Protein phase separation: physical models and phase-separation-mediated cancer signaling

Travis M. Zeigler\textsuperscript{a}, Michael C. Chung\textsuperscript{a}, Om Prakash Narayan\textsuperscript{a} and Juan Guan\textsuperscript{a,b}

\textsuperscript{a}Department of Physics, University of Florida, Gainesville, FL USA; \textsuperscript{b}Department of Anatomy and Cell Biology, University of Florida, Gainesville, FL USA

ABSTRACT
Phase separation is a concept well described in physics where a system spontaneously exhibits two or more distinct yet coexisting phases at equilibrium. This review describes several popular physical models that serve as a theoretical framework to understand protein phase separation in biological systems, a burgeoning area of research with many challenges left to be explored. The principles of statistical mechanics and thermodynamics that encompass phase separation are crucial to understanding the biophysical properties of biomolecular condensates. Representative systems of protein phase separation in several naturally occurring cancer fusion proteins and their implications in cancer mechanisms are discussed to highlight the underappreciated biophysical perspective on cancer. This insight into the driving force for protein condensate assembly may help to identify novel disease mechanisms and open opportunities for further innovative therapeutic strategies.

ARTICLE HISTORY
Received 23 March 2021
Accepted 25 May 2021

1. Introduction
Phase separation is a concept well described in physics where a system exhibits two or more distinct yet coexisting phases \cite{1}, \cite{2}. In biological
systems, it is being increasingly shown that phase separation plays a critical role in a variety of cellular processes. Through high-resolution cell imaging studies, it is shown that under physiologically relevant salt, pH, and temperature conditions there are micrometer-scale structures spontaneously assembled in the cells that are composed of proteins. In sharp contrast to the traditional subcellular compartment and organelles, these protein-based structures have been found to be membraneless. As knowledge is rapidly expanding in this area, it is increasingly recognized that many subcellular structures, such as Cajal bodies, nucleoli, P bodies, and stress granules, are formed through protein phase separation [3–6]. Not only is protein phase separation involved in normal physiological processes, but it also plays key roles in pathological development [7–9].

The concept of phase separation is familiar to any who have experimented with the mixing of oil and vinegar. If the mixture is left to settle, the oil and vinegar will separate from each other. This spontaneous demixing can be ascribed to the energetically favorable homotypic interactions between oil molecules [10]. Therefore, upon reaching equilibrium, you would see regions that vary between high and low concentrations of oil and vinegar molecules. Interestingly, similar behavior on a microscopic scale has been observed with proteins inside cells. Proteins can bind to each other to create assemblies with high local concentration inside cells called macromolecular condensates. Such protein-rich structures often show some typical features of liquid–liquid phase separation, including flow in the presence of a shear stress and the formation of liquid droplets [11]. It is important to note that such assemblies can exist in a continuum of biophysical states. Depending on the specific systems and how the phase separation is induced, the system may exhibit characteristics that are somewhere in between liquid-like, gel-like, and solid-like. It seems that the state of phase separation is important as there is evidence that solid-like condensates are linked to a variety of diseases [12–16]. Protein phase separation plays an important role in normal cell physiology. Various membraneless subcellular structures are known to be assembled through protein phase separation, such as nucleoli and P bodies involved in RNA synthesis and processing. In contrast, misregulated protein phase separation can lead to dire consequences, resulting in neurodegenerative diseases such as amyotrophic lateral sclerosis and frontotemporal dementia [7] by promoting the formation of prion-like protein aggregates and cancers by promoting uncontrolled cell proliferation and growth signaling. This review will first summarize the theoretical frameworks and physical models widely accepted by the field. We will then present intriguing cases where protein phase separation is involved in cancer research, with special emphasis on cancer mutations of fusion proteins.
2. Physical mechanisms for protein phase separation

It is natural to wonder what kind of process or combination of processes could lead to phase separation. After all, in everyday experience things tend to mix, like mixing salt and water in a cup. Once the salt and water solution is made, the salt will not spontaneously precipitate out of the solution, assuming the environment surrounding the solution does not change. How then could phase separation, the opposite of mixing, occur? The answer lies in the sign of the change in Gibbs free energy. The Gibbs free energy is the amount of useful work that a system can do at a constant temperature and pressure. This quantity determines if a particular process occurs spontaneously or not. The process is spontaneous if the change is negative, and it is not spontaneous if the change is positive. The change in the Gibbs free energy is related to the change in enthalpy and entropy of a system by the relation, \( \Delta G = \Delta H - T\Delta S \), where \( \Delta H \) is the change in the enthalpy of the system, \( T \) is the temperature, and \( \Delta S \) is the change in the entropy of the system. There are several models that can be used to calculate these important quantities, each with its own assumptions.

