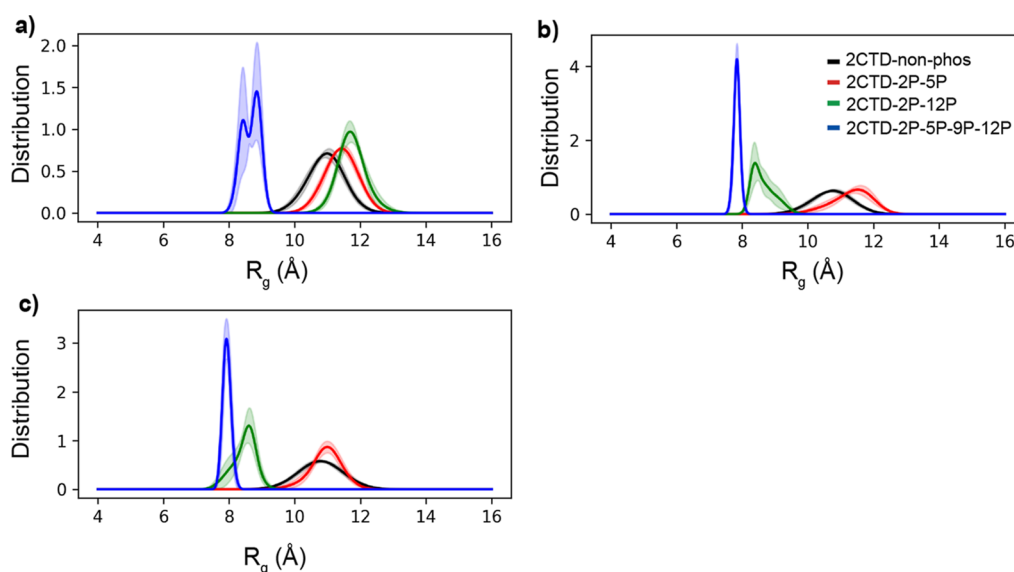
Figure 2. Average radius of gyration (R_g) density distributions for 2CTD crowded systems at 8.3 mM protein and (a) 0, (b) 150, and (c) 300 mM NaCl concentrations. The colors of the curves are the same as in panel (b) for all of the other panels. Standard errors were calculated by splitting the last 800 ns of the 1 μs trajectory into 160 ns of small trajectories for each system.

In order to investigate the conformational space of the 2CTD systems, we first generated the average R_g distributions over the number of peptides for each 2CTD system at three different NaCl concentrations shown in Figure 2. At a 0 mM salt concentration (only in the presence of neutralizing counterions), we observed that 2CTD-2P-5P and 2CTD-2P-12P extended their structures while 2CTD-2P-5P-9P-12P contracted compared to 2CTD-non-phos in Figure 2a. This observation deviates for the 2CTD-2P-12P system from the results in our earlier work,[28] in which we observed contraction for 2CTD-2P-12P compared to 2CTD-non-phos using REMD simulations of isolated single peptides. Expansion of 2CTD-2P-12P in the concentrated system compared to the isolated system suggests that crowding affects the conformation in a way that interpeptide interactions dominate over the interactions with nearby residues (intrapeptide). This conformational change was accompanied by the increased clustering observed in Figure 1a,d, which suggests that clustering and conformational changes interactively influence each other. At higher salt concentrations, 2CTD-2P-12P showed more contracted conformations (Figure 2b,c), suggesting that not only crowding but also salt concentration altered the conformation of this peptide. Average R_g distributions for 17 and 25 mM protein concentrations are shown in Figures S6 and S7, respectively, at three different salt concentrations. The average R_g distribution patterns were similar to the patterns of the 8.3 mM protein concentration (except at a 0 mM NaCl concentration) at all three salt concentrations for all of the crowded systems.

To obtain a deeper understanding of the conformational space of 2CTD-2P-12P compared to other three 2CTD crowded systems, we generated free energy landscapes of PCA using the Cartesian coordinates of backbone atoms of peptides as described in the Methods section and extracted the frames with the low energy conformations shown in Figure 3 for 2CTD-2P-12P, Figures S8, S9, and S10 in the Supporting Information for 2CTD-non-phos, 2CTD-2P-5P-9P-12P and 2CTD-2P-5P, respectively, at three NaCl concentrations. Figure 3a shows the extended peptides for the 2CTD-2P-

12P crowded system at a 0 mM NaCl concentration, which confirmed the observed R_g behavior in Figure 2a. Additionally, we saw that 2CTD-2P-12P peptides are contracted at 150 and 300 mM NaCl concentrations in Figure 3b,c, respectively, compared to at 0 mM NaCl (Figure 3a), which also aligned with the R_g behavior of 2CTD-2P-12P in Figure 2b,c. For the other three 2CTD systems, we confirmed the extended (for 2CTD-non-phos and 2CTD-2P-5P) and contracted (for 2CTD-2P-5P-9P-12P) peptide conformations at three NaCl concentrations in Figures S8–S10.

Hydrogen Bonding and Counterion Interactions Drive the Clustering of 2CTD Systems. In order to determine which interactions drive cluster formation in the 2CTD crowded systems, we first calculated the total number of H-bonds associated with peptides including both intrapeptide and interpeptide H-bonds. Figure 4 represents the distributions of the total number of H-bonds associated with peptides at 0, 150, and 300 mM NaCl concentrations.

