

RESEARCH ARTICLE

Thermodynamic and sequential characteristics of phase separation and droplet formation for an intrinsically disordered region/protein ensemble

Wen-Ting Chu¹, Jin Wang^{2*}

1 State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, People's Republic of China, **2** Department of Chemistry & Physics, State University of New York at Stony Brook, Stony Brook, New York, United States of America

* jin.wang.1@stonybrook.edu



OPEN ACCESS

Citation: Chu W-T, Wang J (2021) Thermodynamic and sequential characteristics of phase separation and droplet formation for an intrinsically disordered region/protein ensemble. *PLoS Comput Biol* 17(3): e1008672. <https://doi.org/10.1371/journal.pcbi.1008672>

Editor: Alexander MacKerell, University of Maryland School of Pharmacy, UNITED STATES

Received: August 27, 2020

Accepted: January 4, 2021

Published: March 8, 2021

Copyright: © 2021 Chu, Wang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All raw data and analysis files of density calculation are available from the Zenodo database (DOI: [10.5281/zenodo.4383333](https://doi.org/10.5281/zenodo.4383333)).

Funding: W.-T.C. received the support from National Natural Science Foundation of China Grants (21603217, 21721003), Youth Innovation Promotion Association CAS Grant (2020231), Ministry of Science and Technology of China Grant (2016YFA0203200), Scientific Instrument Developing Project of the Chinese Academy of

Abstract

Liquid–liquid phase separation (LLPS) of some IDPs/IDRs can lead to the formation of the membraneless organelles *in vitro* and *in vivo*, which are essential for many biological processes in the cell. Here we select three different IDR segments of chaperon Swc5 and develop a polymeric slab model at the residue-level. By performing the molecular dynamics simulations, LLPS can be observed at low temperatures even without charge interactions and disappear at high temperatures. Both the sequence length and the charge pattern of the Swc5 segments can influence the critical temperature of LLPS. The results suggest that the effects of the electrostatic interactions on the LLPS behaviors can change significantly with the ratios and distributions of the charged residues, especially the sequence charge decoration (SCD) values. In addition, three different forms of swc conformation can be distinguished on the phase diagram, which is different from the conventional behavior of the free IDP/IDR. Both the packed form (the condensed-phase) and the dispersed form (the dilute-phase) of swc chains are found to be coexisted when LLPS occurs. They change to the fully-spread form at high temperatures. These findings will be helpful for the investigation of the IDP/IDR ensemble behaviors as well as the fundamental mechanism of the LLPS process in bio-systems.

Author summary

The membraneless organelles caused by liquid–liquid phase separation (LLPS) of IDPs/IDRs are involved in a wide range of biological functions such as RNA processing, ribosome biogenesis, and sequestration of mRNA, proteins, and compacted chromatin. In this study, we focus on the histone H2A-H2B binding partner, Swc5 and investigate the effects of the temperature, sequence length and number of charged residues on the LLPS behaviors. Moreover, we proposed three forms of swc conformation in ensemble according to the phase diagram. These characteristics of conformational changes (from LLPS to no

Sciences Grant (YJKYYQ20180038), and Jilin Province Science and Technology Development Plan Grant (20200301009RQ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

LLPS, from condensed-phase to dilute-phase) are observed in the swc chains with different sequence length and charge pattern, which may be the general property for the IDP/IDR ensembles.

Introduction

In living cells, there is one special kind of organelles which are lack of the enclosing membrane. They are named as membraneless organelles (MLOs) [1–3], such as nucleoli, stress granules, P bodies, pericentriolar material and germ granules [3–6]. The MLOs are widely spread within the cell nucleus and the cytoplasm, and involved in a wide range of biological functions including RNA processing, ribosome biogenesis, and sequestration of mRNA (for later translation), proteins (for signaling), and compacted chromatin (for gene silencing) [7]. The recent studies revealed that the MLOs behave like liquid droplets with higher density than the surroundings, which can be observed directly via microscopy [4, 8]. The distinct subcellular compartments were deduced to be formed through a process known as liquid–liquid phase separation (LLPS) [5, 9, 10].

LLPS can be found in the solution of one or more kinds of proteins, proteins and RNAs mixture, proteins with crowders *in vitro* [11]. One class of these proteins is the folded proteins with multivalent interactions, including the signaling protein WASP [8]; the other class is the proteins with intrinsically disordered regions (IDRs), such as FUS [12], TDP-43 [13], hnRNPA1/hnRNPA2 [6, 14], Ddx4 [2], LAF-1 [15], etc. Experimental results suggest that the LLPS is sensitive to the temperature and salt concentration. The droplets can easily dissolve upon raising the temperature or salt concentration, and can reform when conditions are reverted. The reversibility comes because the two phases, condensed-phase (droplet) and dilute-phase (dispersed), are in thermodynamic equilibrium [10, 16]. Moreover, the reversible LLPS can be changed to the irreversible pathological aggregates by mutation, phosphorylation, methylation or ubiquitination on the proteins, forming fibrillization that may be a key reason of the neurodegenerative diseases such as Alzheimer's, ALS, Parkinson's, and cataract [6, 12, 17, 18]. As a result, protein sequence is also important for the properties of LLPS.

Intrinsically disordered proteins (IDPs) or IDRs contain much more charged residues than folded proteins/regions, thus are often studied in exploring of the relationship of protein dynamics and function such as in signal transduction and gene regulation [19–22]. However, things may be different for the ensemble/solution of many IDP/IDR molecules. In addition, the experiment has revealed that the distribution of the charged residues in Ddx4 determines its LLPS behavior [2]. Combining the Monte Carlo simulations with the extended Flory-Huggins (FH) theory (random-phase-approximation (RPA) polymer theory) [23–25], Das *et al.* constructed a coarse-grained model for the ensemble of the proteins with designed charge distributions to investigate the dependence of the LLPS on the charge pattern of the protein [26, 27]. However, these chains are all ideal ones with identical numbers of positive-charged residues and negative-charged residues. Then Dignon *et al.* discussed the phase characteristics of FUS, LAF-1, hnRNPA2, TDP-43, as well as their variants by using the *Debye-Hückel* model [28, 29] for charge–charge interactions [30, 31].

Swc5 as a chaperon protein, which interacts with the histones dimer H2A-H2B and participates in the eviction step of the exchange between H2A and H2A.Z [32], contains a 147 amino acids IDR (1–147) on its N-terminus. The first 79 amino acids (1–79) with the conserved DEF/Y motifs (DEF/Y-1 and DEF/Y-2) have been found to be essential for the binding with H2A-H2B [33]. The Swc5 is a subunit of SWR essential for its H2A.Z deposition activity *in*

vivo and *in vitro* [32–34]. Recently, the crystal structure of the histone binding domain of Swc5 in complex with an H2A-H2B dimer has been resolved [33], which indicates that the consecutive acidic residues and flanking hydrophobic residues of swc5 form a cap over the histones, excluding histone-DNA interaction. It supports a model in which swc5 acts as a wedge to promote H2A-H2B dimer eviction. Huang *et al.* showed that the IDR of Swc5 may recruit or retain histones within phase-separated liquid droplets where the transcription factors and the chromatin remodeling factors coalesce to regulate transcription [33]. The charged-residues on the DEF/Y motifs of Swc5 may play an important role in the binding process and phase behavior of Swc5. However, by far there is lack of investigations on the phase separation of Swc5 or other Swc family members. In order to find out the role of the IDR and the DEF/Y motifs in the LLPS of Swc5, we select the first 79 amino acids of Swc5 IDR (denoted as swc 1–79) and cut it into one part with the DEF/Y motifs (swc 1–32) as well as the other part without the DEF/Y motifs (swc 33–79). This setup is expected to capture the main function of Swc5 based on the more reliable structure information of segments. In this study we will take efforts to explore two important issues via molecular simulations: the temperature, sequence length, and charge pattern dependence of phase separation and droplet formation of swc chain ensemble; the different conformational forms of swc chain ensemble under special phase (trinity form). Consequently, this study will uncover the underlying mechanisms of LLPS in bio-molecules and provide helpful insights into the further research of the role of Swc5 in the assembly of histone/chromatin and DNA replication.

Materials and methods

Slab simulation model

We developed a coarse-grained model for each swc sequence, representing each residue of the complex by one bead on the C_{α} atom. Here in this study we selected three different kinds of Swc5 segments, swc 1–79, swc 33–79, and swc 1–32 (see Fig 1). Therefore, the sequence length (or the bead number of each chain, denote as L) is 79, 47, and 32, respectively. Then three different simulation systems were built for these swc chains. In each system, 200 number of identical swc chains ($n = 200$) were added in the simulation box. Therefore the total number of beads ($N = n \times L$) in each system is 15800, 9400, and 6400, respectively. Besides, we tested the Langevin dynamics simulations of two model sequences, sv1 and sv15 (from [35] and [26], 50 amino acids in each sequence), with the same simulation settings.

