A thermodynamic description of biomolecular condensates

J. Matthew Dubach*

*dubach.matt@mgh.harvard.edu

Institute for Innovation in Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA

Abstract
Biomolecular condensates are intracellular membrane-less accumulations of proteins and other molecules at a higher concentration than the rest of the cell. The recent characterization of condensates as liquid-like assemblies has stimulated profound interest in how physical properties of condensates impact molecular biology. Intriguingly, condensates have been shown to be essential to multiple different cellular processes and underlying aspects of various diseases. Yet, the physics of condensate formation remains unsolved. Here, it is shown that intrinsically disordered protein-protein binding alone provides energetically favorable thermodynamics for condensate formation. The reduction in free energy achieved through increased binding at high condensate concentrations can overcome the entropic cost of de-mixing. Formation of condensates is governed by the ratio of total protein concentration to binding affinity ($[protein]_{total}/K_d$). Yet, stable condensation is only possible through interactions with rapid binding dynamics. The model prediction and experimental observation that condensates are no longer formed at high $[protein]_{total}/K_d$ ratios redefines our understanding of condensate physics and impact on cellular biology.

Main Text
Biomolecular condensates are localized accumulations of macromolecules at concentrations above the dilute background that lack membrane boundaries. Condensates are characterized by: their ability to rapidly exchange most components with the surrounding environment (1); their liquid-like properties that allow them to flow, move and merge (2); their dependency on environmental conditions (3); the presence of oliqonucleotides (4); the abundance of fully intrinsically disordered proteins or proteins that contain intrinsically disordered regions (IDR) (5); their ability to rapidly form or dissolve (2); and, in cells, inclusion of multiple proteins (6). Condensates arise naturally in cells and can also be recreated *in vitro* in single protein or multiple component systems. Condensates are thought to be essential to cellular biology and critical in myriad human disease (7). However, the physics of condensate formation remains unclear.

Fundamentally, to exist spontaneously, the presence of biomolecular condensates must produce an overall system free energy that is lower than dilute distribution of components. Thermodynamically, condensates produce a lower entropy of mixing than dilute distribution. Therefore, to achieve condensation there must be energetic advantages that overcome this entropic cost. Initial considerations of biomolecular condensates modeled their formation after the phenomenon known as liquid-liquid phase separation (LLPS) (2). LLPS arises when two or more liquids are mixed at concentrations where it is entropically favorable to exist in two phases - the entropic cost of condensation is less than the entropic loss that arises from restricted molecular orientations of the abundant liquid surrounding each of the less-abundant molecules. However, some physical properties of condensates do not fit the LLPS model (8). Thus, pure LLPS is likely not the mechanism of most condensate formation (9).

Gelation (or percolation) is often used to describe condensate formation (10, 11). Because the presence of IDRs are essential for condensate formation (1), models from polymer chemistry have been leveraged to understand the physics of condensates. In describing polymer physics, Flory and Huggins (12, 13) initially showed that, depending on molecular interaction energies, polymers can phase separate into dense and dilute phases (11). In this model, the difference in free energy of polymer-polymer and polymer-solvent interactions can overcome the entropic cost of condensation at a critical concentration for the given system conditions. The energetic advantage of forming condensation through interaction energy is defined as the Flory-Huggins parameter χ:

$$\chi \sim \left[u_{ps} - \frac{1}{2}(u_{pp} + u_{ss})\right]$$

where u is the energy of interaction that arises between protein (p) and solvent (s).

Condensation occurs at very large χ values, suggesting that relatively high energy polymer-solvent interactions (u_{ps}) and/or relatively low energy polymer-polymer (u_{pp}) and solvent-solvent (u_{ss}) interactions are required. While instructive, the Flory-Huggins model is based on homopolymers that are not necessarily representative of the multiple different amino acids found in IDRs (14). Initially, the abundance of polar amino acids found in IDRs for which aqueous solutions are bad solvents, such as G, Q, N and S, were thought to produce a high χ value through high u_{ps} interactions (10, 15). Yet, observations that altered order of amino acids (16) and replacement of charged amino acids (17) prevents condensate formation demonstrate that high u_{ps} is not necessarily a driver of biomolecular condensates for all proteins. Therefore, since u_{ss} is a constant when amino acids are mutated or rearranged in IDRs, low u_{pp} must be a critical factor in condensation through the Flory-Huggins model. Given the flexibility of IDRs it is reasonable to approximate that protein-protein interactions are capable of achieving a minimum energy through amino acid interactions across the length of the IDR. In this scenario, a maximum χ is achieved when the number of interactions with $u_{pp} < u_{ps}$ is high and the number of interactions where $u_{pp} > u_{ps}$ is low. And, integration of the interaction energy for all interactions, u_{pp} and u_{ps}, defines the energy minimum for protein-protein interaction in the solvent. For two protein molecules binding, this integration is also the definition of binding affinity.