When we introduced the phosphorylation, the total number of H-bonds associated with peptides increased at all NaCl concentrations compared to 2CTD-non-phos, except for 2CTD-2P-5P at a 0 mM NaCl concentration (Figure 4a). This might be the reason why 2CTD-2P-5P showed lower cluster sizes compared to 2CTD-nonphos at 0 mM NaCl concentration in Figure 1a. In addition, we observed a higher total number of H-bonds associated with peptides for 2CTD-2P-12P compared to the other three 2CTD crowded systems at 0 and 300 mM NaCl concentrations in Figure 4a,c, respectively. Furthermore, we observed a higher number of H-bonds at 300 mM NaCl compared to 0 mM, which is not consistent with our earlier observation of a higher largest cluster size for 2CTD-2P-12P at 0 mM NaCl compared to 300 mM in Figure 1a. In order to understand the contribution of H-bonds to the clustering, we need to separate the total number of H-bonds associated with peptides into intrapeptide and interpeptide H-bonds as shown in Figure 5 for 2CTD-2P-12P. We observed a higher number of interpeptide H bonds (Figure 5c) compared to intrapeptide H-bonds (Figure 5b) for 2CTD-2P-12P at a 0 mM NaCl concentration compared to
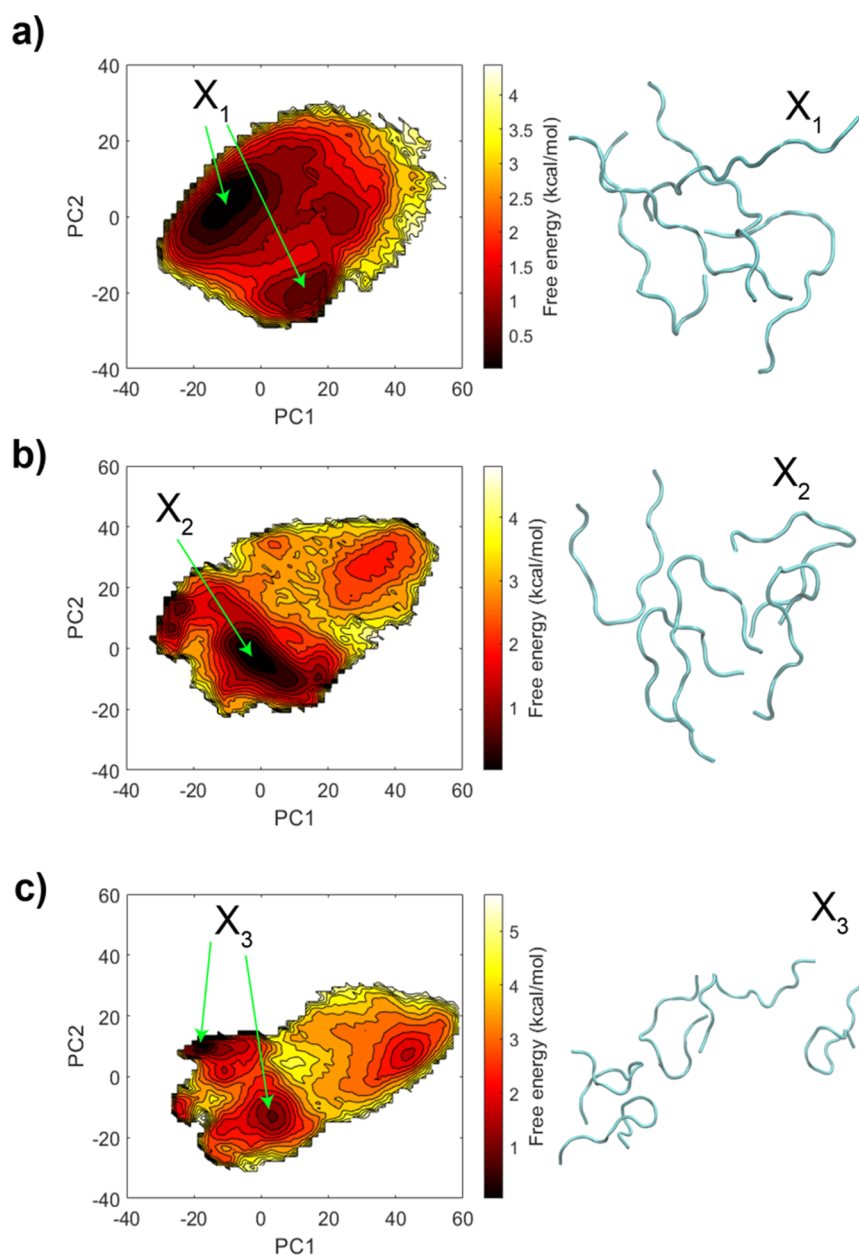
Figure 3. Free energy landscapes using PC1 and PC2 as reaction coordinates from the PCA using Cartesian coordinates for 2CTD-2P-12P crowded models at 8.3 mM protein and (a) 0, (b) 150, and (c) 300 mM NaCl concentrations. In addition, $X_1 - X_3$ represents the frames that have low energy conformations at each salt concentration. X_1 and X_3 have conformations at two energy minima.

300 mM, which validated the higher largest cluster size at 0 mM NaCl concentration. The extended peptides (Figure 3a) and two near-terminal phosphorylated serine residues caused the formation of more interpeptide H-bonds at a 0 mM NaCl concentration. On the other hand, the higher number of intrapeptide H-bonds for 2CTD-2P-12P at 300 mM NaCl concentration (see Figure 5b) was expected as this peptide is more contracted at 300 mM NaCl in Figure 3c.

For the other three 2CTD crowded systems, the separation of intrapeptide and interpeptide H-bonds is shown in Figures S11−S13 in the Supporting Information. We observed a correlation between interpeptide H-bonds and clustering even for the contracted peptide conformations. For example, we saw a higher largest cluster size distribution for the 2CTD-2P-5P-9P-12P system, in which the peptides are contracted (Figure S9b), at a 150 mM NaCl concentration in Figure 1b and this is

in agreement with an increment in the number of interpeptide H-bonds distribution in Figure S12c and clustering of peptides is represented by the low energy conformations in Figure S9b. Moreover, the increments of the number of interpeptide H-bonds for 2CTD-2P-5P at 150 and 300 mM NaCl concentrations compared to the 0 mM NaCl concentration in Figure S13c agree with the largest cluster size distributions in Figure 1b−c, respectively.