In this study we introduce a potential energy function of the system, which includes bonded, pairwise interactions (12-6 Lennard-Jones (LJ) potential), as well as the electrostatic interactions (Debye-Hückel potential). The mean bond length is 3.8 Å ($a = 3.8$ Å). Here in this study, we used the standard Lennard-Jones potential $\varepsilon \left[\left(\frac{a}{r} \right)^{12} - \left(\frac{a}{r} \right)^6 \right]$. The parameters σ and ε were carefully calibrated with the single chain of swc 1–79 before the long-time Langevin dynamics simulation of 200 chains system. Swc5 consists of an intrinsically disordered region (IDR) on the N-terminal half of the 303-amino acid polypeptide [33]. Because there are no structures of *apo* Swc5 available (the only crystal structure 6KBB is H2A-H2B-bound swc 15–32), we built the swc 1–79 model according to the properties of common isolated IDP (random coil). Firstly, parameter σ is the “finite distance”. σ_{ij} is the optimal distance between beads i and j that are in contact with each other. We calculated the non-local distances (e.g. 1–5) of the parts of the swc chains in 6KBB without the secondary structures (coil parts). These distance values are from 8 to 12 Å. Therefore, we set the parameter σ to 10 Å (about 2.6 a , a is the mean bond length). The parameter ε is the well depth and the energy responsible for the strength of a structure. Lower ε will increase the flexibility of the structure. In addition, we performed a

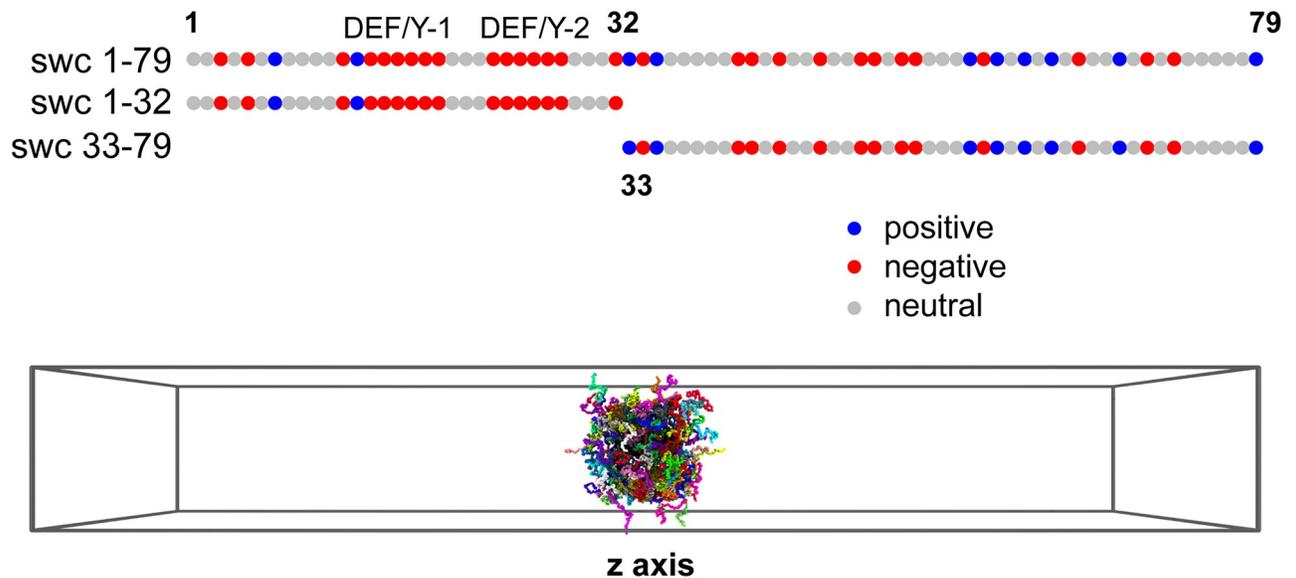


Fig 1. The three swc chains prepared for the simulations (top) and the slab simulation model (bottom). Top: one bead represents one residue of protein. Positive-charged, negative-charged, and neutral residues are shown as blue, red, and gray beads, respectively. The DEF/Y-1 and DEF/Y-2 motifs are labeled on the sequence. Bottom: in the initial state of the long-time simulation, 200 swc chains ($n = 200$, same sequence) were added into the periodic rectangular box (z axis = 300 nm).

<https://doi.org/10.1371/journal.pcbi.1008672.g001>

series of test Langevin dynamics simulations on swc 1–79 and found the LJ potential and coulomb potential at the same level when the parameter ϵ equals to 0.01 kJ/mol ($\sigma = 1.0$ nm). Then the head–end distance (D) show that the isolated swc 1–79 behaves like a flexible IDP when $\sigma = 1.0$ nm and $\epsilon = 0.01$ kJ/mol (see Supporting information S1 Fig).

The electrostatic interaction is calculated by the *Debye–Hückel* model [20, 28, 29, 36, 37], which can be used to quantify the strength of the charge–charge attractions and repulsions in varying ionic strengths:

$$V_{\text{Debye-Hückel}} = K_{\text{coulomb}} B(\kappa) \sum_{i,j} \frac{q_i q_j \exp(-\kappa r_{ij})}{\epsilon r_{ij}} \quad (1)$$

In Eq 1, $K_{\text{coulomb}} = 138.94 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm} \cdot e^{-2}$ is the electric conversion factor; $B(\kappa)$ is the salt-dependent coefficient; κ^{-1} is the Debye screening length, which is directly influenced by the solvent ionic strength (IS)/salt concentration C_{salt} ($\kappa \approx 3.2 \sqrt{C_{\text{salt}}}$); ϵ is the dielectric constant, which was set to 80 during the simulations to mimic the solvent medium (water). In our model, Lys and Arg have a positive point charge (+e), Asp and Glu have a negative point charge (−e). By using this model, the strength of the electrostatic interactions can be affected by the salt solvent through the screening effect. In order to investigate the role of electrostatic interactions, two extreme cases were designed for each system. One is 10 mM salt concentration ($C_{\text{salt}} = 0.01$ M), which represents the system with extremely high strength of electrostatic interactions; the other is infinite high salt concentration, which means that all the electrostatic interactions are screened by the solvent (no charges/electrostatic interactions).

In order to observe the liquid–liquid phase separation (LLPS), determine the phase diagram, and save the computational costs, the slab model was introduced for the simulations [30, 31, 38, 39]. Similar simulation strategy is applied as the previous studies [30, 31]. The slab model has a very long z axis (about 8–12 times of the length of x/y axis), thus the density distributions and the phase transitions can be explored in one dimension. After the initial

minimization, each system was equilibrated for 10 ns at 150 K in NPT ensemble, with periodic boundary conditions and 1.0 bar pressure (Parrinello-Rahman barostat) [40]. The system becomes a packed cubic box after the NPT simulation. Then we change the shape of the periodic simulation box by elongating the z dimension to 300 nm ($z = 300$ nm). Consider the different sequence length, we set three different lengths of x, y axes for swc 1–79 ($x = y = 40$ nm), swc 33–79 ($x = y = 30$ nm), and swc 1–32 ($x = y = 25$ nm) to ensure the similar ratio of bead number and volume. Because the sequence length of both sv1 and sv15 is 50, similar to swc 33–79, the simulation box of sv1/sv15 is the same as that of swc 33–79. Then a 5 μ s long-time Langevin dynamics run in NVT ensemble was performed on each system, keeping the total protein density constant. All simulations were performed with Gromacs 4.5.5 [41] with 2.0 fs time step and 1.0 ps⁻¹ friction coefficient of Langevin thermostat. In this study, we use a simplified coarse-grained with reduced units ($k_B T = 1$). Therefore, the simulation temperature/time is in reduced unit and does not equal to the real temperature/time. The simulation temperature was set from 100 K to 400 K (100, 150, 200, 300, and 400 K) in Gromacs to uncover its effect on LLPS. Here we set the unit temperature (T_0) and unit time (τ) to 100 K and 1 ns in Gromacs. As a result, the simulation temperatures correspond to 1.0, 1.5, 2.0, 3.0, and 4.0 T_0 and the simulation length of each trajectory corresponds to 5000 τ . Generally, for folded protein, one can estimate the physiological temperature through the ratio of experimental melting temperature and the simulated folding temperature [20, 42]. However, Swc5 is a IDP that does not fold at room temperature. As a result, the model used in this study was developed for IDP chains. The simulation temperature cannot be transformed to the real temperature directly. But the trends of simulated T_{cr} values can reveal the probability of LLPS behaviors.

Data analysis

To determine the phase distribution of the swc system, for each frame of the trajectory we select the z axis as the reaction coordinate and cut it into 30 windows (γ , the length of γ is 10 nm). Then we count the number of the beads in each window (denote as m_γ). For each swc system has different total number of beads (N), here we calculate the distribution of the fraction of beads ($P_\gamma = \frac{m_\gamma}{N}$, $\gamma = 1..30$) and find out the highest and the lowest values of this distribution (P_H and P_L). If LLPS occurs, the condensed-phase and dilute-phase are coexisted in the system. As a result, there is a large difference between P_H and P_L . In addition, it should be noted that at the same P_γ , the density of swc 1–79 is much higher than that of swc 33–79 and swc 1–32 because of the large total number of beads (N). As LLPS disappears, the system becomes uniform and the difference between P_H and P_L gets smaller. However, we can not obtain identical P_H and P_L because of the limits of the sampling. Therefore, the P_H and P_L values will get close at high simulation temperature.

For each frame of the simulation trajectory, the head-end distance (D) was calculated for the first and the last bead of each chain (e.g. distance between beads 1 and 79 for swc 1–79). In order to quantify the strength of the electrostatic interactions of the swc chains with and without LLPS, we counted the number of intra-chain electrostatic attractions (E_{intra}) of each chain and inter-chain electrostatic attractions (E_{inter}) of each chain–chain pair. In detail, if the distance between two bead i and j with opposite charges is lower than the cut-off distance (1.2 nm, 1.2 times of σ in LJ potential), we denote that there is one electrostatic attraction between the two beads. If i and j belong to the same chain, this interaction is collected as the intra-chain electrostatic attraction; if i and j belong to different chains, this interaction is collected as the inter-chain electrostatic attraction. In addition, here we only calculated the electrostatic interactions of the system at 10 mM salt concentrations since all the electrostatic interactions are

screened by the solvent in the other case (no charge). Therefore, 200 D , E_{intra} , and E_{inter} values can be collected for each frame of the simulation trajectory.

In order to show the distribution of all the 200 chains, we set the center of the system (condensed-phase with LLPS) at the zero point of x , y , and z axes. All the chains were located on both negative and positive sides of the zero point. Then we calculated the distance/displacement of each chain (center of mass (CM)) to the zero point on the z axis ($|z|_n = |z_n - z_0|$, $n = 1..200$), which can reveal the distribution of the 200 chains from the center of the system. Since the length of the simulation box on the z axis is 300 nm, the largest $|z|$ is 150 nm. We then selected the $|z|$ as the reaction coordinate and cut it into 30 windows. The fraction of the number of the chains that belong to each window, the standard error of the $|z|$ values in each window, as well as the distribution of D , E_{intra} , and E_{inter} along the reaction coordinate.

Here in this study we also calculated the radial distribution functions (rdf). The rdf ($g(r)$) defines the probability of finding a group of atoms at a distance r from another group of atoms. The r value corresponds to the peak value of $g(r)$ curve shows the distance between the two groups with the highest probability. As a result, we simply calculated the $g(r)$ curves between chain 1 and another chain (chain 2, 3, . . . , 200) and collected the r values corresponding to the peak value of $g(r)$ curves (peak1).