Achieving low u_{pp} interactions while avoiding high u_{pp} interactions led to the model of IDR interactions known as “stickers and spacers” (17). In this model, stickers (low u_{pp}) in multivalent macromolecules ranging from single amino acids to structured binding domains, such as SUMO (18), can interact within multivalent molecules to form networks of macromolecules capable of condensing into dense regions with a dependency on the valency (19). Indeed, IDR sequences have been shown to determine the dense phase concentration and material properties of condensates (20, 21). Loss of a single sticker will increase the saturation concentration of condensate formation, but loss of a spacer should have little impact (14), suggesting that low u_{pp} interactions are the drivers of condensation.

Yet, the fundamental question of how gelation-induced phase separation is impacted by complex environments, such as the cytoplasm, remains outstanding (9). This question is particularly acute with single amino acid sticker interactions within the cell where cognate sticker amino acids are abundant in solution and in other protein species. Based on the Flory-Huggins model, at low values of χ, where there are not large overall differences in u_{pp} and u_{ps}, the system will not phase separate, but instead form a macro-gel when higher concentrations are reached (9). In cells there are thousands of types of molecules that could interact with amino acids in IDRs, including molecules that achieve very low energy interaction energy with stickers. Thus, low u_{ps} interactions that have the same energy as low u_{pp} interactions will be prevalent. Therefore, it is likely that to achieve a minimum energy capable of driving condensation, interactions across a large percentage of the IDR need to be integrated.

Here, let us consider that proteins interact through binding regions with a defined binding dissociation constant and not an ensemble of individual stickers and spacers (**Fig. 1A**). Integrating all the sticker interactions into one affinity value in this manner captures the previously observed dependencies on protein composition, valency and order. From this perspective, sequence mutations in disordered proteins that prevent condensate formation (17) do not merely remove sticker interactions, but decrease the overall affinity of protein binding. In a single protein species system, IDRs can interact with multiple proteins, but, here, to understand the impact of binding on condensation, only bivalent interactions (one protein binding to another) are considered. Several factors suggest that bivalent interactions exist, including the ability of all the low u_{pp} interactions to occur between just two IDR protein molecules. As a simple binding interaction defined by the dissociation constant, the degree of binding depends on the concentration (**Fig. 1B**) (22). To generalize this consideration, the amount of binding is a function of the ratio between total protein concentration and the dissociation constant

([protein]$_{total}$/K_d). If a system is at low [protein]$_{total}$/K_d, increasing the protein concentration will profoundly increase the fraction of bound protein (**Fig. 1C**). However, if the system is initially at high [protein]$_{total}$/K_d, increasing the protein concentration has less of an impact on the fraction of bound protein.

When a condensate is formed, the concentration in the condensate is higher than the initial system with no condensation, increasing the fraction of bound protein within the condensate (**Fig. 1D**). Critically, assuming the volume fraction of the condensate is less than the dilute region, the overall fraction of bound protein will be higher when condensates are present (**Fig. S1A**). This increase in total bound fraction upon condensation serves to decrease the effective dissociation constant of the system (**Fig. 1E**). Therefore, condensation generates a lower free energy through increased binding (**Fig. S1B**), defined as $\Delta G_{binding}$:

$$\Delta G_{binding} = RTln\left(\frac{K_{d,effective}}{K_d}\right)$$

where, R is the gas constant, T is the temperature, K_d is the binding affinity, and $K_{d,effective}$ is the effective binding constant upon condensation. And, this shift in free energy is dependent on the total protein in the system and dissociation constant of binding, or [protein]$_{total}$/K_d. Yet, with increased partition coefficient, the entropic cost of condensation will also increase (**Fig. S1C**), with the entropic cost calculated from:

$$\Delta G_{condensation} = RT\left(M_{f,cond} \cdot ln\left(\frac{[protein]_{cond}}{[protein]_{initial}}\right) + M_{f,dilute} \cdot ln\left(\frac{[protein]_{dilute}}{[protein]_{initial}}\right)\right)$$

where, $M_{f,cond}$ is the mole fraction in the condensate, $M_{f,dilute}$ is the mole fraction in the dilute region, [protein]$_{cond}$ is the concentration in the condensate, [protein]$_{dilute}$ is the concentration in the dilute region and [protein]$_{initial}$ is the concentration in the absence of condensation.