2.1 Flory-huggins model

The Flory-Huggins model, originally applied to polymer solutions, is a simple physical model that provides intuitive understanding into the spontaneous de-mixing characteristic of phase separation. This model incorporates ideas similar to the entropy of mixing of ideal gases, with the caveat that it relaxes the assumption that the interaction energies between different pairs of molecules are the same [17,18]. In this model, there are two kinds of molecules: solvent and solute polymers. The solvent molecules and the individual solute monomers are imagined to be at a specific site in a lattice configuration. The number of lattice sites in the solution is \( N = N_1 + rN_2 \), where \( N_1 \) is the number of solvent molecules, \( N_2 \) is the number of solute polymers, and \( r \) is the number of monomers that are present in each solute polymer. A representation of the lattice in the Flory-Huggins model can be seen in Figure 1. The sign of the change in the Gibbs free energy for this system of solute and solvent molecules determines if phase separation happens spontaneously. The important parameters include the strength of interactions between like molecules, (i.e., solvent-solvent interactions and solute–solute interactions), and unlike molecules, solvent–solute interactions. \( \chi \), is an important parameter that depends on the difference between the solvent-solute interaction and the average of the solvent–solvent and solute–solute interactions, \( \chi = \frac{z(E_{\text{unlike}} - E_{\text{like}})}{k_BT} \). In this equation, \( z \) is the number of nearest neighbors for a lattice site, \( E_{\text{unlike}} \) is the solvent–solute interaction energy, \( E_{\text{like}} \) is the average of solute–solute
and solvent–solvent interaction energies, $k_b$ is Boltzmann’s constant, and $T$ is the temperature. The enthalpy change is the change in energy of each solvent–solute interaction compared to the average of solute–solute and solvent–solvent interactions, $E_{\text{Unlike}} - E_{\text{Like}}$, multiplied by the number of interactions, $N_1\phi_2 z$, where $\phi_2 = rN_2/N$ is the volume fraction of the solute monomers. Therefore, $\Delta H = k_bTN_1\phi_2\chi$. The entropy term $\Delta S$ in $\Delta G$ can be found as $S = k_b\ln(\Omega)$, where $\Omega$ is the number of microstates of the system. The number of initial microstates for the center of mass for each solute polymer is $rN_2$, and the number of initial microstates for each solvent molecule is $N_1$. The number of final positions for both solvent molecules and the center of mass of solute polymers is the total number of sites in the lattice, $N$. Using these values, the change in entropy can be calculated as $\Delta S = -k_b[N_1\ln(\phi_1) + N_2\ln(\phi_2)]$, which is the same equation as the entropy of mixing for two ideal gases. Adding this enthalpy change to the standard result obtained in the entropy of mixing yields a formula for change in Gibbs free energy; $\Delta G = RT[n_1\ln\phi_1 + n_2\ln\phi_2 + n_1\phi_2\chi]$, where $R$ is the gas constant, $n_1$ and $n_2$ are the moles of the solvent and solute, respectively, and $\phi_1$ and $\phi_2$ are the volume fractions of the solvent and solute, respectively. To apply this model to protein phase separation, the proteins that phase separate can be treated as the solute polymers and the surrounding cytoplasmic materials as the solvent in the Flory-Huggins

![Figure 1](image-url)

**Figure 1** The Flory-Huggins model for protein phase separation. The blue circles represent the solvent particles, and the red circles represent the protein monomers that make up the protein polymers. The black lines indicate the bonds between protein monomers. For simplicity, each line indicates a single bond. Experimental data has been fit to this simplistic model, and it successfully describes the phenomenon of protein phase separation in various systems.
model. The amino acids that make up the proteins are the individual monomers in the model. One caveat of the Flory-Huggins model is the underlying assumption that there is only one kind of monomer that makes up the polymer. Proteins should be viewed as a co-polymer with twenty different monomer (i.e., twenty distinct amino acids) choices. There would

\[ \text{protein 1 concentration} \]

\[ \text{protein 2 concentration} \]

Figure 2 Multivalent interactions promote protein phase separation. a. The valency can be thought of as the number of interactions that a protein can have with other molecules. Example of two proteins both with a valency of 4. b. The phase boundary is shown with respect to the concentrations of the protein 1 and protein 2. The phase boundary can be shifted depending on the valencies of the proteins. Proteins with higher valencies phase separate at lower concentrations than proteins with lower valencies.

Figure 3 The nephrin–NCK–N-WASP system is an example of a phase separation resulting from multivalent interactions. This figure depicts the interactions formed by nephrin, NCK, and N-WASP. As phosphorylated nephrin is introduced into the system the effective valency is increased and phase separation occurs at lower concentrations. This system shows that multivalent interactions can promote protein phase separation.
be many more interaction energies to account for in a practical biological system. A model based on Flory-Huggins may be a starting point to understand the phase behavior in a given system. Preliminary information may be obtained to create a more sophisticated model that incorporates the intricate behavior of the system. Even though this model may be simplistic, it has been used to successfully fit data for different systems involving protein phase separation [3,17,19–21].

2.2 Multivalent interaction model

Valency, number of bonds or interactions a protein can form with other proteins, plays a central role in determining the phase separation behavior [22,23]. Schematic examples of the valencies of two proteins are shown in Figure 2a. Relevant to protein phase separation, proteins can have well-defined protein-protein interactions between structured domains or have interactions at intrinsically disordered regions (IDRs), where the protein does not have a defined structure and switches between different conformations. It has been suggested that higher valencies have been tied to higher potency of protein phase separation (i.e., the onset of phase separation at lower concentrations) [22,24,25]. In one case, Li et al. studied the ability for the coacervation of SRC homology 3 (SH3) domain and the proline-rich motif (PRM) ligand to phase separate depending on the valency. The number of repeats of SH3 and that of PRM are varied between one and five subunits, creating a well-defined valency for each protein ranging from one to five. The proteins were denoted as $\text{SH3}_m$ and $\text{PRM}_n$, where the $m$ and $n$ in the subscripts represent the number of repeats present in each protein. An equal number of $\text{SH3}_m$ and $\text{PRM}_n$ molecules with the same valency were mixed together. The authors found that phase separation occurred at lower concentrations of both SH3 and PRM when the valency of both were higher [22]. The effects of protein concentration and the valency of proteins on protein phase separation can be seen in Figure 2b. These proteins can also form phase-separated condensates in cells. Li et al. made the fusion proteins mCherry-SH3$_5$ and eGFP-PRM$_5$ and expressed these two proteins in HeLa cells. Cells that expressed both fusion proteins developed small puncta in the cytoplasm that contained both mCherry-SH3$_5$ and eGFP-PRM$_5$. In contrast, puncta did not form with cells that only expressed one of the fusion proteins or cells that had mCherry-SH3$_5$ and eGFP-PRM$_5$. This suggests that condensates were only able to form through multivalent interactions between two proteins with high valency [22].