Results

Temperature, sequence length, and charge pattern affect the LLPS of swc

Three different kinds of swc chains were selected for the simulations of the assembling and diffusing processes at different conditions. As shown in Fig 1, first 79 amino acids of swc (denote as swc 1–79) is rich in charged amino acids, including 29 negative-charged amino acids and 10 positive-charged amino acids. The front 32 amino acids form the shorter part of swc 1–79, swc 1–32, which includes the conserved DEF/Y-1 and DEF/Y-2 motifs. Swc 1–32 has only 2 positive-charged amino acids, but many negative-charged amino acids (16). The back-end 47 amino acids were truncated into the longer part of swc 1–79, swc 33–79, which has 13 negative-charged amino acids and 8 positive-charged amino acids.

We use one bead to represent one residue of proteins. To observe the phase behaviors of different swc chains, a model with many identical swc chains was built for the molecular simulation. For each simulation, 200 chains ($n = 200$) of swc (swc 1–32 or 33–79 or 1–79) were added into the periodic rectangular box to build the slab simulation model (see Fig 1), which has short x and y axes but an elongated z axis [30, 31]. The simulation boxes of swc 1–32, swc 33–79, swc 1–79 has the same length of z axis (300 nm, about 750 a), but different lengths of x and y axes ($x = y = 25, 30$, and 40 nm). Each protein can be considered as a polymer, which randomly walks in the simulation box but show regular behaviors under special conditions.

Simulations were performed under different temperatures from 1.0 to 4.0 T_0 (1.0, 1.5, 2.0, 3.0, and 4.0 T_0 , in reduced units) and different salt concentrations of solvent (10 mM and extremely high salt concentration, corresponding to high and no strengths of electrostatic interactions in the system, denoted as 10 mM and no charge below), respectively. After a long-time equilibration of 5000 τ simulation, we can observe LLPS of swc chains at low temperatures in varying degrees. The analysis of the radial distribution function curves ($g(r)$) can show the general distribution of chain–chain distance. It is obvious that the chain–chain distances of the system with LLPS (low temperature) tend to be lower than that without LLPS (high temperature), see Supporting information S2–S4 Figs. Here we calculate the distribution of the fraction of the swc residues ($P_\gamma = \frac{m_\gamma}{N}$, $\gamma = 1..30$, see the Materials and methods section) along the long z axis, as well as the highest and the lowest probability of this distribution (P_H and P_L). If there is large difference between P_H and P_L , LLPS occurred in the system. As shown in Fig 2,

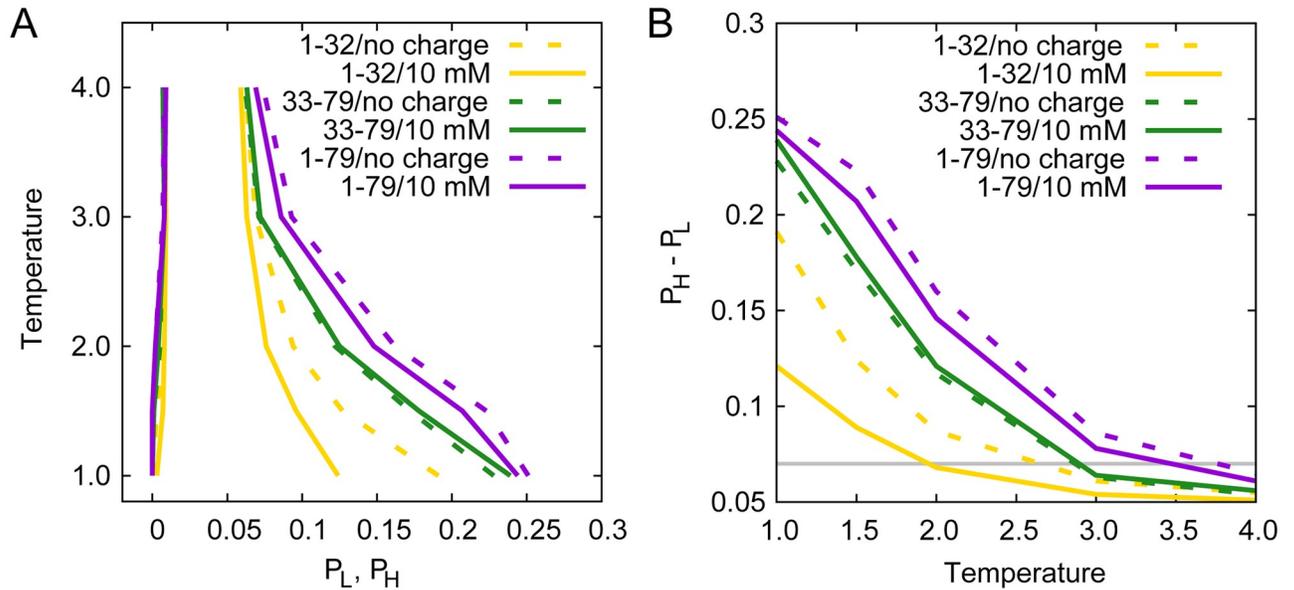


Fig 2. The appearance and disappearance of phase separation. (A) Phase diagram of swc chains in different solvents. (B) $P_H - P_L$ changes with temperature and solvent. Here P_H and P_L are the highest and the lowest points of swc residue distribution along the z axis. The P_H and P_L values are calculated with the last 1000 τ simulation data, as $P_H - P_L$ of all the simulations reaches equilibrium after 3000 τ (see Supporting information S5–S7 Figs). The $P_H - P_L$ data are listed in Supporting information S1 Table.

<https://doi.org/10.1371/journal.pcbi.1008672.g002>

the simulation results suggest the obvious appearance of LLPS for swc 1–79 and swc 33–79 at the temperatures below $1.5 T_0$ ($P_H - P_L > 0.15$). LLPS can be found for swc 1–32 at $1.0 T_0$ if there are no charge interactions in the system (no charge). However, charge interactions of swc 1–32 can have a large effect on its distribution, leading to no obvious LLPS at 1.0 and $1.5 T_0$ of swc 1–32 at 10 mM salt concentration.

We should notice that P_γ represents the fraction of all the polymer beads/protein residues. The total number of beads of swc 1–79 ($N = 15800$) is much higher than that of swc 33–79 ($N = 9400$) and swc 1–32 ($N = 6400$). Therefore, it seems that there are more beads in each window γ of swc 1–79 than that of swc 33–79 and swc 1–32. As shown in Fig 3, at $1.0 T_0$, LLPS is obvious in the system swc 1–79 and swc 33–79, but not obvious in the system 1–32. However, in order to keep the same total density, the three systems have the same length of z axis but different lengths of x and y axes. As a result, the distribution of density on z axis (density in each window: $D_\gamma = \frac{m_\gamma}{V_\gamma}$, V_γ is the volume of each window) has the same trends with the distribution of P_γ among these systems.

The LLPS will disappear at high temperature (critical temperature, T_{cr}) when the density difference is not found in the protein solution. The T_{cr} can be obtained by fitting to $\rho_H - \rho_L = A(T_{cr} - T)^\beta$ [30, 31] or the Flory-Huggins theory [23, 24, 43]. Here we use $P_H - P_L$ to roughly estimate T_{cr} because of the different sequence length of the three swc chains. When $P_H - P_L = 0$, the temperature is equal to T_{cr} . However, we can not obtain identical P_H and P_L at high temperature in the simulations because of the limits of the sampling. If we set the threshold of LLPS to be 0.07 (LLPS disappears when $P_H - P_L < 0.07$), the T_{cr} values of swc 1–79 and swc 33–79 are about $3.5 T_0$ and $3.0 T_0$, respectively (see Fig 2B). The salt concentration of the solvent has a significant impact on the T_{cr} of swc 1–32. The charge interactions will decrease the T_{cr} of swc 1–32 for about $0.5 T_0$ at 10 mM and swc 1–79 for about $0.2 T_0$ at 10 mM, and enhance the T_{cr} a bit of swc 33–79. The changes of T_{cr} on charge interactions seem to have something to do