In total, the energetic impact of condensation will be the sum of the increased energy of de-mixing and decreased energy from binding (**Fig. 1F**) - here it is assumed that there are no enthalpic contributions to the energetic cost of de-mixing. If this sum is negative, the system will be thermodynamically stable when condensates are present. The value of the energetic advantage is also dependent on the volume fraction of the condensate (**Fig. S1D**). At a volume fraction of 1% and ignoring any concerns about condensate concentration saturation, the total energy minimum, and corresponding entropic cost and change in binding energy, can be calculated for multiple [protein]$_{total}$/K_d values (**Fig. 1G & S1E**). At higher [protein]$_{total}$/K_d the energetic advantage of condensation is lost as a substantial fraction of the protein is bound in the absence of condensation and, thus, the entropic cost outweighs the change in binding energy. Conversely, the total molar energy of condensation plateaus around a [protein]$_{total}$/K_d ratio of 10^{-3} and remains at that point at increasingly lower ratios. However, there are physical limitations at very low ratios. To achieve low [protein]$_{total}$/K_d ratios, either the protein concentration needs to be low or the dissociation constant needs to be high. For low protein concentrations the total change in energy becomes extremely low relative to the energy of the system and the overall change in free energy will approach thermal fluctuations making condensation unstable (**Fig. 1H**). Similarly, at large dissociation constants the relative binding affinity compared to other molecular interactions becomes small and the energetic binding advantage will disappear (**Fig. 1I**). Therefore, in the binding model for homotypic interactions, there exists a range of [protein]$_{total}$/K_d values that enable condensate formation, bounded by a distinct high [protein]$_{total}$/K_d ratio and a system dependent low ratio.

However, the binding model is also reliant on the binding dynamics. Fluorescence microscopy experiments show that proteins readily exchange between the condensate and dilute region (23-29). In the binding model, at an energetic minimum, the condensate consists of bound and unbound protein, which are both at higher concentrations than in the dilute region. For unbound protein in the condensate, the entropic force to leave is balanced by the energetic advantage of existing at a high condensate concentration to increase the likelihood of binding (**Fig. 2A**). However, the bound protein does not have any energetic advantage to remain within the condensate and is only subject to the gain

in entropy upon diffusing into the dilute region. The change in protein concentration in an established condensate can be expressed as: $\Delta[protein]_{cond} = unbound_{enter} + 2 \bullet bound_{enter} - (unbound_{leave} + 2 \bullet bound_{leave})$. The thermodynamic forces governing bound proteins suggest that $bound_{leave} >> bound_{enter}$, and at steady state $\Delta[protein]_{cond} = 0$. Therefore, at steady state, $unbound_{enter} = unbound_{leave} + 2 \bullet bound_{leave}$, creating a mass imbalance at equal diffusion rates.

Yet, diffusion out of the condensates is a dynamic process, and over time, a bound protein pair will both diffuse and dissociate (**Fig. 2B**). If dissociation is fast relative to diffusion, when considering condensate protein flux, unbound protein is essentially equivalent to bound protein. And, within the condensate protein is more likely to be unbound than bound (**Fig. 1F**), suggesting that on the time scale of diffusion all protein is effectively unbound. Thus, the relative rates of diffusion and dissociation impact the ability of a condensate to exist (**Fig. 2C**). Recent measurements of IDR protein binding demonstrate that IDRs can interact through extremely fast dynamics, with k_{off} rates ranging from 10^0 to 10^5 s^{-1} (30-32). Furthermore, in cells the dissociation rate will likely be faster than *in vitro* measurements due to the presence of competing binding molecules, which have been shown to increase k_{off} by 20x *in vitro* (33). Critically, the dissociation rate has profound effects on binding dynamics. Stochastic simulations using the Gillespie algorithm (34) demonstrate the dynamics of protein interactions at different binding rates, despite generating the same steady state bound fraction (**Fig. 2D**). Conversely, measured diffusivity rates inside condensates are on the order of 10^{-3} to 10^{-1} $\mu m^2/s$ (23-28). Therefore, it is expected that IDR systems with fast binding rates exist at dynamics where bound protein dissociates before significant diffusion relative to the size of the typical condensate is achieved (**Fig. 2E**). This rapid dissociation, which is unique to IDR interactions, enables condensate formation through simple binding.

Two well established aspects of condensate formation *in vitro* are; 1) a saturation protein concentration (C_{sat}) that defines the low concentration boundary at which condensates form, and 2) altering system properties, such as increasing the salt concentration or heating the sample, among other biophysical changes, will impact the ability of condensates to form. From the perspective of the binding model presented here, the concentration dependency suggests that at low, non-condensate forming concentrations the system resides at very low $[protein]_{total}/K_d$ ratios, and increasing the protein concentration increases the $[protein]_{total}/K_d$ ratio into a range where condensation is thermodynamically favorable (**Fig. 3A**). While, the biophysical changes that impact condensates all have an effect on the binding affinity, leading to a change in $[protein]_{total}/K_d$ ratio that can occur in either direction.