A recent study by Flores-Solis. et al.,[17] which was based on both two-dimensional NMR experiments and MD simulations, showed that tyrosine−proline (Y−P) interactions play an important role in driving the LLPS in human Pol II CTD and other low-complexity proteins at physiological salt concentration without phosphorylation. On that note, we calculated contact maps to obtain a perspective of residue−residue interactions, which contributed to the clustering. Figure 6
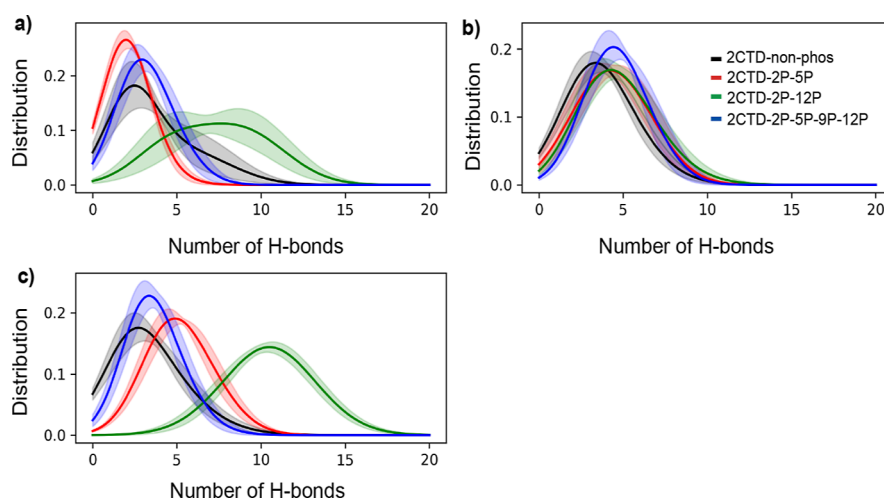
Figure 4. Distributions of the total number of H-bonds associated with peptides (both intra and inter) for 2CTD systems at 8.3 mM protein and (a) 0, (b) 150, and (c) 300 mM NaCl concentrations. The colors of the curves are the same as in panel (b) for all the other panels. Standard errors were calculated by splitting the last 800 ns of the 1 μs trajectory into 160 ns small trajectories for each system.

represents the contact maps for the four 2CTD crowded systems, which include only the interpeptide contacts at the physiological salt concentration (150 mM). For 2CTD-2P-5P-9P-12P, 2CTD-2P-12P, and 2CTD-2P-5P, we observed a large number of interpeptide contacts between residues specifically around phosphorylated serine residues in Figure 6b–d, respectively. This confirms that even with contracted peptide conformations, crowded systems can form interpeptide contacts to generate large cluster sizes. In addition to that, we observed additional enrichment of tyrosine interactions (including Y−P interactions) for the 2CTD-non-phos system rescaled in Figure S14b. To further understand the effects of each amino acid on intermolecular interactions and clustering, we performed three mutation simulations as control. In the first, second, and third controls, all the Y, T, and P in 2CTD-non-phos were mutated into A. We observed a decreased number of clustering when Y or P were mutated (Figure S15) and the number of contacts was overall decreased for these mutations (Figure S16). These results are consistent with the high number of contacts observed for Y and P in the contact maps for the wild-type 2CTD-non-phos simulations and also with the study of Flores-Solis. et al.,[17] which suggested the role of Y−P interactions in LLPS formation. On the other hand, we observed an increase in clustering for the T-A mutations. Contact map analysis shows that substituting T with A caused increased interactions of Y in the second repeat, especially with P, suggesting that T partly hinders Y−P interactions, potentially due to steric effects.
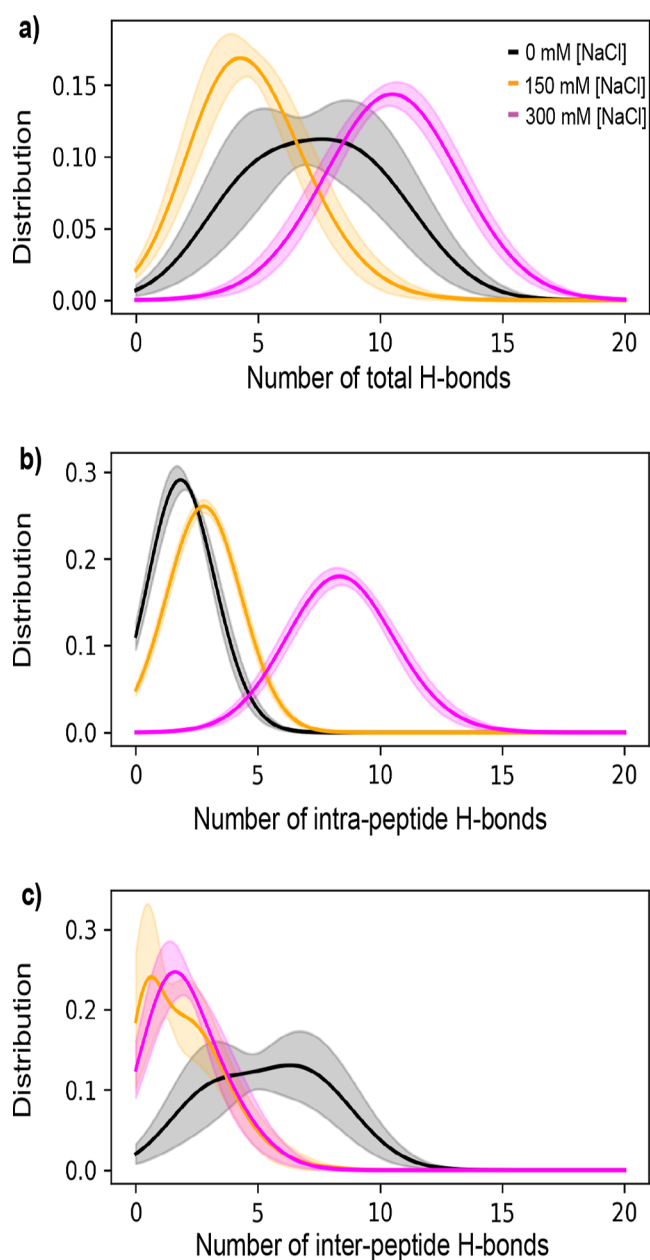
Next, we analyzed the concentration effect on the interpeptide interactions. Figure S14 shows the contact maps of 2CTD-non-phos at different CTD concentrations. The contact maps of 8.3 mM and larger concentrations (17 and 25 mM) have very similar trends, as we observed higher contacts of Y and P residues. This suggests that the interactions observed for small clusters (dimers or trimers) also play a role in the formation of larger clusters and potentially in the LLPS formation. However, in the low concentrations (2.5 and 1 mM), the number of contacts decreased substantially. For 2.5 mM, we still observed some contacts mostly for Y and P, while for 1 mM, all the CTDs stayed as monomers throughout the simulations, and no contacts were observed. This suggests that

the 1 mM concentration falls below the critical concentration for 2CTD-non-phos to form clusters.