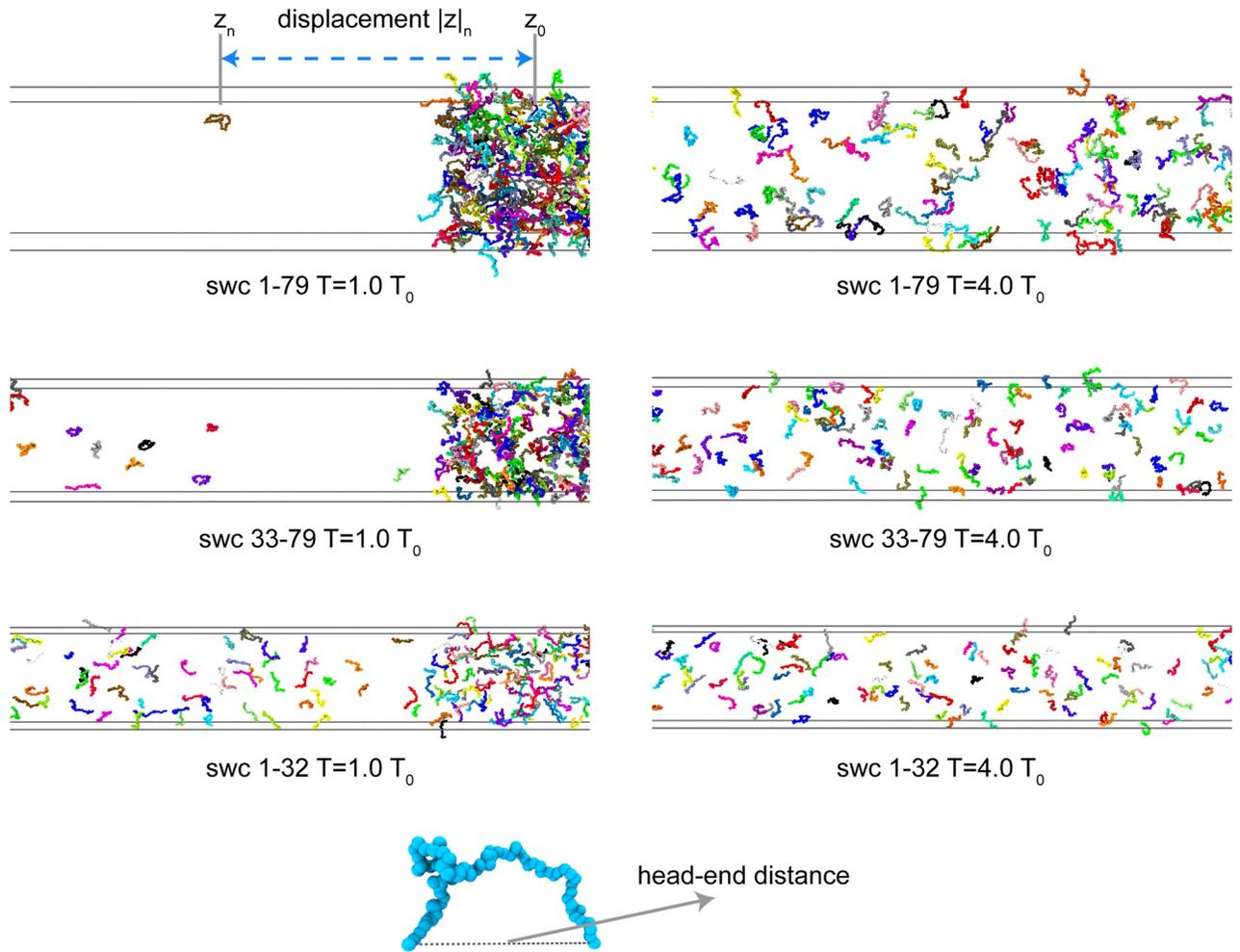


Fig 3. The last frame of the simulations of the swc systems at 1.0 and 4.0 T_0 . Here we illustrate part of the long z axis. The head–end distance (D) and the displacement $|z|$ are labeled in this figure. Only the simulations with 10 mM salt concentration are shown in this figure.

<https://doi.org/10.1371/journal.pcbi.1008672.g003>

with the sequence charge decoration (SCD) value ($SCD = \frac{1}{N} \sum_{i=1}^N \sum_{j=i+1}^N \sigma_i \sigma_j \sqrt{j-i}$, see Ref [44]).

SCD with large absolute value means the sequence with charge-blocky pattern (like sv15 in Ref [44]); SCD with small absolute value means the sequence with charge-scrambled pattern (like sv1 in Ref [44]). The SCD values of swc 1–32, swc 33–79, and swc 1–79 are 7.77, -0.32, and 7.40. The SCD can be considered as 0.00 when there are no charge interactions (neutral polymers). Recently, Das *et al.* discussed the influence of charge distribution on the behaviors and phase diagrams of two sequences sv1 and sv15 (with the same length and total charge number) [26]. We have performed the simulations using the sv1 and sv15 sequences and the obtained same result with Das *et al.* that the T_{cr} of sv15 is higher than that of sv1 (see the Discussion section). However, in our study the systems have different lengths, total charge numbers, and numbers of the charged residues. Then we compare swc 1–79 with swc 1–32, which have similar SCD values. The results suggest that the charge interactions have a more significant effect on the sequences with shorter length and fewer numbers of the charged residues.

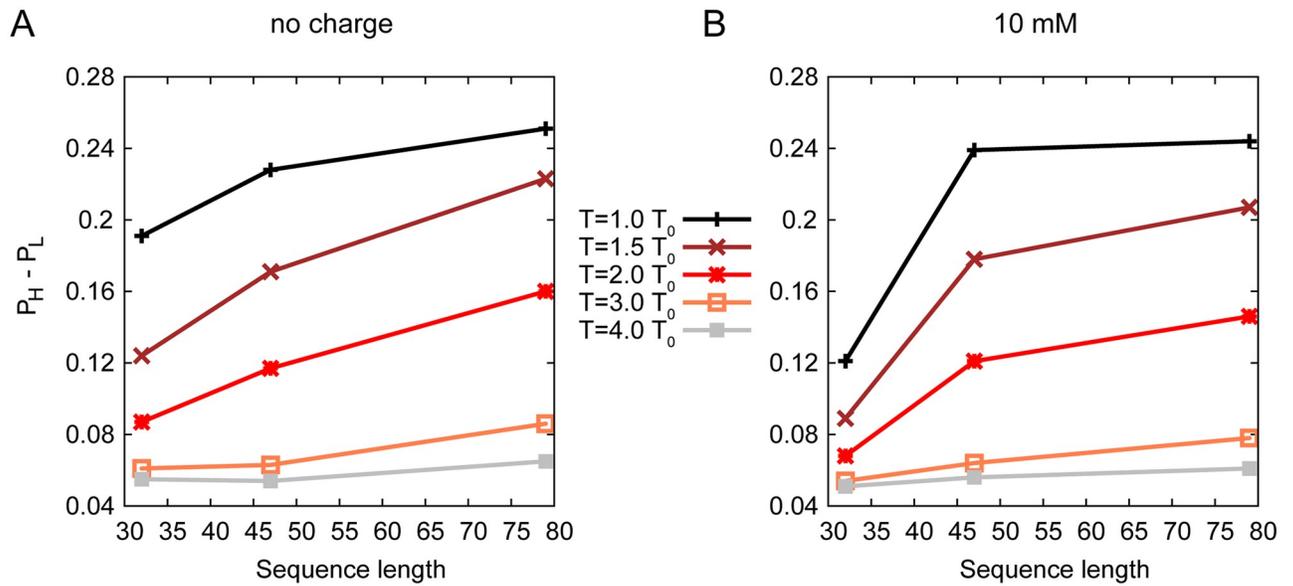


Fig 4. $P_H - P_L$ as a function of swc sequence length. 32 on x axis corresponds to swc 1–32, 47 corresponds to swc 33–79, 79 corresponds to swc 1–79.

<https://doi.org/10.1371/journal.pcbi.1008672.g004>

Moreover, here swc without charge interactions can be considered as neutral polymers, the behaviors of which depend mainly on the length of the polymer chain. Therefore we present a figure to show the influence of the sequence length on the $P_H - P_L$. As shown in Fig 4A, without charge interactions, the extent of LLPS (measured by $P_H - P_L$) and the length of swc chain are linearly correlated. Long polymers prefer to show the LLPS at the same temperature. However, the charge interactions decrease the $P_H - P_L$ of swc 1–32 and 1–79, but increase the $P_H - P_L$ of swc 33–79 (see Fig 4B). At temperature $1.0 T_0$, two sequences with different lengths (swc 1–79 and swc 33–79) have similar extent of LLPS under the influence of the charge interactions at 10 mM. Consequently, the distribution of the charged residues will affect the extent of LLPS.

The droplet is dynamical even at low temperatures

In order to analyze the behavior of the individual chain during the simulation, we calculated the displacement $|z|$ of each chain on the z axis ($|z|_n = |z_n - z_0|$, $n = 1..200$, see Fig 3), which can reveal the movement of each chain relative to the center of the system (condensed-phase with LLPS) during the simulation. As shown in Supporting information S8–S10 Figs, at low temperatures, the composition of the condensed-phase or the dilute-phase is not invariant after equilibration, in spite of the large and stable $P_H - P_L$. Some swc chains move from the condensed-phase to the dilute-phase, or from the dilute-phase to the condensed-phase. This is consistent with the observation of the dynamical LLPS in the experiments [6, 43]. This kind of movement changes to be frequent in the simulation time at high temperatures, suggesting the lost of the LLPS behavior.

We then selected the $|z|$ as the reaction coordinate and cut it into 30 windows and calculated the fraction of the number of the chains that belong to each window, as well as the standard error of the $|z|$ values in each window. As shown in the Supporting information S11 Fig, high value of the probability of low $|z|$ corresponds to the condensed-phase, which is not obvious at high temperature. Intriguingly, at $1.0 T_0$, the region of the condensed-phase of swc 33–79 is larger than that of swc 1–79 and swc 1–32. In addition, we used the standard error of the $|z|$ values in each window to show the fluctuation. The fluctuation of the chains in dilute-phase

is much higher than those in the condensed-phase (see Supporting information S11 Fig). Moreover, at $1.0 T_0$, the fluctuation of the condensed-phase of swc 1–79 is a bit lower than that of swc 33–79 and swc 1–32, which means that the condensed-phase of swc 1–79 is more stable than that of the other two sequences.

Temperature, sequence length, and charge pattern play different roles on the conformation of swc

For each swc chain, we record its head–end distance (see Fig 3, denoted as D) to show the conformational changes in the system during the simulation time. As shown in Fig 5A and 5B, for neutral polymers (without charge interactions), the head–end distance (D) and the length of swc chain (L) are linearly correlated. Compared with the sequence length L , the temperature influences the D value slightly. Swc chains tend to be a bit more expanded at lower temperatures when there are no charge interactions. The strange behavior observed for the swc ensemble is different from that of the free IDPs which unfold (expanded) at high temperatures. We will explain it in detail in the next section. In addition, the charge interactions of swc 33–79 and sv15 can significantly decrease the D value (see Fig 5A and 5C, Supporting information S2 Table). The contribution of the electrostatic attractions is higher than that of the repulsions in the swc 33–79 and sv15 system, making it to be more “folded/collapsed” like at lower temperature. The charge interactions of swc 1–32 and 1–79 have the opposite effect on the D value. Especially for swc 1–32, the repulsions straighten the protein significantly at low temperatures. Swc 1–32 have 16 negative-charged and 2 positive-charged residues. As illustrated in Fig 1, the tandem DEF/Y-1 and DEF/Y-2 motifs compose of the continuous negative-charged residues, which will enhance the repulsions significantly at low salt concentrations.

The ensemble of chains suggest the three forms of swc

Aiming to find out the difference between the chains in the condensed-phase and that in the dilute-phase, we calculated the distributions of the swc head–end distance (referred as D) along the displacement $|z|$. In the meantime, the standard error of D was collected for each window along the $|z|$, denoted as σ (see Fig 6). The large and small D values correspond to the extended and collapsed structure of IDPs. In addition, the σ value shows the flexibility of IDP structure.