However, one area where the binding model presented here differs from existing condensate theory is the existence of a high $[protein]_{total}/K_d$ ratio where condensates will not form (**Fig. 1G & 3A**). The existing theory suggests that condensates form at any concentration above C_{sat} (9), therefore the vast majority of condensate research does not explore protein concentrations above the C_{sat} concentration. To test the binding model, full length human TDP-43 in PBS was imaged at increasing concentrations. Condensates were visible at concentrations up to 50 μM, however at 100 μM they were no longer present (**Fig. 3B**), suggesting the system reached a $[protein]_{total}/K_d > 0.1$, and that the TDP-43 binding K_d is between 500 μM and 1 mM. Human Tau, using PEG as a molecular crowder (35), showed a similar loss of condensates at a higher $[protein]_{total}/K_d$ ratio (**Fig. 3B**). This loss of condensation phenomenon has also been observed with condensate-inducing molecules (4, 36, 37). In these systems, protein reentrance occurs with the addition of components ranging from small molecules to RNA. Introduction of these molecules forms condensates at IDR protein conditions below C_{sat}, however, as the concentration of these condensate-inducing molecules increases, condensates eventually disappear. Based on previous findings (37), here, concentrations of Bis-ANS up to 50 μM formed condensates with 5 μM TDP-43 in 20 mM potassium phosphate, pH 6, but condensates were no longer present at 100 μM Bis-ANS (**Fig. 3C**). The driver of IDR proteins reentering the dilute region at high concentrations of condensate-inducing molecules is thought to be the concentration ratio of IDR protein to molecule (4, 38). Therefore, it would be expected that increasing the concentration of TDP-43 at 100 μM Bis-ANS would cause condensate formation, however, this is not the case (**Fig. 3C**). Furthermore, increasing the concentration of TDP-43 at 50 μM Bis-ANS caused condensates to disappear. These results suggest that loss of condensation at high condensate-inducing molecule concentrations is driven by the high concentration alone - agreeing with heterotypic interaction results that are predicted by the binding model (**Fig. S2A**).

Condensate formation in the binding model is also dependent on the affinity of the binding interaction. Experimentally, TDP-43 condensate formation is dependent on the salt concentration (39), suggesting that salt changes the relative binding affinity. Indeed, when 100 mM NaCl was added to condensate-forming Bis-ANS and TDP-43 concentrations, condensates were no longer present (**Fig. 3D**), suggesting that NaCl shifted the binding affinity outside the $[protein]_{total}/K_d$ condensation range. These results also demonstrate that the binding affinity governing condensation is relative and system dependent.

Lastly, recent observations that the dilute region concentration and partition coefficient are dependent on the total protein concentration do not agree with the existing regular solution theory (40). The existing theory predicts that the partition coefficient remains constant and this experimental deviation was hypothesized to be driven by intracellular heterotypic multicomponent interactions (40). Yet, calculations from the binding model for homotypic interaction show both the partition coefficient and dilute region concentration are dependent on total concentration, suggesting that binding driven condensation produces the experimentally observed results (**Fig. 3E**). Here, to fit published experimental results (40), the total concentration was modeled from 1 μM to 100 μM, the volume fraction increased from 0.5% at 1 μM to 2% at 100 μM and the dissociation constant was fixed at 1 mM.

The binding model also describes heterotypic binding interactions (**Fig. S2A**). Heterotypic interactions often have higher binding affinities than homotypic interactions (41, 42), which suggests this is a route for cellular protein condensation to occur at low protein concentrations. The binding model can also describe the binding energetic advantage of condensation where multiple molecules can bind to a single IDR or oligonucleotide (**Fig. S2B**). Here, the high $[protein]_{total}/K_d$ ratio where condensation is no longer energetically favorable depends on the number of monomers that can bind a polymer and the relative concentration of each component. Furthermore, it is energetically favorable for client proteins or molecules (18) with cognate binding regions on scaffold condensed IDR proteins to be incorporated into the condensates, provided the binding region does not overlap with the condensate forming binding region. The local accumulation of binding sites on IDRs presents a reduction in client binding energy that overcomes entropic cost of client condensation. Thus, for client molecules, there is no binding affinity or concentration limitation, and the energetic minimum depends on the number of client binding sites in the condensate (**Fig. S2C**).

Overall, the binding model suggests condensation occurs within a range of energetically favorable $[protein]_{total}/K_d$ ratios (**Fig. 4A**). This range occurs for both homotypic and heterotypic interactions, where, for 1:1 binding, exceeding a maximum $[molecule]_{total}/K_d$ ratio based on the concentration of either interacting molecule will prevent condensation (**Fig. S2A**). The condensation $[protein]_{total}/K_d$ range relies upon the ability to generate a decrease in binding free energy that overcomes the entropic cost of condensation (**Fig. 4B**). Yet, because the dissociation must be fast and the binding reaction energetically significant, there exists a restricted range of dissociation constants and concentrations capable of generating condensates (**Fig. 4C**).

Similar to the Flory-Huggins model, the binding affinity depends on amino acid interactions and will thus be dependent on the system. Experimentally, condensation has been shown to be dependent on properties such as system temperature (43), IDR phosphorylation (44), and buffering of electrostatic interactions (5, 45, 46), all of which alter the binding affinity. It has been previously suggested that condensates are assemblies that maintain disorder and form through transient protein-protein contacts (23). The binding model presented here describes the thermodynamic forces that enable these assemblies. Furthermore, the binding model suggests that rather than physical crosslinks among multivalent proteins being the cause of condensate formation (47), binding interactions provide enough thermodynamic stability to enable condensates. Interactions with multivalency will impart the same thermodynamic advantage as dimerization since, ultimately, regardless of the number of proteins each individual protein binds, the protein will be bound. And, the finding that at high $[protein]_{total}/K_d$ ratios condensates are no longer formed agrees with intracellular observations that high concentrations will dissolve condensates (48), suggesting that further exploration of high concentrations or properties that increase binding affinity will reveal more of the complex roles condensates play in molecular biology and disease.