A more sophisticated model system with biological relevance has also been demonstrated [22,26]. The schematics of the protein system are depicted in Figure 3. NCK has SH3 domains, which can bind to the six PRMs on N-WASP. The authors found that N-WASP and NCK formed
droplets when mixed due to multivalent interactions. A third protein nephrin has three tyrosine phosphorylation sites where the SH2 domains of NCK can bind and serve as an underlying scaffold for NCK to increase its valency. It was shown that adding a diphosphorylated nephrin (2pTyr) tail peptide lowers the phase separation concentration threshold for NCK and N-WASP. The phase separation concentration threshold was lowered even more when a 3pTyr peptide (i.e., three potential valencies) was added [22]. These results show that valency is an important factor in determining if and at what concentration phase separation occurs. It was plausible for the cell to regulate the phase transition by tuning the amount of phosphorylation of nephrin in signaling pathways involving these three proteins [22].

Next, a theoretical framework for how the valencies of proteins lead to phase separation will be developed. As previously discussed, the number of bonds in a protein polymer is important in determining its ability to phase separate at certain concentrations. Therefore, to get a more accurate picture of protein phase separation, the contribution of valency should be included. The Flory-Huggins model did include interactions with multiple monomers in the formation of the protein polymer, but it does not consider the different ways of forming such a polymer. A physical model including valency has been developed by Li et al [22]. In this model, there exists A molecules with n interaction sites (denoted as $\alpha$ sites) and B molecules also with n sites (denoted as $\beta$ sites). Bonds form between $\alpha$ and $\beta$ sites. It is
assumed that there is 100% bond formation of all the different sites; that is, every site on an A molecule is bonded to a site on a B molecule and vice versa. Combining an A molecule and a B molecule leads to the formation of an AB molecule. The AB molecule is analogous to the monomer in the Flory-Huggins model. This model of combining two molecules to form new molecule can be used as a slightly more complicated model to explain protein phase separation. The process of combining two monomers into a dimer will be developed, then the same process can be generalized for the synthesis of any number of monomers in the polymer. In this case, the change in entropy will come from the different number of ways a dimer can be made from two monomers and the loss in translational and rotational freedom from formation of larger polymers [22]. Since the number of sites on each A and B molecule is n, the number of unique ways of bonding the two molecules together to form an AB molecule is n!. Similarly, the number of unique ways of combining two A and two B molecules is (2n)!. There is one small caveat; the (2n)! number of ways stated includes ways that form two separate AB molecules. Only the number of ways to form the A₂B₂ molecule matter. Therefore, the number of ways to form two separate AB molecules needs to be determined and subtracted from (2n)! to give the number of ways to form only the A₂B₂ molecule. As stated before, each AB molecule can form in an n! number of ways, and there are two ways to pick an A and a B molecule to form a bond. Thus, the number of ways to for two AB molecules from two A and two B molecules is 2(n!)(n!). Subtracting this value from (2n)! gives (2n)! − 2(n!)(n!) ways to form an A₂B₂ molecule. This kind of process can be continued for combining more than 2 AB molecules by using an iterative process, using the formula of the previous number to determine the formula for the next one. For large values of polymer length, the number of ways of forming an AₘBₘ molecule is approximately (mn)!. Therefore, the change in the entropy due to the difference in number of microstates can be found as ∆S_{\text{configuration}} \approx k_b \ln((mn)! − m \ln(n!)) \approx mn\ln(m), where the last result is obtained via Sterling’s approximation. This contribution to the entropy change is a positive quantity that is linear in the valency of the proteins and linear in the number of A and B molecules. The other contribution for the entropy change is ∆S_{\text{diffusion}} = k_b [(m − 1)\Delta S], where ∆S is the change in entropy due to the assimilation of one AB molecule into the larger polymer. This contribution is a negative quantity that depends on the concentration but not on the valency [22]. A change in the effective valency of a system will have a noticeable effect on ∆S_{\text{configuration}}, and therefore is an important concept regarding the formation of phase separated condensates. This framework provides a way for the incorporation of polymerization of individual monomers into a model for protein phase separation.
2.3 Patchy colloid model

The patchy colloid model can be thought of as an extension to the multivalent interaction model. In this model, molecules are hard spheres with interacting regions represented as physical patches on their surface [27]. In the paper by Zhou and Nguemaha, they form a model for liquid–liquid phase separation, which will be summarized next [28]. First, they distinguish between protein-like (P) and regulatory (R) particles; each is a hard sphere with attractive patches on its surface. A schematic of a protein-like particle (P) and a regulatory particle (R) is shown in Figure 4. In their model, the number of patches on P particles is 4, and the number of patches on R particles is 2. The patches on each particle cover the same fraction of the surface area of each sphere. The patches on the P particles are arranged in a tetrahedron pattern, while the patches on the R particles lie on the poles. A bond in this context is formed when the patches on two particles touch each other. The interaction energy between two particles is defined as \( U_{ij} = U_{sw}(r_{ij})U_{pp}(r_{ij}, \Omega_i, \Omega_j) \) where \( r_{ij} \) is the distance between the centers of two particles, \( r_i \) is the unit vector between the centers, and \( \Omega_i \) and \( \Omega_j \) correspond to the orientation of particles i and j, respectively. The first term corresponds to the square well potential that is defined in equation (1).