As shown in Fig 6, at low temperatures with LLPS, the swc chains are more flexible in the dilute-phase (high $|z|$) than in the condensed-phase (low $|z|$). For example, at $1.0 T_0$, all the σ values of swc 1–79 chains outside the condensed-phase ($|z| > 200 \text{ \AA}$) are higher than 0.9 \AA (with and without charge interactions), while almost all the σ values of swc 1–79 chains inside the condensed-phase ($|z| < 200 \text{ \AA}$) are lower than 0.1 \AA (with and without charge interactions, see Fig 7). At high temperatures without LLPS ($4.0 T_0$), all swc 1–79 chains have similar but slightly higher σ value (about 0.2 \AA) as those in the condensed phase at $1.0 T_0$ (with and without charge interactions). In addition, the distribution of the D values of swc 1–79 is more concentrated at $4.0 T_0$ than those at lower temperatures. And the mean D value of swc 1–79 at $4.0 T_0$ is lower than those at lower temperatures (see Fig 5B and 5C). The swc 33–79 and swc 1–32 have similar trends of the σ values along the $|z|$. These results suggest that the ensemble of swc chains exhibit different conformations with and without LLPS. When LLPS occurs, the lower σ values in the condensed-phase mean that swc chains are tightly packed, and the higher σ values in the dilute-phase mean that swc chains flexibly moving. At high temperatures without LLPS, swc chains are fully spread/distributed in the simulation box with relatively stable D and small σ values, which suggests that the high temperatures can decrease the flexibility of swc chains compared with those in the dilute-phase at low temperatures. As a result, we can classify

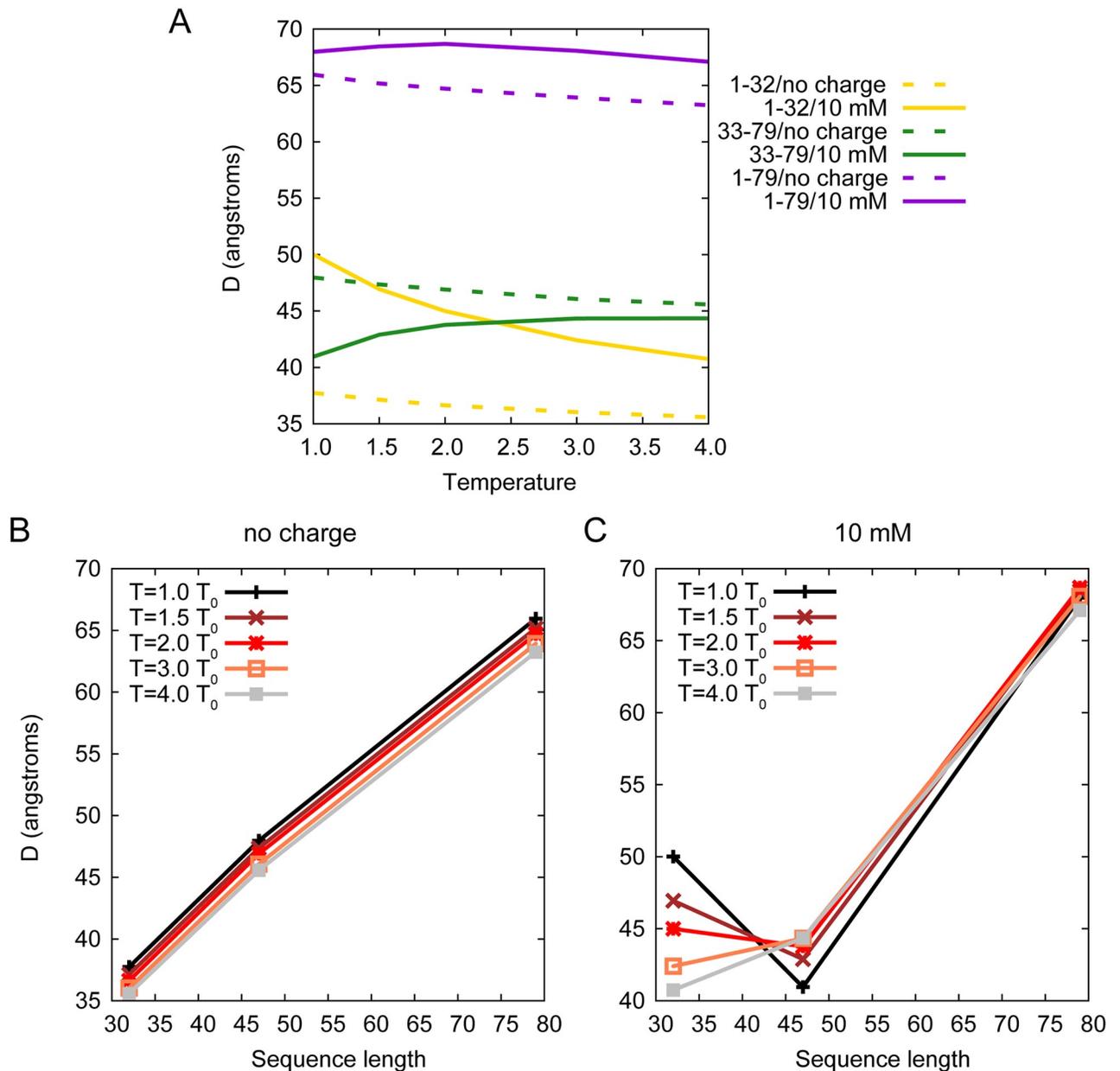


Fig 5. Head-end distance (D) distributions. (A) Head-end distance (D) of swc chains as a function of temperature. (B–C) Head-end distance (D) of swc chains as a function of sequence length during simulations without charges (B) and at 10 mM (C). In panel B and C, 32 on x axis corresponds to swc 1–32, 47 corresponds to swc 33–79, 79 corresponds to swc 1–79. The D value is calculated as the mean value of the 200 chains in the system during the last 1000 τ simulation data.

<https://doi.org/10.1371/journal.pcbi.1008672.g005>

the chains into three main forms (see Fig 7): packed form (P) in the condensed-phase with low σ value (low temperature with LLPS); dispersed form (D) in the dilute-phase with high σ value (low temperature with LLPS); fully-spread form (F) with low σ value (high temperature without obvious LLPS). Moreover, these three forms of swc chains can be distinguished on the phase diagram of D and σ (see the left panels of Fig 7). At low temperatures, swc chains behave in the dilute-phase as free IDPs/IDRs because there are almost no inter-chain interactions and the D value is flexible (high σ). The D value is stable in the condensed-phase because of the

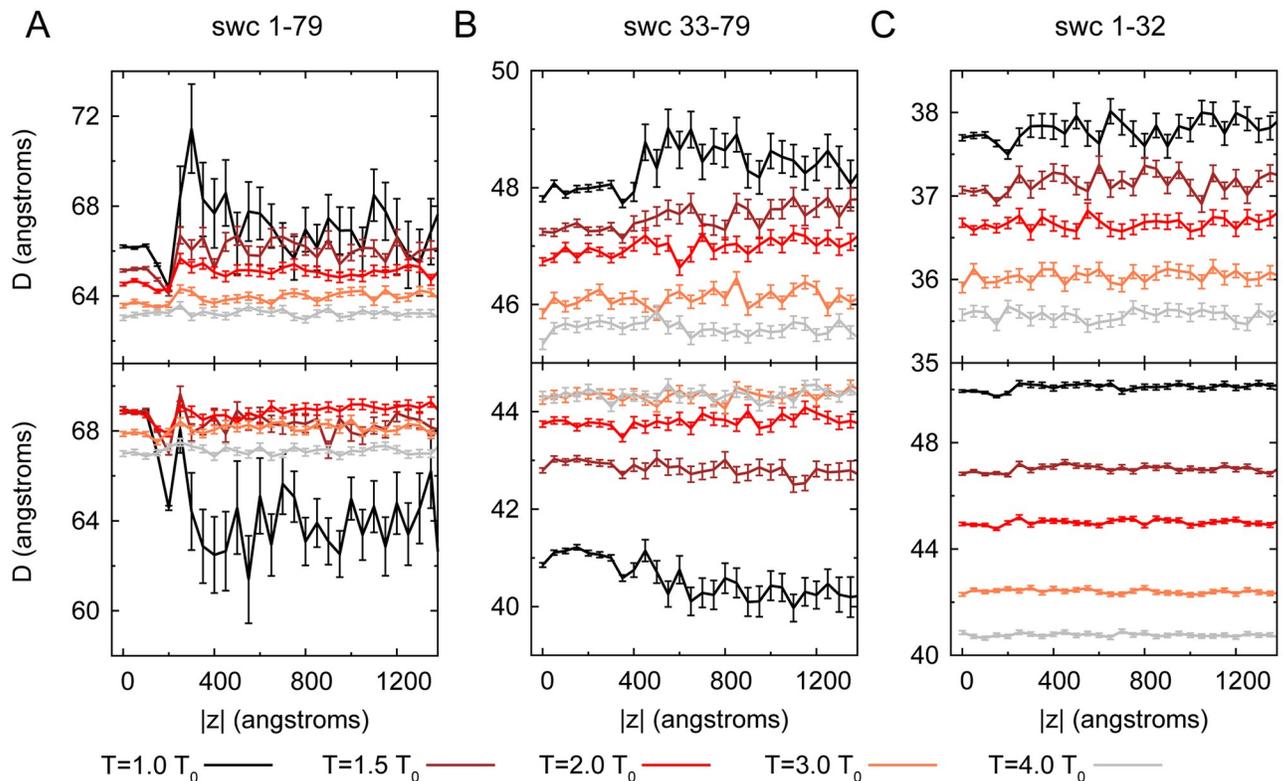


Fig 6. Head-end distance (D) of swc 1–79 (A), swc 33–79 (B), and swc 1–32 (C) as a function of displacement z . Data without charges and at 10 mM are shown in the up panels and the bottom panels, respectively. Mean D and standard error (σ) values during the last 1000 τ simulation data are illustrated in this figure. Considering the effect of boundary, the data with displacement z higher than 1400 are not calculated.