Figures

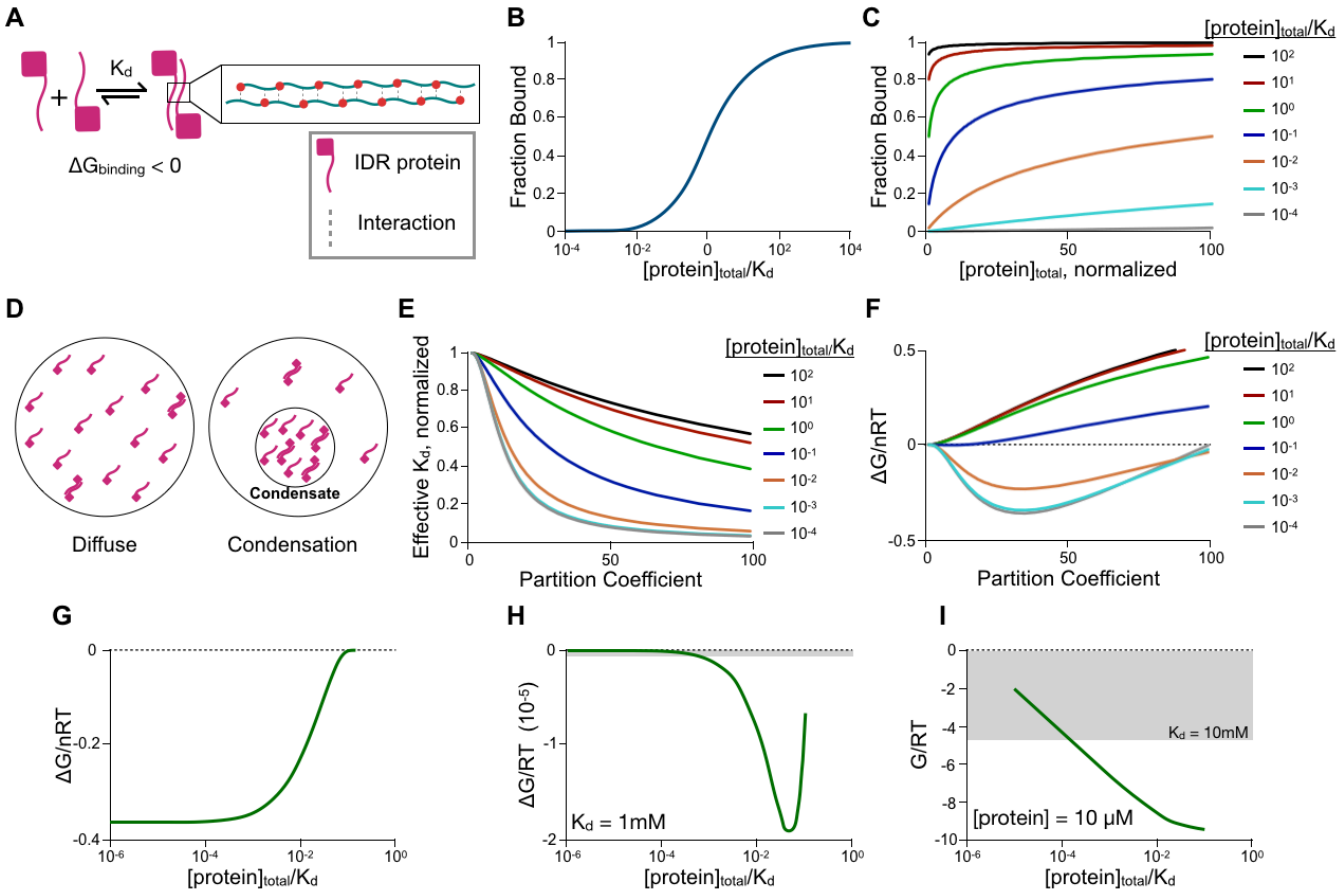

Figure 1: IDR binding provides an energetic advantage for condensation. A) Binding of IDR proteins can be described by a single dissociation constant that captures all the amino acid interactions. **B)** The fraction of bound protein is a function of the $[protein]_{total}/K_d$ ratio. **C)** The fraction of bound protein increases at higher protein concentration, dependent on the $[protein]_{total}/K_d$ ratio. **D)** Shifting from a diffuse state to one with condensates creates high local concentrations where the fraction of bound protein will increase. **E)** The effective K_d (calculated from the total bound and unbound protein) in a condensed state normalized to the diffuse state as a function condensate partition coefficient (ratio of protein concentration in the condensate to the dilute region) for multiple $[protein]_{total}/K_d$ ratios. **F)** Change in total energy as a function of condensate partition coefficient for multiple $[protein]_{total}/K_d$ ratios. **G)** Total energy per mole of protein minimum for condensate formation as a function of $[protein]_{total}/K_d$. **H)** Total energy or condensate formation as a function of $[protein]_{total}/K_d$ at a K_d of 1 mM. The gray area is a region of very low total energy arising from low protein concentration. **I)** Absolute energy of binding as a function of $[protein]_{total}/K_d$. at a protein concentration of 10 μM. The gray area is binding energy for K_d values > 10 mM. For all results, the condensate volume fraction is 1%.