\[
U_{sw}(r_{ij}) = \begin{cases} 
\infty, & r_{ij} \leq \sigma \\
-\varepsilon_{ij}, & \sigma < r_{ij} \leq \sigma + \lambda \\
0, & r_{ij} \geq \sigma + \lambda 
\end{cases}
\]  

(1)

where \( \varepsilon_{ij} \) is the strength of the interparticle attraction, \( \lambda \) is the spatial length of the attraction, and \( \sigma \) is the diameter of both particles. The second term describes the particle–particle interaction, equation (2).

\[
U_{pp}(r_{ij}, \Omega_i, \Omega_j) = \begin{cases} 
1, & \text{if an } i-j \text{ bond is formed} \\
0, & \text{otherwise}
\end{cases}
\]  

(2)

In the last two equations, i and j can be either P or R for the two types of particles. The interaction between P and R particles is expected to depend upon the particular molecules involved in the system. Zhou and Nguemaha set the interaction between P particles to be strongly attractive, and the interaction between R particles to be repulsive. After this theoretical framework was set, the authors used Gibbs-ensemble Monte Carlo simulations to produce phase diagrams for various ratios of P to R particles. In the simulations, the authors found that the R particles can either promote or inhibit liquid–liquid phase separation, depending on the molar ratio of the P and R particles and the ratio of protein–protein interactions and protein–regulator interactions. For weak protein–regulator interactions, the regulator particles cause liquid–liquid phase separation by acting as a volume-exclusion crowder. This effect is amplified when the
regulator to protein ratio is high. On the other hand, when the protein–regulator interactions are strong, the regulator protein promotes liquid-liquid phase separation at low R to P molar ratios, and then at higher ratios the regulators start to inhibit liquid-liquid phase separation [28]. In this model, the P and R particles are treated as rigid bodies; however, for many proteins their constituent molecules can bend and change shape, as is the case for proteins with IDRs. The general patchy colloid model incorporates the geometry of particles present in a system, and thus gives researchers the ability to finely adjust their unique model to represent the experimental system more accurately. There are other models developed for various systems that also use patchy colloid theory [7,29]. Typically, the differences are attributed to modifications to the interactions between particles and the number of patches on the particles.

2.4 Amino acid composition model

The interactions that lead to phase separation discussed thus far have been generic and abstract. Interactions between various molecules have been discussed, but the specific molecules or regions of molecules have not been identified. In this section, we will go into more detail about the specific molecules involved in phase separation in the FUS family of proteins. FUS is a family of proteins that have similar domains, an intrinsically disordered prion-like domain (PLD) and an RNA-binding domain (RBD) [30,31]. Wang et al. have discovered that interactions between these two types of domains can drive phase separation [32]. In their paper, the authors were interested in the interactions between tyrosine in the PLD and arginine in the RBD since there is evidence that cation-π interactions, noncovalent interactions between electron rich systems with π bonds (e.g., benzene, ethylene, acetylene) and cations, may drive protein phase separation [3,13,33,34]. To study the importance of these interactions, they precisely mutated these two amino acids in FUS. Tyrosine residues in the PLD of FUS were replaced with serine residues and most arginine residues were replaced with glycine residues in the RBD. It was observed that phase separation for the altered proteins did not occur at concentrations up to 30 µM, compared to the saturation concentration for the original FUS protein was about 2 µM. Therefore, the interactions between tyrosine residues in the PLD and arginine residues in the RBD are important factors promoting phase separation. Additional variants of FUS with mutated amino acids were created to study the difference between generic cation-π interactions and the specific interaction between tyrosine and arginine. It was found that generic cation-π interactions are not the only reason tyrosine and arginine interactions promote protein phase separation; the chemical structure of the tyrosine and arginine side chains is also important [32]. The authors also highlighted
the complex types of interactions and their sometimes-conflicting effects on protein phase separation. For example, they found that electrostatic interactions are important for phase separation of proteins with both PLD and RBD domains. However, these electrostatic interactions tend to inhibit phase separation between PLD domains alone. More research of this level of sophistication and thoroughness is required to determine the important interactions for phase separation in other proteins.

2.5 Biophysical simulations

While theoretical models provide crucial insight on phase separation, simulations of biophysical protein properties and phase separation based on phenomenological models are equally important. Rohit Pappu and his lab have developed tools that can analyze aspects of proteins in biological systems and predict how the proteins will behave. We will briefly discuss two of the simulation tools based on phenomenological models. The first tool CIDER, Classification of Intrinsically Disordered Ensemble Relationships, can analyze the amino acid sequences of intrinsically disordered proteins (IDP) to simulate their biophysical properties. CIDER groups IDPs into five different categories based on the ratios of positive