<https://doi.org/10.1371/journal.pcbi.1008672.g006>

packing of the chains. P form (ordered behavior) and D form (disordered behavior) of swc chains can coexist at low temperatures. As the increase of the temperature, the boundary between the condensed-phase and the dilute-phase disappears (no LLPS). The simulation box is fully filled with swc chains. However, the thermodynamical motions are enhanced at high temperatures. The swc chains have to be rolled up a bit to avoid colliding with each other, leading to a lower D value at high temperatures.

The three forms of swc chains (P, D, and F) can also be recognized in the distribution of E_{inter} and E_{intra} . As shown in Supporting information S12 and S13 Figs, at the low temperatures when LLPS occurs, chains in the packed form (P) have much higher E_{inter} and lower E_{intra} than those in the dispersed form (D). In addition, the standard error values of E_{intra} in D form (higher than 0.6 for swc 1–79 at $1.0 T_0$) are significantly higher than those in the P form (lower than 0.07 for swc 1–79 at $1.0 T_0$), which indicates that swc chains behave as free IDPs/IDRs in the dilute-phase. At high temperatures without LLPS, swc chains behave as the fully-spread form (F) with a bit higher E_{inter} (about 0.3 for swc 1–79 at $4.0 T_0$) than those in D form at low temperatures (close to 0 for swc 1–79 at $1.0 T_0$). In addition, the E_{intra} values of F form at high temperatures (ranging from 12.4 to 12.8 for swc 1–79 at $4.0 T_0$) show large differences with that of D form at low temperatures (ranging from 14.2 to 17.5 for swc 1–79 at $1.0 T_0$). The results suggest that at high temperature without LLPS, the chains in F form do not behave as free IDPs/IDRs.

The charge interactions have different impacts on the three swc segments. In the condensed-phase (packed form) at low temperatures, swc 1–79 and swc 1–32 are more expanded

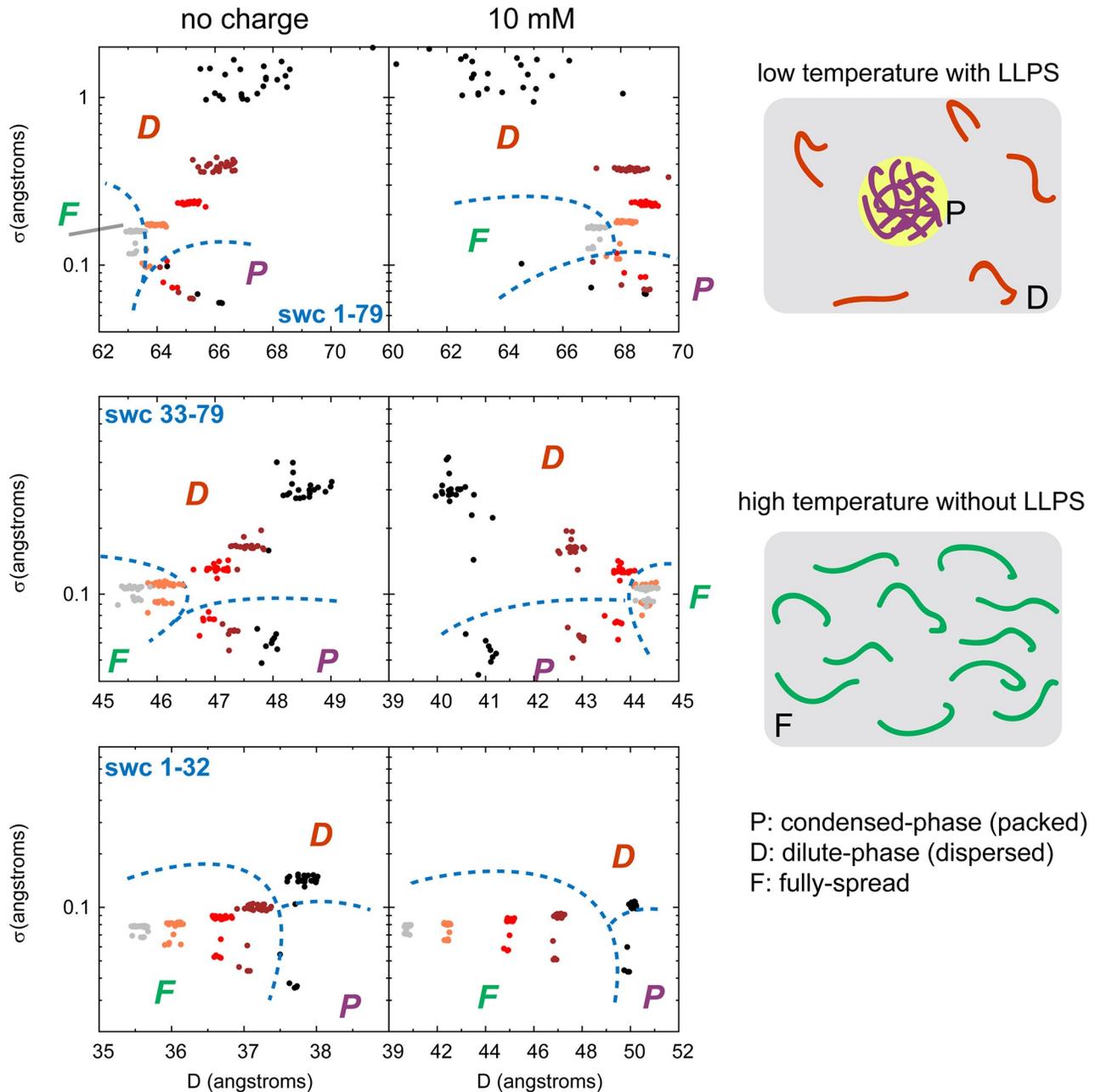


Fig 7. Predicted phase diagrams. Phase diagrams of each system with the distribution of D and σ (left), as well as the schematic diagrams for three forms of swc chains (right). The data in the left panels are extracted from the Fig 6, with the same coloring method. Three forms of swc chains (P, D, and F) are labeled in this figure. P represents the packed form of IDP in the condensed-phase (droplet) at low temperatures; D represents the dispersed form of IDP in the dilute-phase (outside the droplet) at low temperatures; F represents the fully-spread form of IDP at high temperatures without LLPS.

<https://doi.org/10.1371/journal.pcbi.1008672.g007>

(larger D) at 10 mM than that without charge interactions, while swc 33–79 is more folded (smaller D) at 10 mM (see Fig 6). The results suggest that the more expanded conformation at 10 mM does not favor the stable packing in the condensed-phase because the T_{cr} values of both swc 1–79 and swc 1–32 are decreased at 10 mM. In addition, the large number of continuous negative-charged residues in swc 1–79 and swc 1–32 (the DEF/Y motifs) can form high

inter-chain repulsive interactions. The high strength of electrostatic interactions (at low salt concentration) may not be helpful for the LLPS.

Intriguingly, in some systems with high extent of LLPS (e.g. swc 1–79 and 33–79 at $1.0 T_0$), it is obvious that the D value decreases slightly in the dilute-phase compared to that in the condensed-phase at 10 mM salt concentration, but increases slightly without charge interactions when LLPS occurred (see Fig 6). At $1.0 T_0$ without the charge interactions, the mean D of swc 1–79 is 66.00 Å in the condensed-phase ($|z| < 200$ Å) and 67.25 Å in the dilute-phase ($|z| > 200$ Å). At 10 mM, the mean D of swc 1–79 is 68.37 Å in the condensed-phase ($|z| < 200$ Å) and 64.03 Å in the dilute-phase ($|z| > 200$ Å). For swc 33–79, these values are 47.94 Å (condensed-phase, $|z| < 400$ Å) and 48.53 Å (dilute-phase, $|z| > 400$ Å) without charge interactions, 41.01 Å (condensed-phase, $|z| < 400$ Å) and 40.37 Å (dilute-phase, $|z| > 400$ Å) at 10 mM. As a result, the large amount of inter-chain interactions limit the proteins in the condensed-phase into a different form (P form), which is stable and different from the free IDP conformation (D form), keeping the shape of the droplet (MLO) stable. When LLPS is obvious, it seems that the vdW inter-chain interactions (no charges) will lead to a lower D in the condensed-phase than that in the dilute-phase. Combined vdW and charged inter-chain interactions (10 mM) will lead to a higher D in the condensed-phase than that in the dilute-phase. In addition, the fluctuation of the condensed-phase and the dilute-phase increases and decreases as the increase of the temperature. The difference between the condensed-phase and the dilute-phase becomes smaller until both of them transfer to the F form (LLPS disappears). And these behaviors above are found in the swc chains with different sequence length and charge pattern.

Discussion

In this study, we measure the phase diagram of the three swc chains and show the influence of electrostatic interactions on the LLPS of these swc chains. We find that without electrostatic interactions, the three swc chains all show LLPS at low temperature ($1.0 T_0$, see Fig 2). It is consistent with the theoretical conclusions of Dignon and Das *et al.* [27, 31]. The simulations of FUS did not show the ionic strength dependence [31]. Das *et al.* claimed that the general inter-residue LJ attraction favors LLPS but diminishes sequence specificity of LLPS [27]. We obtain similar conclusion because these proteins are neutral polymers with identical beads with no electrostatic interactions. The extent of LLPS is only relevant with the sequence length.