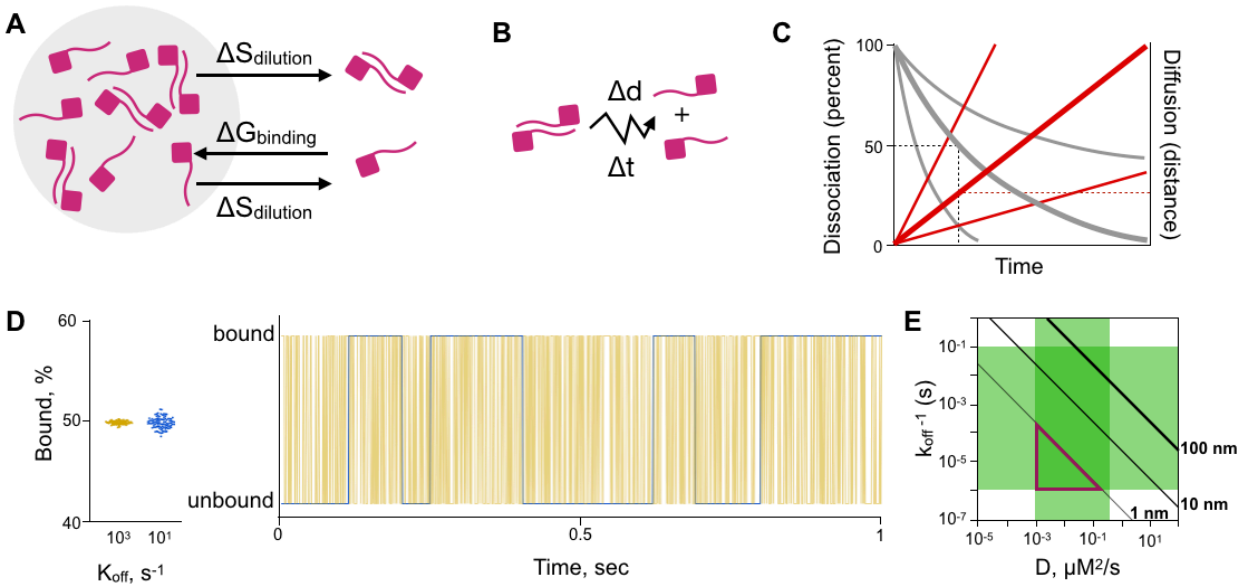

Figure 2: Binding dynamics enable stable condensate formation. A) Protein freely diffuses into and out of a condensate. Unbound protein is subject to entropic forces to leave the condensate but an energetic advantage to enter or remain. Bound protein is only subject to the entropic force to leave the condensate. **B)** Bound protein will diffuse during the bound lifetime before dissociating. **C)** Both dissociation (gray) and diffusion distance (red) (mean squared displacement) are rates that have distinct values that will govern the diffusion of bound proteins. At the intermediary rates (thick lines), dissociation will occur at an average time, which dictates how far the bound protein will diffuse dependent on the diffusivity. **D)** Bound fraction of stochastic simulations (n=100) for proteins with different binding rates, but the same K_d, and a single simulation showing the bound state of a single protein over time. **E)** A plot of $1/K_{off}$ vs. diffusivity, with expected condensate values in green. Shown are lines for mean squared displacement at multiple length scales. The purple triangle represents expected condensate values where bound protein will dissociate before it diffuses 1 nm.