Figure 5 The phase separated condensates of EWS-FLI1 impact tumorigenic gene transcription in Ewing’s sarcoma. a. EWS-FLI1 fusion protein made from joining partial EWS and FLI1 proteins. b. This diagram represents the impact that condensates of the fusion protein EWS-FLI1 has on the transcription of tumorigenic genes in Ewing’s sarcoma. The FLI1 domain of the fusion protein can bind to the DNA at GGAA microsatellites. BAF protein complexes can bind to EWS-FLI1 or to wild type EWSR1. The protein condensates act to enhance the transcription of the tumorigenic gene.
and negative residues in the protein. There are a couple of assumptions in CIDER’s implementation. First, CIDER assumes that the IDPs have fixed charge states. The other assumption is that the IDP sequence has low hydrophobicity and proline content. The properties of IDPs can then be extrapolated using phenomenological models incorporated within CIDER [35]. The team also developed a second tool called LASSI (Lattice simulation engine for Sticker and Spacer Interactions), which can simulate phase separation in multivalent proteins [36]. This tool uses the model of stickers and spacers to perform simulations of protein phase separation. The stickers can be regions on a protein or short strings of amino acids in an IDR that form interactions with other regions of proteins, while spacers are typically IDRs that control inhomogeneities in the densities of stickers around each other [36,37]. LASSI uses this model on a cubic lattice with boundary conditions that recapitulate real systems. The start of a simulation begins with randomly placing the proteins on the cubic lattice; solvent molecules occupy the sites not occupied by proteins. Simulations can then be conducted in which the conformational change of proteins is examined in a dense or dilute phase. Both CIDER and LASSI have been used by researchers to better understand various systems [33,38]. Other methods such as molecular dynamics simulations can also elucidate the phase behavior of proteins. Molecular dynamics simulations are a strategy to study the movements of atoms or molecules in a system using the forces that act on each individual particle. The forces can be derived from the potential energy which is found through interatomic potentials and molecular structure [39]. Molecular dynamics simulations have been used to study several proteins that can phase separate, such as the N-terminal disordered region of DDX4 protein [40]. The process of creating tools and simulations such as these is important as they help formulate future hypotheses and inspire experiments related to investigating the properties of various proteins and their phase separation behavior.

3. Protein phase separation as an unconventional biophysical mechanism in cancers driven by protein fusion mutations

Cancer has traditionally been studied in cell biology and biochemistry contexts as dysregulation of cell proliferation caused by protein mutations or disruption of the cellular signaling network. Recently, the physics of cancer has been gaining recognition to play complementary and critical roles in driving cancer progression [41,42]. There are some essential features present in all types of cancer and hence define it. The defining characteristics of a cancer cell include cell growth and proliferation that are not controlled by normal cell mechanisms and the invasion of territory typically occupied by cells other than the cancer cells. The way in which cancer cells
achieve these two properties can vary widely in different cancer types. For example, genetic mutations, heredity, and environmental factors all can play a role in the development of cancer in a patient. Oncogenes promote cell growth and proliferation in an unregulated way. Tumor suppression genes inhibit cell growth or can induce cell death. Cancer can arise by the creation of oncogenes through mutations or by the repression of tumor suppression genes. Similarly, mutations in caretaker genes, which originally produce proteins responsible for genome repair and integrity, can also cause cancer. We will discuss several examples of oncogene creation whose cancer-driving mechanisms were recently discovered to utilize protein phase separation. In particular, we will focus on a broad class of cancer mutations known as fusion protein mutations. Fusion proteins are made through deleterious chromosomal rearrangements joining two partial genes that would normally code for separate proteins. Our discussion of the recent work on cancer driven by fusion protein phase separation offers a different biophysical perspective of cancer. The understanding of this new perspective may lead to novel therapeutic treatment strategies that complement existing therapies to prevent or delay the development of cancer resistance, which is the bottleneck in current cancer research.

### 3.1 Protein phase separation in gene transcription

Before discussing the specific details of protein phase separation in gene transcription, we quickly summarize how gene expression is regulated at the transcription level. Transcription is a process that involves transcribing genetic information from DNA to RNA. During transcription, transcription regulatory proteins can bind to specific DNA sequences to regulate the expression (activation or inhibition) of a specific gene. Activation of transcription can occur by activating a paused RNA polymerase, and inhibition of transcription can occur by several methods such as stopping the assembly of general transcription factors or through inhibitor proteins preventing activator proteins from carrying out their function. There are regions on the genomic DNA called cis-regulatory sequences, also called enhancers where transcription regulatory proteins can bind to influence the rate of transcription of a gene. Similarly, superenhancers are regions in the DNA that contain multiple enhancers that increase the probability that transcription will occur for a particular gene. It has been shown that protein clusters of activator proteins can form at superenhancers [43–47]. Whether protein phase separation at enhancers creates an increase in transcription of a gene involved with tumorigenesis has recently been studied in Ewing’s sarcoma [48].

Ewing’s sarcoma is a well-known example of protein phase separation involved in cancer growth [48–50]. Ewing’s sarcoma is a pediatric cancer
caused by mutations in mesenchymal cells which form connective tissue such as bone, cartilage, etc. This type of cancer presents an intriguing case where protein phase separation can impact gene transcription in a drastic way with dire consequences [48]. The driver mutation for this cancer is often the fusion of two proteins EWSR1 and FLI, called EWS-FLI1, as a result from chromosomal translocation [51,52]. The EWSR1 protein is a part of a particular family of proteins called FET. This family of proteins has intrinsically disordered prion-like domains, and they have been shown to form liquid-liquid phase separation [16,53,54]. FLI1 is an ETS, Erythroblast Transformation Specific, transcription factor. In the fusion product, the intrinsically disordered domain of EWSR1 is kept with the ability to cause protein phase separation and the DNA binding domain of FLI1 is kept. Through the retained FLI1 interaction with the genome, the fusion protein EWS-FLI1 can specifically attach to a particular region of DNA called a GGAA microsatellite which is a region of DNA that consists of short repeating pattern of nucleotides [48]. The microsatellite regions here act as enhancers. Once this fusion protein has attached to the DNA, a BAF (BRG1/BRM-associated factor) complex protein is attracted to the EWS-FLI1 protein binding site. The BAF complex is a chromatin remodeling protein that unwinds DNA in nucleosomes, so that other transcription factors can bind to the DNA to initiate transcription of a particular gene. In this case, the genes transcribed lead to uncontrolled cell proliferation and growth and ultimately tumors. Figure 5 depicts the macromolecular complex that disrupts transcription in Ewing’s sarcoma. Interestingly, EWS-FLI1 can recruit BAF complexes to GGAA microsatellites, whereas wild-type FLI1 cannot. It has been shown that this recruitment of BAF to EWS-FLI1 binding site is through the bridging effect of wild type EWSR1: wild type EWSR1 present in the cells can bind BAF; and wild type EWSR1 can bind the intrinsically disordered domain of EWS portion of the fusion protein [48]. The wild type EWSR1 interacts with the fusion protein EWS-FLI1 to from phase-separated macromolecular complexes around the microsatellite enhancers. Accordingly, the transcription rate of the gene corresponding to the microsatellite is consistently increased compared to the transcription rate without the macromolecular complex. Therefore, the presence of the fusion protein in a phase-separated state is the key element to promote tumors.