In addition, our simulation results suggest that the electrostatic interactions can increase the critical temperature T_{cr} of swc 33–79 a bit, but decrease the T_{cr} of swc 1–79 and 1–32 (Fig 2). Intriguingly, it is different from the simulation results of Das *et al.* that T_{cr} values of sv1 and sv15 all decrease when the electrostatic interactions are screened [27]. In that study, both sv1 and sv15 are ideal polyampholytes with equal number of positive-charged residues and negative-charged residues ($f_+ = f_- = 0.5$, f_+ and f_- are the fractions of positive-charged and negative-charged residues, respectively). In the present study, the conserved DEF/Y-1 and DEF/Y-2 motifs enhance the f_- values of swc 1–79 and swc 1–32 significantly (see Fig 1). Both swc 1–79 and 1–32 have high NCPR (net charge per residue, $|f_+ - f_-|$) that can be considered as polyelectrolytes according to the classification of Das and Pappu [35]. We then performed the simulation using the sv1 and sv15 sequences with the same settings. As shown in the Supporting information S14 Fig, the T_{cr} of sv1 seems barely changed when the electrostatic interactions are strong or screened. The T_{cr} of sv15 is about $3.5 T_0$ at 10 mM, and decreases to about $3.0 T_0$ when the electrostatic interactions are screened. And the electrostatic interactions influence the $P_H - P_L$ of sv15 significantly at low temperatures. We obtain the same result with Das *et al.* that the T_{cr} of sv15 is higher than that of sv1. The differences on the value and magnitude could result from the different simulation model and box. The SCD values of both sv1 and

sv15 are negative, -0.41 and -4.35. In addition, the SCD values of swc 1–32, swc 33–79, and swc 1–79 are 7.77, -0.32, and 7.40. Combined with the results of swc sequences, we can conclude that for the sequence with high negative SCD value (e.g. sv15 with -4.35), the electrostatic interactions will favor the LLPS. On the other hand, for the sequence with high positive SCD value (e.g. swc 1–32 with 7.77), the electrostatic interactions will hamper the LLPS. Moreover, though swc 33–79 and sv1 have similar sequence lengths and SCD values, the number and distribution of their charged residues are all different (sv1 and sv15 are fully charged polyampholytes with zero net charge), which may lead to different density distribution and LLPS behaviors.

Some experiments have demonstrated that the arginine-rich positively charged sequences (e.g., RGG boxes, RRM, SR repeats) drive the LLPS in an ionic strength dependent manner of Ddx4, LAF-1, and hnRNPA1 [2, 6, 7, 15, 45]. Besides, both methylation and phosphorylation can change the phase separation threshold for certain proteins such as Ddx4, tau441, MEG-3/MEG-4 [2, 45, 46]. Therefore, the LLPS behavior in different solutions/environments will be more complicated *in vivo* or in the cell.

We have summarized the behavior of swc ensemble under the influence of temperature and found a different phenomenon from the conventional free IDPs/IDRs. Swc chains tend to be a bit more expanded at lower temperatures when there are no charge interactions. In the swc ensemble, the chains behave as two forms coexisted with different D and σ values at low temperatures, and have to be rolled up a bit to avoid colliding with each other at high temperatures, leading to a lower D value compared with the extended structure. In the previous studies [19, 47, 48], the discovery of IDPs has extended the traditional sequence-structure relationship and led to the protein trinity model, which suggests that the native protein structure includes not only the ordered state (solid-like), but also molten globules (MG, liquid-like) and the random coils (pre-MG, gas-like). Free IDPs are found to behave like molten globules (collapsed-disordered) or random coils (extended-disordered). However, the trinity property is suitable for the single IDP or the IDPs without inter-chain interactions between each other (free IDPs). Here in this study, we discussed the structure and order of the whole ensemble of swc chains. Three different forms of swc chains can be distinguished on the phase diagrams of D and σ values (see Fig 7): packed form (P) in the condensed-phase with low σ value (low temperature with LLPS); dispersed form (D) in the dilute-phase with high σ value (low temperature with LLPS); fully-spread form (F) with low σ value (high temperature without obvious LLPS). The swc conformations can be linked with the temperature and the ensemble. P and D forms can coexist at low temperatures. They disappear into another form (F) at high temperatures. In addition, some characteristics, such as the D and σ values inside and outside the droplet, are found in the swc chains with different sequence lengths and charge patterns. It suggests that similar behaviors may be observed in other IDPs/IDRs. Different structures can be linked to different protein functions. Likewise, different forms of IDP/IDR ensemble may facilitate the essential processes in cell.

There are significant investigations that have discussed the biological functions of the LLPS. Huang *et al.* predicted that the IDRs of Swc5 may be involved in the LLPS with the similar mode as the carboxy-terminal domain (CTD) of RNA polymerase (Pol) II [33]. For Pol II, the CTD (IDR) can undergo LLPS *in vitro*, which is based mainly on the weak hydrophobic interactions and influenced by the CTD length and the CTD phosphorylation [49]. Boehning *et al.* [49] proposed that the CTD can recruit the core of Pol II and the transcription factors on the condensed-phase (droplet), leading to the Pol II hub, which may provide a reservoir of Pol II, enabling high initiation rates during the activated transcription. Moreover, as a histone chaperon, the Swc5 binds with H2A-H2B tightly with its IDR [33]. The DEF/Y motifs are essential for the binding. Therefore, the IDR of Swc5, especially the DEF/Y motifs, may recruit or retain

histones within phase-separated liquid droplets where transcription factors and chromatin remodeling factors coalesce to regulate transcription [33]. Our investigations show the LLPS of the Swc5 IDR, which may facilitate the recruitment, deposit, and assembly of the histones. In addition, our results agree with the properties of the CTD about the LLPS, suggesting the similar assembly mode and biological function of the IDR of Swc5.

Supporting Information

S1 Fig. Parameters calibration. The mean head–end distance (D) with different parameters σ and ε (left panel: $\sigma = 1.0$ nm, $\varepsilon = 0.01, 0.1, 0.8$ kJ/mol; right panel: $\sigma = 0.6, 0.8, 1.0$ nm, $\varepsilon = 0.01$ kJ/mol).

(TIF)

S2 Fig. The distribution of the peak1 of swc 1–79. Peak1 is the r value corresponding to the peak value of $g(r)$ curve between chain 1 and another chain (199 peak1 values in total).

(TIF)

S3 Fig. The distribution of the peak1 of swc 33–79. Peak1 is the r value corresponding to the peak value of $g(r)$ curve between chain 1 and another chain (199 peak1 values in total).

(TIF)

S4 Fig. The distribution of the peak1 of swc 1–32. Peak1 is the r value corresponding to the peak value of $g(r)$ curve between chain 1 and another chain (199 peak1 values in total).

(TIF)

S5 Fig. $P_H - P_L$ of swc 1–32 as a function of simulation time. The upper panel shows the simulations without charge interactions (no charge); the bottom panel shows the simulations at 10 mM salt concentration.

(TIF)

S6 Fig. $P_H - P_L$ of swc 33–79 as a function of simulation time. The upper panel shows the simulations without charge interactions (no charge); the bottom panel shows the simulations at 10 mM salt concentration.

(TIF)

S7 Fig. $P_H - P_L$ of swc 1–79 as a function of simulation time. The upper panel shows the simulations without charge interactions (no charge); the bottom panel shows the simulations at 10 mM salt concentration.

(TIF)

S8 Fig. Swc chain movements. The displacement of one swc chain to the CM of all 200 swc 1–79 chains on the z axis (displacement z) during the simulation. Considering the periodic simulation box, the displacement is the minimum of the distance values between this representative chain and any of the system CM mirrors.

(TIF)

S9 Fig. Swc chain movements. The displacement of one swc chain to the CM of all 200 swc 33–79 chains on the z axis (displacement z) during the simulation. Considering the periodic simulation box, the displacement is the minimum of the distance values between this representative chain and any of the system CM mirrors.

(TIF)

S10 Fig. Swc chain movements. The displacement of one swc chain to the CM of all 200 swc 1–32 chains on the z axis (displacement z) during the simulation. Considering the periodic

simulation box, the displacement is the minimum of the distance values between this representative chain and any of the system CM mirrors.

(TIF)

S11 Fig. The distribution of the probability (A) and the standard error (B) as a function of displacement $|z|$. The coloring method is the same as that in Fig 2.

(TIF)

S12 Fig. Mean number of inter-chain (E_{inter}) and intra-chain (E_{intra}) electrostatic attractions of swc 1–79 (A), swc 33–79 (B), and swc 1–32 (C) as a function of $|z|$. Average and standard error values during the last 1000 τ simulation data are illustrated in this figure.

(TIF)

S13 Fig. Phase diagrams of each system with the distribution of E_{inter} and E_{intra} . The data points are extracted from the figure above, with the same coloring method. Three forms of swc chains (P, D, and F) are labeled in this figure. P represents the packed form of IDP in the condensed-phase (droplet) at low temperatures; D represents the dispersed form of IDP in the dilute-phase (outside the droplet) at low temperatures; F represents the fully-spread form of IDP at high temperatures without LLPS.

(TIF)

S14 Fig. The appearance and disappearance of phase separation. (A) Phase diagram of sv1 and sv15 in different solvents. (B) $P_H - P_L$ changes with temperature and solvent. Here P_H and P_L are the highest and the lowest points of sv residue distribution along the z axis.

(TIF)

S1 Table. The $P_H - P_L$ data of different sequences at different condition (temperature 1.0, 1.5, 2.0, 3.0, 4.0 T_0 ; 10 mM salt concentration (10) or without charge interactions (no)).

These data are the mean values calculated with the last 1000 τ trajectory.

(DOCX)

S2 Table. The head–end data of different sequences at different condition (temperature 1.0, 1.5, 2.0, 3.0, 4.0 T_0 ; at 10 mM salt concentration (10) or without charge interactions (no)). These data are the mean values calculated with the last 1000 τ trajectory.

(DOCX)

Acknowledgments

W.-T.C. thanks Network and Computing Center, Changchun Institute of Applied Chemistry, Chinese Academy of Science and Computing Center of Jilin Province for computational support.

Author Contributions

Conceptualization: Jin Wang.

Data curation: Wen-Ting Chu, Jin Wang.

Formal analysis: Wen-Ting Chu, Jin Wang.

Funding acquisition: Wen-Ting Chu, Jin Wang.

Investigation: Wen-Ting Chu, Jin Wang.

Methodology: Wen-Ting Chu, Jin Wang.

Project administration: Wen-Ting Chu, Jin Wang.

Resources: Wen-Ting Chu, Jin Wang.

Software: Wen-Ting Chu, Jin Wang.

Supervision: Jin Wang.

Validation: Wen-Ting Chu, Jin Wang.

Visualization: Wen-Ting Chu.

Writing – original draft: Wen-Ting Chu, Jin Wang.

Writing – review & editing: Wen-Ting Chu, Jin Wang.