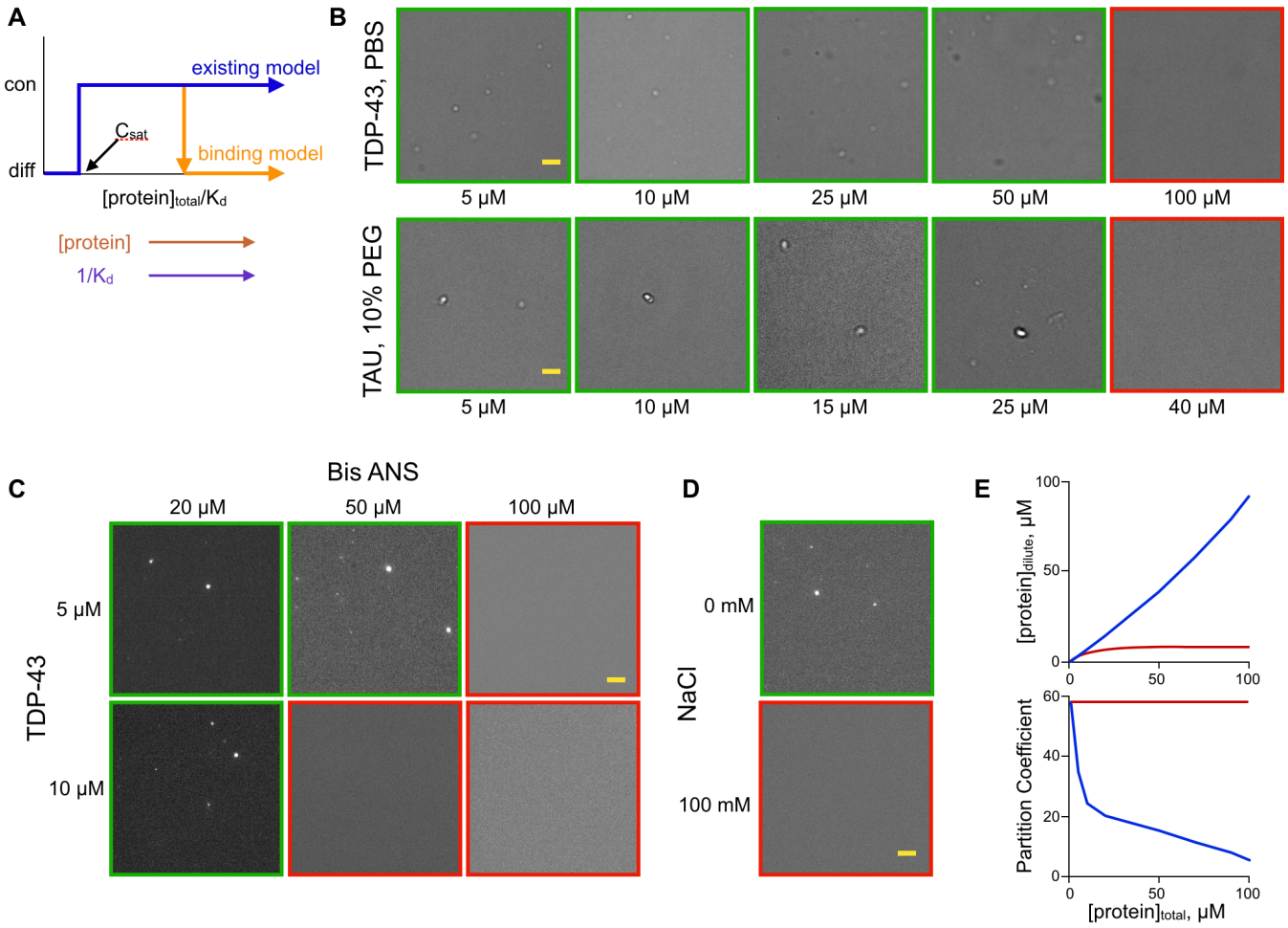

Figure 3: The binding model accurately predicts condensate properties. A) Both the existing model and the binding model predict a C_{sat} concentration where condensates spontaneously form. However, the existing model predicts condensates will be stable at any concentration above C_{sat}, while the binding model predicts a high $[protein]_{total}/K_d$ ratio where condensation is no longer favorable. The binding model also predicts that binding affinity will impact condensate formation. **B)** DIC images of full length recombinant human TDP-43 in PBS or recombinant human Tau in PBS with 10% PEG at different concentrations, green outlines indicate condensate formation while red outlines indicate a lack of condensates. **C)** Fluorescent images of Bis-ANS and TDP-43 in 20 mM potassium phosphate at pH 6. **D)** Fluorescent images of 50 μM Bis-ANS and 5 μM TDP-43 in 20 mM potassium phosphate at pH 6 with or without 100 mM NaCl. **E)** Binding model calculations (blue) of homotypic interaction forming condensates showing dilute region concentration (top) and partition coefficient (bottom) as a function of total protein concentration. Shown in red are the regular solution theory expected values. All images, scale bar = 10 μm.

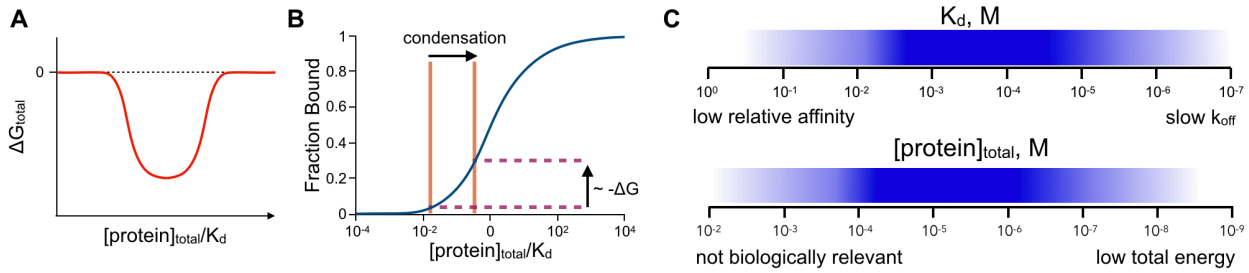

Figure 4: Condensation constraints based on the binding model. A) The binding model demonstrates there exists a range of $[protein]_{total}/K_d$ ratios where IDRs can form condensates with favorable thermodynamic energy. **B)** Condensation increases the fraction of protein binding, which causes a decrease in total energy. **C)** IDR systems capable of forming homotypic or heterotypic 1:1 binding condensates are limited to certain dissociation constant values and and protein concentrations.