To understand how electrostatic interactions play a role in the clustering of 2CTD peptides, we focus on phosphate groups and their interactions with Na$^+$ ions. In our earlier study,[28] we observed that Na$^+$ interacts with negatively charged oxygens, forming a bridge between multiple phosphate groups. To quantify Na$^+$ ion bridging for the phosphate groups, we generated the distributions of distances between interpeptide P atoms for the three phosphorylated 2CTD systems at the physiological salt concentration. Figure 7a shows that for all three phosphorylated 2CTD systems, there is a peak around 5 Å, which provides an ideal distance for Na$^+$ to form coordination between the phosphate groups of two different peptides. Figure 7b shows a snapshot with low energy conformations according to the PCA plot in Figure S9b for the 2CTD-2P-5P-9P-12P system. The Na$^+$ ions are in close distance with the phosphate groups from two segments (peptides) that are within around 5 Å distance, confirming the Na$^+$ ion bridging of the interpeptide phosphates. These electrostatic interactions help to form interpeptide contacts and stabilize the large clusters in phosphorylated 2CTD systems at the physiological salt concentration.

Overall Effects of Salt Concentration and Phosphorylation on Clustering. In general, for higher protein concentrations (17 and 25 mM), large clusters formed regardless of the phosphorylation state, especially at high salt concentrations with minimal deviations within the cluster distributions (Figures S2–S3). However, there are complex effects on clustering from phosphorylation and salt concentration specifically at the lower protein concentration. Figure 8a displays a heatmap representation of how the phosphorylation and salt concentration affect the average largest cluster size for the 2CTD systems at an 8.3 mM protein concentration. The average largest cluster size for 2CTD-non-phos was around below 3 at every NaCl concentration. Moreover, all three phosphorylated 2CTD systems formed higher average largest cluster sizes compared to the non-phosphorylated state at the physiological salt concentration (150 mM). Additionally, the clustering was decreased or unchanged with increasing salt concentration from 150 to 300
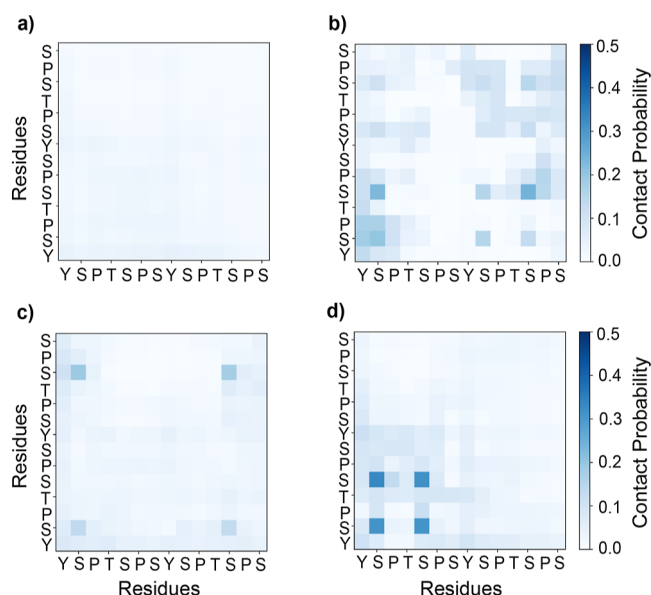
Figure 5. Distributions of the number of (a) total, (b) intrapeptide, and (c) interpeptide H-bonds for 2CTD-2P-12P systems at 8.3 mM protein and different salt concentrations. The colors of the curves are the same as in panel (a) for all of the other panels. Standard errors were calculated by splitting the last 800 ns of the 1 μs trajectory into 160 ns small trajectories for each system.
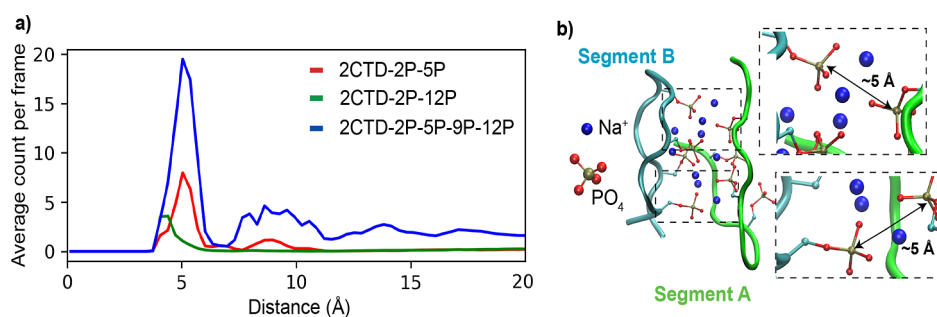

Figure 6. Contact maps for interpeptide contacts between residues within 5 Å at 8.3 mM protein and 150 mM NaCl concentrations for (a) 2CTD-non-phos, (b) 2CTD-2P-5P-9P-12P, (c) 2CTD-2P-12P, and (d) 2CTD-2P-5P systems. Contacts were averaged for each residue pair over the trajectory and the number of peptide pairs.