Another example of phase separation mediated gene transcription is show by Nucleoporin 98 (NUP998). NUP98 is one of many proteins that make up the nuclear pore complex (NPC) which acts as a channel to allow molecules to cross the nuclear envelope. NUP98 has many fusion partners such as HOAX proteins, PMX1, HHEX, and TOP [55]. Fusion proteins involving NUP98 correspond to a variety of different hematopoietic malignancies such as acute myeloid leukemia (AML) and myelodysplastic
syndromes (MDS). It has been shown that fusion proteins with NUP98 can act as aberrant transcription factors [55]. A prevalent fusion protein in leukemia involving NUP98, NUP98-KDM5A, has been found by Terlecki-Zaniewicz et al. to form puncta in the nucleus instead of associating with the NPC as wildtype NUP98 does [56]. Their work suggests that the fusion between the N terminus of NUP98 and the C terminus of fusion partners involved in gene control is sufficient to initiate oncogenic gene expression.

### 3.2 Cell signaling pathways and protein phase separation

Cell signaling is a series of biochemical reactions that influence cellular metabolism and gene expression, leading to changes such as cell proliferation, differentiation, transformation, and programmed cell death. There are a flurry of recent studies showing the involvement of protein phase separation in normal physiological processes such as the regulation of transcription [53,57–60]. One example is the T cell signaling pathway. Su et al. showed that upon activation of T cell receptor (TCR), this pathway leads to effector proteins to phase separate, which allows internal access to some proteins and excludes others [19]. This accelerates the activation of downstream signaling proteins, suggesting that protein phase separation plays a critical role in mounting an immune response.

In contrast, dysregulation of cell signaling through protein phase separation can give rise to cancerous proliferation and growth [61]. In a type of non-small cell lung cancer (NSCLC) that especially affects non-smokers, a driver mutation EML4-ALK is a fusion protein between echinoderm microtubule-associated protein-like 4 (EML4) and anaplastic lymphoma kinase (ALK). ALK belongs to a family of receptor tyrosine kinase (RTK) proteins which are tightly regulated in normal cell cycles through extracellular growth factor stimulations. The conventional RTK signaling starts from the plasma membrane of the cell when the transmembrane RTK dimerizes upon the binding of corresponding growth factors. The dimerized RTK proteins will then phosphorylate each other to create phosphor-tyrosine motif to recruit downstream signaling effectors. Among the RTK downstream signaling pathways, RAS/MAPK is important for cell growth and proliferation, and numerous cancers have mutations significantly upregulating signaling activities in this pathway. This RTK/RAS/MAPK signaling axis was previously thought to exclusively initiate from cell membranes. However, recent work shows the cancer-causing fusion protein EML4-ALK employs an alternative signaling mode via phase-separated protein granules [62]. In the paper by Tulipule et. al., EML4-ALK is not found in the plasma membrane; instead, it is found in the cytoplasm forming micrometer-scale protein assemblies. These cytoplasmic protein granules locally enrich relevant signaling proteins to partition into granules and act effectively as signaling hotspots for the RAS/
MAPK pathway. The phase-separated protein granules are not only necessary but also sufficient for cancer growth signaling: disrupting granule assembly greatly attenuates signaling while forcing granule formation enhances cancer signaling. The ultimate result of EML4-ALK granules can be seen in Figure 6. Furthermore, the work shows that granules initiate the same signaling pathways with the same biochemical signature in conventional membrane-based signaling such as phosphor-tyrosine motifs. This alternative mode of signaling through cytoplasmic protein granules is speculated to be a general mechanism widely applicable to RTK cancer fusion mutations. For example, the authors demonstrate similar signaling platform in another naturally-occurring example of CCDC6-RET and a prototype EGFR artificial granule. Therefore, this perspective of protein phase separation points to a currently understudied area in cell signaling and cancer research with exciting avenues for mechanistic understanding and new therapeutic strategies.

4. Conclusions

The key concepts of the physical mechanisms underlying protein phase separation lies in statistical mechanics. Entropy and interaction strength are the two important factors that determine if mixing or phase separation occurs in biological systems. We have shown several models of protein phase separation, each with its own level of sophistication. The Flory-Huggins model provides intuitive understanding of the phase separation
behavior of proteins in solution by considering a lattice-based binary mixture. The change in entropy due to mixing of the molecules and the interaction energies of molecules are combined to find the change in the Gibbs free energy, the value of which determines the behavior of the protein and solvent system. The concept of valency can be used to find the entropy change due to different conformations between protein monomers. This model gives the entropy change due to polymerization of protein monomers, which is not included in the Flory-Huggins model. The patchy colloid model takes the idea of valency one step further; it deals with the exact locations on molecules that form interactions not just how many interactions exist. The various models stated in this paper can give researchers more methods to study and understand protein phase separation. The work done by Pappu’s lab in developing CIDER and LASSI uses theoretical frameworks to predict the behavior of IDPs and the phase separation behavior in multivalent proteins, respectively. Tools of this kind are essential to further analyze proteins and their affinity for phase separation in biological systems.