References

1. Sleeman JE, Trinklemulcahy L. Nuclear bodies: new insights into assembly/dynamics and disease relevance. *Current Opinion in Cell Biology*. 2014; 28:76–83. <https://doi.org/10.1016/j.ceb.2014.03.004> PMID: 24704702
2. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*. 2015; 57(5):936–947. <https://doi.org/10.1016/j.molcel.2015.01.013> PMID: 25747659
3. Marnik EA, Updike DL. Membraneless organelles: P granules in *Caenorhabditis elegans*. *Traffic*. 2019; 20(6):373–379. <https://doi.org/10.1111/tra.12644> PMID: 30924287
4. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, et al. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*. 2009; 324(5935):1729–1732. <https://doi.org/10.1126/science.1172046> PMID: 19460965
5. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, et al. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*. 2016; 165(7):1686–1697. <https://doi.org/10.1016/j.cell.2016.04.047> PMID: 27212236
6. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*. 2015; 163(1):123–133. <https://doi.org/10.1016/j.cell.2015.09.015> PMID: 26406374
7. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*. 2016; 14(1):1–20. <https://doi.org/10.1186/s12964-015-0125-7> PMID: 26727894
8. Li P, Banjade S, Cheng H, Kim S, Chen B, Guo L, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. 2012; 483(7389):336–340. <https://doi.org/10.1038/nature10879> PMID: 22398450
9. Hyman AA, Weber C, Julicher F. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology*. 2014; 30(1):39–58. <https://doi.org/10.1146/annurev-cellbio-100913-013325> PMID: 25288112
10. Brangwynne CP, Tompa P, V PR. Polymer physics of intracellular phase transitions. *Nature Physics*. 2015; 11:899–904. <https://doi.org/10.1038/nphys3532>
11. Mittag T, Parker R. Multiple Modes of Protein-Protein Interactions Promote RNP Granule Assembly. *Journal of Molecular Biology*. 2018; 430(23):4636–4649. <https://doi.org/10.1016/j.jmb.2018.08.005> PMID: 30099026
12. Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, et al. Phosphorylation of the FUS low complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO Journal*. 2017; 36(20):2951–2967. <https://doi.org/10.15252/emboj.201696394> PMID: 28790177
13. Li H, Chiang W, Chou P, Wang W, Huang J. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *Journal of Biological Chemistry*. 2018; 293(16):6090–6098. <https://doi.org/10.1074/jbc.AC117.001037> PMID: 29511089
14. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*. 2013; 495(7442):467–473. <https://doi.org/10.1038/nature11922> PMID: 23455423
15. Elbaumgarfinkle S, Kim Y, Szczepaniak KJ, Chen CC, Eckmann CR, Myong S, et al. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics.

- Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(23):7189–7194. <https://doi.org/10.1073/pnas.1504822112>
16. Zhou H, Nguemaha V, Mazarakos KP, Qin S. Why Do Disordered and Structured Proteins Behave Differently in Phase Separation. *Trends in Biochemical Sciences*. 2018; 43(7):499–516. <https://doi.org/10.1016/j.tibs.2018.03.007> PMID: 29716768
 17. Patel A, Lee HO, Jawerth L, Maharana S, Jahnelt M, Hein MY, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. 2015; 162(5):1066–1077. <https://doi.org/10.1016/j.cell.2015.07.047> PMID: 26317470
 18. Milovanovic D, Wu Y, Bian X, De Camilli P. A liquid phase of synapsin and lipid vesicles. *Science*. 2018; 361(6402):604–607. <https://doi.org/10.1126/science.aat5671> PMID: 29976799
 19. Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E, et al. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(40):E3743–E3752. <https://doi.org/10.1073/pnas.1308381110> PMID: 24043820
 20. Chu WT, Clarke J, Shammas SL, Wang J. Role of non-native electrostatic interactions in the coupled folding and binding of PUMA with Mcl-1. *PLoS computational biology*. 2017; 13(4):e1005468. <https://doi.org/10.1371/journal.pcbi.1005468> PMID: 28369057
 21. Chu WT, Chu X, Wang J. Investigations of the underlying mechanisms of HIF-1 α and CITED2 binding to TAZ1. *Proceedings of the National Academy of Sciences*. 2020; 117(11):5595–5603. <https://doi.org/10.1073/pnas.1915333117> PMID: 32123067
 22. Chu WT, Shammas SL, Wang J. Charge Interactions Modulate the Encounter Complex Ensemble of Two Differently Charged Disordered Protein Partners of KIX. *Journal of Chemical Theory and Computation*. 2020; 16(6):3856–3868. <https://doi.org/10.1021/acs.jctc.9b01264> PMID: 32325001
 23. Flory PJ. Thermodynamics of High Polymer Solutions. *Journal of Chemical Physics*. 1941; 10(1):51–61. <https://doi.org/10.1063/1.1723621>
 24. Huggins ML. Some Properties of Solutions of Long-chain Compounds. *The Journal of Physical Chemistry*. 1942; 46(1):151–158. <https://doi.org/10.1021/j150415a018>
 25. Lin Y, Brady JP, Formankay JD, Chan HS. Charge pattern matching as a ‘fuzzy’ mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New Journal of Physics*. 2017; 19(11):115003. <https://doi.org/10.1088/1367-2630/aa9369>
 26. Das S, Eisen A, Lin Y, Chan HS. A Lattice Model of Charge-Pattern-Dependent Polyampholyte Phase Separation. *Journal of Physical Chemistry B*. 2018; 122(21):5418. <https://doi.org/10.1021/acs.jpcc.7b11723>
 27. Das S, Amin AN, Lin Y, Chan HS. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *Physical Chemistry Chemical Physics*. 2018; 20(45):28558–28574. <https://doi.org/10.1039/C8CP05095C> PMID: 30397688
 28. Azia A, Levy Y. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *Journal of molecular biology*. 2009; 393(2):527–542. <https://doi.org/10.1016/j.jmb.2009.08.010> PMID: 19683007
 29. Givaty O, Levy Y. Protein sliding along DNA: dynamics and structural characterization. *Journal of molecular biology*. 2009; 385(4):1087–1097. <https://doi.org/10.1016/j.jmb.2008.11.016> PMID: 19059266
 30. Dignon GL, Zheng W, Best RB, Kim YC, Mittal J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(40):9929–9934. <https://doi.org/10.1073/pnas.1804177115> PMID: 30217894
 31. Dignon GL, Zheng W, Kim YC, Best RB, Mittal J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Computational Biology*. 2018; 14(1):e1005941. <https://doi.org/10.1371/journal.pcbi.1005941> PMID: 29364893
 32. Sun L, Luk E. Dual function of Swc5 in SWR remodeling ATPase activation and histone H2A eviction. *Nucleic Acids Research*. 2017; 45(17):9931–9946. <https://doi.org/10.1093/nar/gkx589> PMID: 28973436
 33. Huang Y, Sun L, Pierrakeas L, Dai L, Pan L, Luk E, et al. Role of a DEF/Y motif in histone H2A-H2B recognition and nucleosome editing. *Proceedings of the National Academy of Sciences of the United States of America*. 2020; 117(7):3543–3550. <https://doi.org/10.1073/pnas.1914313117> PMID: 32001508
 34. Tramantano M, Sun L, Au C, Labuz D, Liu Z, Chou M, et al. Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *eLife*. 2016; 5:e14243. <https://doi.org/10.7554/eLife.14243> PMID: 27438412

35. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(33):13392–13397. <https://doi.org/10.1073/pnas.1304749110> PMID: 23901099
36. Chu X, Wang Y, Gan L, Bai Y, Han W, Wang E, et al. Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone Chz1 and histone H2A. Z-H2B. *PLoS Comput Biol*. 2012; 8(7):e1002608. <https://doi.org/10.1371/journal.pcbi.1002608> PMID: 22807669
37. Wang Y, Gan L, Wang E, Wang J. Exploring the dynamic functional landscape of adenylate kinase modulated by substrates. *Journal of Chemical Theory and Computation*. 2012; 9(1):84–95. <https://doi.org/10.1021/ct300720s> PMID: 26589012
38. Blas FJ, Macdowell LG, De Miguel E, Jackson G. Vapor-liquid interfacial properties of fully flexible Lennard-Jones chains. *Journal of Chemical Physics*. 2008; 129(14):144703. <https://doi.org/10.1063/1.2989115> PMID: 19045161
39. Silmore KS, Howard MP, Panagiotopoulos AZ. Vapour-liquid phase equilibrium and surface tension of fully flexible Lennard-Jones chains. *Molecular Physics*. 2017; 115(3):320–327. <https://doi.org/10.1080/00268976.2016.1262075>
40. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981; 52(12):7182–7190. <https://doi.org/10.1063/1.328693>
41. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*. 2008; 4(3):435–447. <https://doi.org/10.1021/ct700301q> PMID: 26620784
42. Jackson J, Nguyen K, Whitford P. Exploring the balance between folding and functional dynamics in proteins and RNA. *International Journal of Molecular Sciences*. 2015; 16(4):6868–6889. <https://doi.org/10.3390/ijms16046868> PMID: 25822873
43. Brady JP, Farber P, Sekhar A, Lin Y, Huang R, Bah A, et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114(39):E8194–E8203. <https://doi.org/10.1073/pnas.1706197114> PMID: 28894006
44. Sawle L, Ghosh K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *Journal of Chemical Physics*. 2015; 143(8):085101. <https://doi.org/10.1063/1.4929391> PMID: 26328871
45. Wegmann S, Eftekhazadeh B, Tepper K, Zoltowska KM, Bennett RE, Dujardin S, et al. Tau protein liquid–liquid phase separation can initiate tau aggregation. *The EMBO Journal*. 2018; 37(7):e98049. <https://doi.org/10.15252/embj.201798049> PMID: 29472250
46. Wang JT, Smith J, Chen B, Schmidt H, Rasoloson D, Paix A, et al. Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in *C. elegans*. *eLife*. 2014; 3:e04591. <https://doi.org/10.7554/eLife.04591> PMID: 25535836
47. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*. 2001; 19(1):26–59. [https://doi.org/10.1016/S1093-3263\(00\)00138-8](https://doi.org/10.1016/S1093-3263(00)00138-8) PMID: 11381529
48. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein science*. 2002; 11(4):739–756. PMID: 11910019
49. Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature structural & molecular biology*. 2018; 25(9):833–840. <https://doi.org/10.1038/s41594-018-0112-y> PMID: 30127355