DISCUSSION

CTD of Pol II has been suggested to form clusters that lead to phase separation.[4,16,17] In this study, we investigated the clustering behavior of CTD models using all-atom MD simulations. The CTD models used in this study consisted of two heptapeptide repeats as simple models to provide insights into the intermolecular interactions that would potentially cause clustering in longer CTDs in humans or yeast. For the nonphosphorylated CTD, we observed cluster formation in a lesser amount than the phosphorylated CTD models. The most prevalent intermolecular interactions were observed with the tyrosine residues in the heptapeptide repeats when the serine residues are not phosphorylated. Since the intermolecular H-bond contribution is not substantial (Figure S11), hydrophobic interactions between the tyrosine ring and other residues may be the main contributor to the cluster formation observed for the 2CTD-non-phos. In addition, our mutation studies suggest that both tyrosine and proline contribute to the clustering of 2CTD-non-phos since mutating them into alanine reduced the cluster formation (Figure S15). The study by Flores-Solis et al. on CTD clustering by both experiments and simulations suggested that tyrosine interactions are abundant in CTD clustering.[17] They especially focused on Tyr−Pro interactions that are also observed by MD simulations in their study. Tyrosine interactions were also observed between CTD and the mediator complex that forms condensates with Pol II.[54] The predominance of the tyrosine interactions in the 2CTD-non-phos from our simulations is aligned with the literature that reports an abundance of tyrosine interactions in CTD clustering.

Nonphosphorylated CTDs were shown to form liquid droplets in vitro and phosphorylation caused the dispersion of these droplets.[4] Phosphorylated CTDs were also recruited to the droplets formed by other proteins, especially observed by kinases of positive transcription factor b (P-TEFb).[16] Such interchanges between different condensates as a function of

mM, suggesting that charge screening impacts the clustering behavior.

Figure 8b shows the general model which represents how the interpeptide interactions and ion coordinations induce the formation of the clusters. Extended peptides form clusters using both interpeptide interactions and ion coordination while contracted peptides form clusters using mainly ion coordination upon phosphorylation. Ion coordination induces interpeptide interactions in contracted peptides as confirmed by the contact maps in Figure 6. In summary, the clustering of 2CTD peptides depends on the phosphorylation patterns that support a combination of interpeptide H-bondings and electrostatic interactions formed via counterion bridging.

Figure 7. (a) Distributions for distances between interpeptide P atoms of phosphorylated 2CTDs at 8.3 mM protein and 150 mM NaCl concentrations. The average count per frame for each system was calculated by counting the number of occasions of two different peptides occupying the distance between P atoms at each specific distance bin and dividing it by the total number of frames. (b) Snapshot of the 2CTD-2P-5P-9P-12P system at 8.3 mM protein and 150 mM NaCl concentrations with low energy conformations representing the distances between two P atoms of two different peptides (segment A and segment B) allowing the phosphate groups to coordinate Na$^+$ ions to form clusters. Only Na$^+$ ions between segments within 3 Å of both segments are shown.
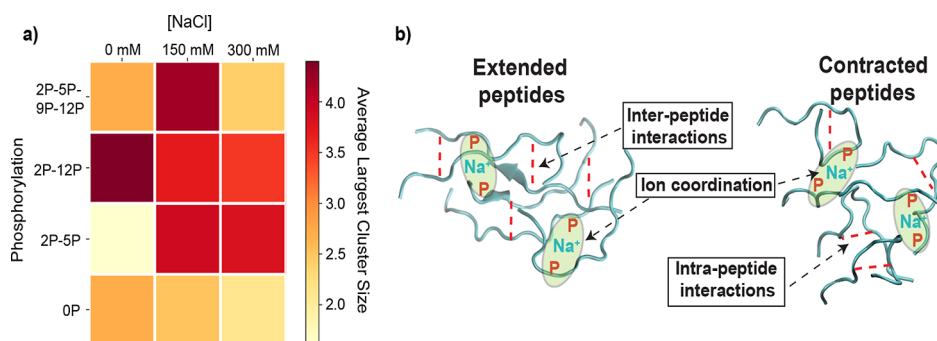
Figure 8. (a) Heatmap for the overall effect of phosphorylation level and NaCl concentration on the clustering of 2CTD systems at an 8.3 mM protein concentration. (b) General model representation of how the interpeptide interactions and ion coordination induce the formation of clusters with extended and contracted peptides.

phosphorylation level have been proposed as a mechanism by which post-translational modifications regulate the transition from initiation to the elongation stage.[21] However, the molecular interactions that change the clustering behavior of CTD upon phosphorylation are understudied. In our study, MD simulation results suggest that phosphorylation causes the formation of clusters via electrostatic interactions with Na$^+$ ions that bridge intermolecular phosphate groups. These newly formed electrostatic interactions increase the number of clusters compared to the 2CTD-non-phos system. Earlier studies suggested an interchange of the interactions from hydrophobic to electrostatics in nature upon phosphorylation of CTD.[4,16] However, these studies indicated the formation of electrostatic interactions between negatively charged phosphorylated CTD and other positively charged IDPs, without mentioning any electrostatic interactions mediated by ions. Our observation of Na$^+$ ion bridging could be related to the size of the CTD that a human or yeast CTD with higher numbers of heptapeptide repeats may behave differently compared to our 2CTD model with two heptapeptide repeats. On the other hand, we also note that ion-mediated condensations were observed for highly charged nucleic acids in the presence of the Mg^{2+} ion more prevalently,[55−57] but also observed for Na$^+$ ions in high concentration in crowded systems.[55,57] Therefore, the role of counterion interactions in clustering and condensate formation by phosphorylated CTDs might be underestimated and needs to be considered in future studies.

The ion-mediated interactions observed in our previous study[28] caused more compact structures for phosphorylated CTDs through bending to coordinate the counterion by multiple phosphate groups. Other computational works on IDPs showed similar observations that counterions mediate the inter- and intramolecular interactions.[58,59] All these studies were conducted using AMBER or CHARMM force fields, which have been systematically parametrized and extensively validated against experimental observations,[41,60−62] suggesting their strong ability to provide accurate results for intermolecular interactions. However, in a highly charged concentrated system, the static charges in the standard force fields may have limitations in describing the electrostatics upon the polarization of charged atoms induced by the environment. To account for charge redistribution due to the highly charged concentrated systems and to increase the accuracy of the results, polarizable force fields[63,64] could be used instead of standard force fields with fixed charges. Additionally, our observations of counterion-mediated clustering could be indirectly validated using experimental methods such as NMR or FRET. Increased clustering upon phosphorylation may decrease the NMR relaxation times or increase the FRET efficiencies, and the ion dependency could be monitored by modifying the ion concentration.