Protein phase separation and its role in cancer is a rapidly evolving area of study in which there is still much to learn. The physical principles and cancer examples of protein phase separation presented in this paper provide a brief overview on the driving force of protein phase separation in biological systems and how it can cause cancer. Current cancer treatments typically rely upon the removal of the cancerous cells through surgery, disrupting biochemical processes, such as cell signaling, or dysregulation in transcription that cause cancer or downregulating proteins that are important for cancer formation. Since the underlying mechanism of these cancer systems employ a higher-order protein assembly to promote uncontrolled cell proliferation and growth signaling, a phase-separation-based treatment will disrupt or reduce cancer-promoting protein assemblies. In connection to the physical models discussed here, these treatment strategies may involve blocking multivalent self-associative protein interactions or specific amino acid interactions in prion-like domains. Therefore, the studies of cancer biophysics provide alternative cancer treatment strategies through targeting the biophysical properties of protein phase separation. A better understanding, gained by the models and simulations stated in this review, of the role of protein phase separation in processes of cancer development can potentially lead to alternative cancer treatment strategies that complement existing therapies to prevent or delay the development of cancer resistance, which is critical to combat cancer [63].
Disclosure of potential conflicts of interest

No potential conflict of interest was reported by the author(s).

References

[1] Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell. 2019;176:419–434.
[2] Banani SF, Lee HO, Hyman AA, et al. Biomolecular Condensates: organizers of Cellular Biochemistry. Nat Rev Mol Cell Biol. 2017;18:285–298.
[3] Nott TJ, Petsalaki E, Farber P, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Mol Cell. 2015;57:936–947.
[4] Kedersha N, Anderson P. Mammalian Stress Granules and Processing Bodies. Methods Enzymol. 2007;431:61–81.
[5] Feric M, Vaidya N, Harmon TS, et al. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell. 2016;165:1686–1697.
[6] Mitrea DM, Cika JA, Guy CS, et al. Nucleophosmin Integrates within the Nucleolus via Multi-Modal Interactions with Proteins Displaying R-Rich Linear Motifs and RNA. eLife. 2016;5:e13571.
[7] Gögelein C, Nägele G, Tuinier R, et al. A Simple Patchy Colloid Model for the Phase Behavior of Lysozyme Dispersions. J Chem Phys. 2008;129:085102.
[8] Babinchak WM, Surewicz WK. Liquid–Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. J Mol Biol. 2020;432:1910–1925.
[9] Shin Y, Brangwynne CP. Liquid Phase Condensation in Cell Physiology and Disease. Science. 2017;357:eaaf4382.
[10] Hyman AA, Weber CA, Jülicher F. Liquid-Liquid Phase Separation in Biology. Annu Rev Cell Dev Biol. 2014;30:39–58.
[11] Brangwynne CP, Eckmann CR, Courson DS, et al. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science. 2009;324:1729–1732.
[12] Mateju D, Franzmann TM, Patel A, et al. An Aberrant Phase Transition of Stress Granules Triggered by Misfolded Protein and Prevented by Chaperone Function. EMBO J. 2017;36:1669–1687.
[13] Lee KH, Zhang P, Kim HJ, et al. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell. 2016;167:774–788.e17.
[14] Molliex A, Temirov J, Lee J, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell. 2015;163:123–133.
[15] Murakami T, Qamar S, Lin JQ, et al. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron. 2015;88:678–690.
[16] Patel A, Lee HO, Jawerth L, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell. 2015;162:1066–1077.
[17] Wei M-T, Elbaum-Garfinkle S, Holehouse AS, et al. Phase Behaviour of Disordered Proteins Underlying Low Density and High Permeability of Liquid Organelles. Nat Chem. 2017;9:1118–1125.
[18] Brangwynne CP, Tompa P, Pappu RV. Polymer Physics of Intracellular Phase Transitions. Nat Phys. 2015;11:899–904.

[19] Chang L-W, Lytle TK, Radhakrishna M, et al. Sequence and Entropy-Based Control of Complex Coacervates. Nat Commun. 2017;8:1273.

[20] Zhou HX, Nguemaha V, Mazarakos K, et al. Why Do Disordered and Structured Proteins Behave Differently in Phase Separation? Trends Biochem Sci. 2018;43:499–516.

[21] Lin YH, Chan HS. Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. Biophys J. 2017;112:2043–2046.

[22] Li P, Banjade S, Cheng HC, et al. Phase Transitions in the Assembly of Multivalent Signalling Proteins. Nature. 2012;483:336–340.

[23] Martin EW, Holehouse AS, Peran I, et al. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. Science. 2020;367:694–699.

[24] Feng Z, Chen X, Wu X, et al. Formation of Biological Condensates via Phase Separation: characteristics, Analytical Methods, and Physiological Implications. J Biol Chem. 2019;294:14823–14835.

[25] Banani SF, Rice AM, Peeples WB, et al. Compositional Control of Phase-Separated Cellular Bodies. Cell. 2016;166:651–663.

[26] Banjade S, Rosen MK. Phase Transitions of Multivalent Proteins Can Promote Clustering of Membrane Receptors. eLife. 2014;3:e04123.

[27] Bianchi E, Blaak R, Likos CN. Patchy Colloids: state of the Art and Perspectives. Phys Chem Chem Phys. 2011;13:6397–6410.