Earlier experimental studies showed that phosphorylation of CTD resulted in the dispersion of phases formed by CTD[4] or disassociation of CTD from mediator condensates.[21] However, those studies did not rule out the possibility of the formation of smaller clusters of phosphorylated CTDs that would not

induce phase separation and therefore could not be detected by microscopic techniques. Related to that note, recent studies suggested that driving forces for clustering and phase separation may not be fully correlated, and cluster formation was predicted at subsaturated solutions.[65,66] In light of these studies and based on our results, we suggest that phosphorylated CTDs with longer lengths may form clusters by ion-mediated interactions even in dispersed solutions, which can be analyzed by spectroscopic techniques like dynamic light scattering or FRET.

We performed this study to understand the basic interaction networks that cause clustering of CTD models at different phosphorylation levels. However, we note that our systems have two heptapeptide repeats, which are much shorter than the CTDs observed in humans and yeast, which have 52 and 26 repeats, respectively. Studies suggest that phase separation occurs when CTD has around more than 10 repeats, which is still far longer than our model systems.[4,22] However, our cluster analysis suggested that phase separation may occur for the 2CTD-non-phos system with a 25 mM protein concentration (Figure S5). Below this concentration, we observed cluster formation without any indication of phase separation, while at 25 mM protein concentration, we observed the coexistence of small and large clusters suggesting a formation of phase separation. Although this finding requires experimental validation and additional support from follow-up computational studies, phase separation of CTD with only two repeats at high concentrations may indicate a higher critical concentration for this system than CTD with ten or more heptapeptide repeats. To understand the effects of heptapeptide repeats in cluster formation and phase separation, we need to extend this study to the CTDs with more repeats. However, there are computational limitations in studying phase separation of longer CTD systems at the atomistic detail as the system sizes increase substantially with the increased number of repeats. One way to circumvent this limitation is to switch to coarse-grained representation as it was routinely done to study the phase separation of IDPs.[67−70]

CONCLUSIONS

In this computational study, we report the cluster formation of the 2CTD models upon phosphorylation at different salt concentrations. This study helped us to obtain a fundamental understanding of the intermolecular interactions that are potentially linked to LLPS formation of the Pol II CTD, which was reported previously in the literature. Phosphorylation caused complex effects on the clustering specifically at low protein concentrations depending on the salt concentration. For phosphorylated CTDs, clusters are either formed with contracted peptides or extended peptides by forming interpeptide interactions (H-bonds) and coordination with Na$^+$ ions in between phosphate groups of different peptides. At higher salt concentrations, charge screening occurs, especially for the multiple phosphorylated CTD systems. Future directions will focus on developing and applying coarse-grained models to study phase separation of full human RNA Pol II CTD in a computationally accessible time scale.

ASSOCIATED CONTENT

Data Availability Statement

We provided the scripts for the analyses of the simulations and the snapshots with the lowest energy conformations for each simulation at 8.3 mM protein concentrations in the GitHub repository (https://github.com/bercemd/ctd_clustering).

◆ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.4c04457.

MSD vs lag-times plot for 2CTD-non-phos; cluster distribution and cluster size percentages in 2CTD crowded models at 17 mM and 25 mM protein concentrations at different salt concentrations; comparison of cluster distribution and cluster size percentages for low concentrations and for all concentrations at logarithmic scale; average R_g density distributions for 2CTD crowded models at 17 mM and 25 mM protein concentrations at different salt concentrations; free energy landscapes from PCA and sets of low energy conformations for 2CTD-non-phos, 2CTD-2P-5P-9P-12P, and 2CTD-2P-5P at an 8.3 mM protein concentration and different salt concentrations; distributions of the number of total, intrapeptide, interpeptide H-bonds at an 8.3 mM protein concentration and at different salt concentrations; contact map for 2CTD-non-phos at 2.5, 8.3, 17, and 25 mM protein concentrations and 150 mM NaCl concentrations; cluster distribution and cluster size percentages of 2CTD-non-phos mutation simulations; contact map for 2CTD-non-phos mutation simulations; and self-diffusion coefficients of 17 and 25 mM 2CTDs at different salt concentrations and at different protein concentrations at a 150 mM salt concentration (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Bercem Dutagaci − Department of Molecular and Cell Biology, University of California, Merced, California 95343, United States; Health Sciences Research Institute, University of California, Merced, California 95343, United States; ⊙ orcid.org/0000-0003-0333-5757; Phone: 209-228-3603; Email: bdutagaci@ucmerced.edu

Authors

Weththasinghage D. Amith − Department of Molecular and Cell Biology, University of California, Merced, California 95343, United States

Vincent T. Chen − Department of Molecular and Cell Biology, University of California, Merced, California 95343, United States

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.4c04457

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Bernecky, C.; Plitzko, J. M.; Cramer, P. Structure of a transcribing RNA polymerase II−DSIF complex reveals a multi-dentate DNA−RNA clamp. *Nat. Struct. Mol. Biol.* **2017**, *24* (10), 809−815.

(2) Hantsche, M.; Cramer, P. The structural basis of transcription: 10 years after the Nobel Prize in chemistry. *Angew. Chem., Int. Ed.* **2016**, *55* (52), 15972−15981.

(3) Eick, D.; Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **2013**, *113* (11), 8456−8490.

(4) 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. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **2018**, *25* (9), 833−840.

(5) Conaway, R. C.; Bradsher, J. N.; Conaway, J. W. Mechanism of assembly of the RNA polymerase-II preinitiation complex - evidence for a functional interaction between the carboxyl-terminal domain of the largest subunit of RNA polymerase-II and a high molecular mass form of the TATA factor. *J. Biol. Chem.* **1992**, *267* (12), 8464−8467.