[28] Nguemaha V, Zhou H-X. Liquid-Liquid Phase Separation of Patchy Particles Illuminates Diverse Effects of Regulatory Components on Protein Droplet Formation. Sci Rep. 2018;8:6728.

[29] Hvozd T, Kalyuzhnyi YV, Vlachy V. Aggregation, Liquid-Liquid Phase Separation, and Percolation Behaviour of a Model Antibody Fluid Constrained by Hard-Sphere Obstacles. Soft Matter. 2020;16:8432–8443.

[30] Crozat A, Áman P, Mandahl N, et al. Fusion of CHOP to a Novel RNA-Binding Protein in Human Myxoid Liposarcoma. Nature. 1993;363:640–644.

[31] King OD, Gitler AD, Shorter J. The Tip of the Iceberg: RNA-Binding Proteins with Prion-like Domains in Neurodegenerative Disease. Brain Res. 2012;1462:61–80.

[32] Wang J, Choi JM, Holehouse AS, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018;174:688–699.e16.

[33] Pak CW, Kosno M, Holehouse AS, et al. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol Cell. 2016;63:72–85.

[34] Boeynaems S, Bogaert E, Kovacs D, et al. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell. 2017;65:1044–1055.e5.

[35] Holehouse AS, Das RK, Ahad JN, et al. CIDER: resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J. 2017;112:16–21.

[36] Choi J-M, Dar F, Pappu RV. LASSI: a Lattice Model for Simulating Phase Transitions of Multivalent Proteins. PloS Comput Biol. 2019;15:e1007028.

[37] Choi JM, Holehouse AS, Pappu RV. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annu Rev Biophys. 2020;49:107–133.

[38] Ruff KM, Dar F, Pappu RV. Ligand Effects on Phase Separation of Multivalent Macromolecules. Proc Nat Acad Sci. 2021;118:e2017184118.
[39] Hospital A, Goñi JR, Orozco M, et al. Molecular Dynamics Simulations: advances and Applications. Adv Appl Bioinform Chem. 2015;8:37–47.

[40] Paloni M, Bailly R, Ciandrini L, et al. Unraveling Molecular Interactions in Liquid-Liquid Phase Separation of Disordered Proteins by Atomistic Simulations. J Phys Chem B. 2020;124:9009–9016.

[41] White FM, Gatenby RA, Fischbach C. The Physics of Cancer. Cancer Res. 2019;79:2107–2110.

[42] Michor F, Liphardt J, Ferrari M, et al. What Does Physics Have to Do with Cancer? Nat Rev Cancer. 2011;11:657–670.

[43] Boija A, Klein IA, Sabari BR, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018;175:1842–1855. e16.

[44] Peng L, Li E-M, Xu L-Y. From Start to End: phase Separation and Transcriptional Regulation. Biochim Biophys Acta Gene Regul Mech. 2020;1863:194641.

[45] Sabari BR, Dall’Agnese A, Boija A, et al. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. Science. 2018;361:eaa3958.

[46] Chong S, Dugast-Darzacq C, Liu Z, et al. Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription. Science. 2018;361:eaa2555.

[47] Cai D, Feliciano D, Dong P, et al. Phase Separation of YAP Reorganizes Genome Topology for Long-Term YAP Target Gene Expression. Nat Cell Biol. 2019;21:1578–1589.

[48] Boulay G, Sandoval GJ, Riggi N, et al. Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. Cell. 2017;171:163–178.e19.

[49] Delattre O, Zucman J, Plougastel B, et al. Gene Fusion with an ETS DNA-Binding Domain Caused by Chromosome Translocation in Human Tumours. Nature. 1992;359:162–165.

[50] Jiang S, Fagman JB, Chen C, et al. Protein Phase Separation and Its Role in Tumorigenesis. eLife. 2020;9:1–27.

[51] Crompton BD, Stewart C, Taylor-Weiner A, et al. The Genomic Landscape of Pediatric Ewing Sarcoma. Cancer Discov. 2014;4:1326–1341.

[52] Tirode F, Surdez D, Ma X, et al. Genomic landscape of ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. 2014;4:1342–1353.

[53] Altmeyer M, Neelsen KJ, Teloni F, et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-Ribose). Nat Commun. 2015;6:8088.

[54] Li YR, King OD, Shorter J, et al. Stress granules as crucibles of ALS pathogenesis. J Cell Biol. 2013;201:361–372.

[55] Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. Blood. 2011;118:6247–6257.

[56] Terlecki-Zaniewicz S, Humer T, Eder T, et al. Biomolecular Condensation of NUP98 Fusion Proteins Drives Leukemogenic Gene Expression. Nat Struct Mol Biol. 2021;28:190–201.

[57] Case LB, Ditlev JA, Rosen MK. Regulation of transmembrane signaling by phase separation. Annu Rev Biophys. 2019;48:465–494.

[58] Boehning M, Dugast-Darzacq C, Rankovic M, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol. 2018;25:833–840.
[59] Cho WK, Spille JH, Hecht M, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science. 2018;361:412–415.

[60] Hnisz D, Shrinivas K, Young RA, et al. A phase separation model for transcriptional control. Cell. 2017;169:13–23.

[61] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–674.

[62] Tulpule A, Guan J, Neel DS, et al. Kinase-mediated ras signaling via membraneless cytoplasmic protein granules. Cell. 2021;184:2649–2664.e18.

[63] Boeynaems S, Alberti S, Fawzi NL, et al. Protein phase separation: a new phase in cell biology. Trends Cell Biol. 2018;28:420–435.