(6) Lu, H.; Flores, O.; Weinmann, R.; Reinberg, D. The nonphosphorylated form of RNA polymerase-II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (22), 10004−10008.

(7) Hengartner, C. J.; Myer, V. E.; Liao, S. M.; Wilson, C. J.; Koh, S. S.; Young, R. A. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* **1998**, *2* (1), 43−53.

(8) Liao, S. M.; Zhang, J. H.; Jeffery, D. A.; Koleske, A. J.; Thompson, C. M.; Chao, D. M.; Viljoen, M.; Vanvuuren, H. J. J.; Young, R. A. A kinase-cyclin pair in the RNA-polymerase-II holoenzyme. *Nature* **1995**, *374* (6518), 193−196.

(9) Bartkowiak, B.; Greenleaf, A. L. Phosphorylation of RNAPII To P-TEFb or not to P-TEFb? *Transcription* **2011**, *2* (3), 115−119.

(10) Parada, C. A.; Roeder, R. G. Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. *Nature* **1996**, *384* (6607), 375−378.

(11) Hirose, Y.; Tacke, R.; Manley, J. L. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **1999**, *13* (10), 1234−1239.

(12) Rosonina, E.; Blencowe, B. J. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *RNA* **2004**, *10* (4), 581−589.

(13) Ansari, A.; Hampsey, M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **2005**, *19* (24), 2969−2978.

(14) Richard, P.; Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes Dev.* **2009**, *23* (11), 1247−1269.

(15) Harlen, K. M.; Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (4), 263−273.

(16) Lu, H.; Yu, D.; Hansen, A. S.; Ganguly, S.; Liu, R.; Heckert, A.; Darzacq, X.; Zhou, Q. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **2018**, *558* (7709), 318−323.

(17) Flores-Solis, D.; Lushpinskaia, I. P.; Polyansky, A. A.; Changiarath, A.; Boehning, M.; Mirkovic, M.; Walshe, J.; Pietrek, L. M.; Cramer, P.; Stelzl, L. S.; et al. Driving forces behind phase separation of the carboxy-terminal domain of RNA polymerase II. *Nat. Commun.* **2023**, *14* (1), 5979.

(18) Guo, C. H.; Luo, Z. J.; Lin, C. Q. Phase separation properties in transcriptional organization. *Biochemistry* **2022**, *61* (22), 2456−2460.

(19) Nandana, V.; Schrader, J. M. Roles of liquid−liquid phase separation in bacterial RNA metabolism. *Curr. Opin. Microbiol.* **2021**, *61*, 91−98.

(20) Shao, W.; Bi, X. J.; Pan, Y. X.; Gao, B. Y.; Wu, J.; Yin, Y. F.; Liu, Z. M.; Peng, M. Y.; Zhang, W. H.; Jiang, X.; et al. Phase separation of RNA-binding protein promotes polymerase binding and transcription. *Nat. Chem. Biol.* **2022**, *18* (1), 70−80.

(21) Guo, Y. E.; Manteiga, J. C.; Henninger, J. E.; Sabari, B. R.; Dall'Agnese, A.; Hannett, N. M.; Spille, J. H.; Afeyan, L. K.; Zamudio, A. V.; Shrinivas, K.; et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **2019**, *572* (7770), 543−548.

(22) Kwon, I.; Kato, M.; Xiang, S. H.; Wu, L.; Theodoropoulos, P.; Mirzaei, H.; Han, T.; Xie, S. H.; Corden, J. L.; McKnight, S. L. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **2013**, *155* (5), 1049−1060.

(23) Portz, B.; Shorter, J. Switching condensates: The CTD code goes liquid. *Trends Biochem. Sci.* **2020**, *45* (1), 1−3.

(24) Gibbs, E. B.; Lu, F. Y.; Portz, B.; Fisher, M. J.; Medellin, B. P.; Laremore, T. N.; Zhang, Y. J.; Gilmour, D. S.; Showalter, S. A. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* **2017**, *8*, 15233.

(25) Janke, A. M.; Seo, D. H.; Rahmanian, V.; Conicella, A. E.; Mathews, K. L.; Burke, K. A.; Mittal, J.; Fawzi, N. L. Lysines in the RNA polymerase II C-terminal domain contribute to TAF15 fibril recruitment. *Biochemistry* **2018**, *57* (17), 2549−2563.

(26) Portz, B.; Lu, F. Y.; Gibbs, E. B.; Mayfield, J. E.; Rachel Mehaffey, M.; Zhang, Y. J.; Brodbelt, J. S.; Showalter, S. A.; Gilmour, D. S. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* **2017**, *8*, 15231.

(27) Meinhart, A.; Kamenski, T.; Hoeppner, S.; Baumli, S.; Cramer, P. A structural perspective of CTD function. *Genes Dev.* **2005**, *19* (12), 1401−1415.

(28) Amith, W. D.; Dutagaci, B. Complex conformational space of the RNA polymerase II C-terminal domain upon phosphorylation. *J. Phys. Chem. B* **2023**, *127* (43), 9223−9235.

(29) Bhattacharya, A.; Kim, Y. C.; Mittal, J. Protein−protein interactions in a crowded environment. *Biophys. Rev.* **2013**, *5* (2), 99−108.

(30) Speer, S. L.; Stewart, C. J.; Sapir, L.; Harries, D.; Pielak, G. J. Macromolecular crowding is more than hard-core repulsions. *Annu. Rev. Biophys.* **2022**, *51* (1), 267−300.

(31) Szasz, C.; Alexa, A.; Toth, K.; Rakacs, M.; Langowski, J.; Tompa, P. Protein disorder prevails under crowded conditions. *Biochemistry* **2011**, *50* (26), 5834−5844.

(32) Cubuk, J.; Soranno, A. Macromolecular crowding and intrinsically disordered proteins: A polymer physics perspective. *ChemSystemsChem.* **2022**, *4* (5), No. e202100051.

(33) Nawrocki, G.; Wang, P. H.; Yu, I.; Sugita, Y.; Feig, M. Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B* **2017**, *121* (49), 11072−11084.

(34) Harada, R.; Sugita, Y.; Feig, M. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* **2012**, *134* (10), 4842−4849.

(35) Heo, L.; Sugita, Y.; Feig, M. Protein assembly and crowding simulations. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102340.

(36) Javanainen, M.; Ollila, O. H. S.; Martinez-Seara, H. Rotational diffusion of membrane proteins in crowded membranes. *J. Phys. Chem. B* **2020**, *124* (15), 2994−3001.

(37) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.

(38) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859−1865.

(39) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405−413.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(41) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14* (1), 71−73.

(42) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.

(43) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.

(44) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; et al. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* **2013**, *9* (1), 461−469.

(45) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319−2327.

(46) Dutagaci, B.; Nawrocki, G.; Goodluck, J.; Ashkarran, A. A.; Hoogstraten, C. G.; Lapidus, L. J.; Feig, M. Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. *Elife* **2021**, *10*, No. e64004.

(47) Calandrini, V.; Pellegrini, E.; Calligari, P.; Hinsen, K.; Kneller, G. R. nMoldyn - Interfacing spectroscopic experiments, molecular dynamics simulations and models for time correlation functions. *Collection SFN* **2011**, *12*, 201−232.

(48) Buyl, P. tidynamics: A tiny package to compute the dynamics of stochastic and molecular simulations. *J. Open Source Softw.* **2018**, *3* (28), 877.

(49) Grossfield, A. *WHAM: the weighted histogram analysis method.* version 2.0.11.

(50) *MATLAB optimization toolbox* The MathWorks, Inc., Natick, Massachusetts, United States.

(51) Garaizar, A.; Espinosa, J. R.; Joseph, J. A.; Collepardo-Guevara, R. Kinetic interplay between droplet maturation and coalescence modulates shape of aged protein condensates. *Sci. Rep.* **2022**, *12* (1), 4390.

(52) Muramatsu, N.; Minton, A. P. Tracer diffusion of globular-proteins in concentrated protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85* (9), 2984−2988.

(53) Ranganathan, S.; Shakhnovich, E. The physics of liquid-to-solid transitions in multi- domain protein condensates. *Biophys. J.* **2022**, *121* (14), 2751−2766.

(54) Chen, X. Z.; Yin, X. T.; Li, J. B.; Wu, Z. H.; Qi, Y. L.; Wang, X. X.; Liu, W. D.; Xu, Y. H. Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* **2021**, *372* (6546), No. eabg0635.

(55) Onuchic, P. L.; Milin, A. N.; Alshareedah, I.; Deniz, A. A.; Banerjee, P. R. Divalent cations can control a switch-like behavior in heterotypic and homotypic RNA coacervates. *Sci. Rep.* **2019**, *9*, 12161.

(56) Tom, J. K. A.; Onuchic, P. L.; Deniz, A. A. Short polyA RNA homopolymers undergo Mg2+-mediated kinetically arrested condensation. *J. Phys. Chem. B* **2022**, *126* (46), 9715−9725.

(57) Van Treeck, B.; Protter, D. S. W.; Matheny, T.; Khong, A.; Link, C. D.; Parker, R. RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (11), 2734−2739.

(58) Man, V. H.; He, X. B.; Han, F. Y.; Cai, L. J.; Wang, L. X.; Niu, T. Y.; Zhai, J. C.; Ji, B. H.; Gao, J.; Wang, J. M. Phosphorylation at Ser289 enhances the oligomerization of Tau repeat R2. *J. Chem. Inf. Model.* **2023**, *63* (4), 1351−1361.

(59) Marien, J.; Prevost, C.; Sacquin-Mora, S. nP-collabs: Investigating counterion mediated bridges in the multiply phosphorylated tau-R2 repeat. *J. Chem. Inf. Model.* **2024**, *64*, 6570−6582.

(60) Mu, J. X.; Liu, H.; Zhang, J.; Luo, R. A.; Chen, H. F. Recent force field strategies for intrinsically disordered proteins. *J. Chem. Inf. Model.* **2021**, *61* (3), 1037−1047.

(61) Stoppelman, J. P.; Ng, T. T.; Nerenberg, P. S.; Wang, L. P. Development and validation of AMBER-FB15-compatible force field

parameters for phosphorylated amino acids. *J. Phys. Chem. B* **2021**, *125* (43), 11927−11942.

(62) Zerze, G. H.; Zheng, W. W.; Best, R. B.; Mittal, J. Evolution of all-atom protein force fields to improve local and global properties. *J. Phys. Chem. Lett.* **2019**, *10* (9), 2227−2234.

(63) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An empirical polarizable force field based on the classical drude oscillator model: Development history and recent applications. *Chem. Rev.* **2016**, *116* (9), 4983−5013.

(64) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable force fields: History, test cases, and prospects. *J. Chem. Theory Comput.* **2007**, *3* (6), 2034−2045.

(65) Kar, M.; Dar, F.; Welsh, T. J.; Vogel, L. T.; Kühnemuth, R.; Majumdar, A.; Krainer, G.; Franzmann, T. M.; Alberti, S.; Seidel, C. A. M.; et al. Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (28), No. e2202222119.

(66) Li, P. L.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483* (7389), 336−340.

(67) Dignon, G. L.; Zheng, W. W.; Mittal, J. Simulation methods for liquid-liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.* **2019**, *23*, 92−98.

(68) Regy, R. M.; Zheng, W. W.; Mittal, J. Using a sequence-specific coarse-grained model for studying protein liquid-liquid phase separation. *Methods Enzymol.* **2021**, *646*, 1−17.

(69) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (44), No. e2111696118.

(70) Valdes-Garcia, G.; Heo, L.; Lapidus, L. J.; Feig, M. Modeling concentration-dependent phase separation processes involving peptides and RNA via residue-based coarse-graining. *J. Chem. Theory Comput.* **2023**, *19* (2), 669−678.