Methods and Materials

Binding model calculations.

All calculations were performed in Microsoft excel using equations that define the dissociation constant and the laws of mass balance.

Protein condensates.

Recombinant human TDP-43 and recombinant human Tau (R and D systems) were concentrated and solvent exchanged into either PBS or 20 mM potassium phosphate, pH 6.0, using 10 kDa MWCO microcentrifuge protein concentrator PES (Pierce). Solvent exchange was performed 4 times. Protein concentration was determined by BCA assay (Pierce). Bis-ANS (Caymen Chemical) was dissolved in at 1 mM in 20 mM potassium phosphate, pH 6.0 and diluted as needed. Each sample was bath sonicated and imaged immediately in a glass bottom 96 well plate (CellVis). Images were acquired on a Nikon Ti-2 widefield microscope with a 60x/1.49NA oil objective. Condensates were visualized with DIC. Bis-ANS was excited with a 405 laser and emission was collected through a quad laser TIRF filter cube.

Stochastic simulation.

Stochastic simulations of protein binding were performed using the Gillespie algorithm in Matlab. Simulations were run at a concentration of 1 μM protein and a heterotypic binding dissociation constant of 1 μM.

Funding: This work was supported by a grant from the National Institutes of Health (R01CA241179) and utilized equipment funded by the Massachusetts Life Science Center.

Competing interests: No conflicts of interest are declared.

References

1. S. Chong *et al.*, Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
2. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).
3. Timothy J. Nott *et al.*, Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell* **57**, 936-947 (2015).
4. S. Maharana *et al.*, RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918-921 (2018).
5. A. Molliex *et al.*, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-133 (2015).
6. W. Xing, D. Muhlrad, R. Parker, M. K. Rosen, A quantitative inventory of yeast P body proteins reveals principles of composition and specificity. *eLife* **9**, e56525 (2020).
7. S. F. Banani *et al.*, Genetic variation associated with condensate dysregulation in disease. *Dev Cell* **57**, 1776-1788 e1778 (2022).
8. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev* **33**, 1619-1634 (2019).
9. T. Mittag, R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol Cell* **82**, 2201-2214 (2022).
10. Clifford P. Brangwynne, P. Tompa, Rohit V. Pappu, Polymer physics of intracellular phase transitions. *Nature Physics* **11**, 899-904 (2015).
11. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife* **6**, (2017).
12. P. J. Flory, Thermodynamics of high polymer solutions. *The Journal of chemical physics* **10**, 51-61 (1942).
13. M. L. Huggins, Some properties of solutions of long-chain compounds. *The Journal of Physical Chemistry* **46**, 151-158 (1942).
14. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics* **49**, 107-133 (2020).
15. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **32**, 102-112 (2015).
16. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699 (2020).
17. J. Wang *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e616 (2018).
18. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).
19. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).
20. M. T. Wei *et al.*, Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat Chem* **9**, 1118-1125 (2017).
21. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
22. I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, How to measure and evaluate binding affinities. *eLife* **9**, e57264 (2020).
23. Kathleen A. Burke, Abigail M. Janke, Christy L. Rhine, Nicolas L. Fawzi, Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular Cell* **60**, 231-241 (2015).
24. S. Elbaum-Garfinkle *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* **112**, 7189-7194 (2015).
25. H. Zhang *et al.*, RNA Controls PolyQ Protein Phase Transitions. *Molecular Cell* **60**, 220-230 (2015).
26. P. P. Gopal, J. J. Nirschl, E. Klinman, E. L. F. Holzbaur, Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proceedings of the National Academy of Sciences* **114**, E2466-E2475 (2017).

27. N. O. Taylor, M.-T. Wei, H. A. Stone, C. P. Brangwynne, Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophysical Journal* **117**, 1285-1300 (2019).
28. A. W. Folkmann, A. Putnam, C. F. Lee, G. Seydoux, Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**, 1218-1224 (2021).
29. J. Miné-Hattab *et al.*, Single molecule microscopy reveals key physical features of repair foci in living cells. *eLife* **10**, e60577 (2021).
30. A. Borgia *et al.*, Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61-66 (2018).
31. S. S. Stadmiller, J. S. Aguilar, C. A. Waudby, G. J. Pielak, Rapid Quantification of Protein-Ligand Binding via 19F NMR Lineshape Analysis. *Biophysical Journal* **118**, 2537-2548 (2020).
32. P. O. Heidarsson *et al.*, Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. *Nat Chem* **14**, 224-231 (2022).
33. A. Sottini *et al.*, Polyelectrolyte interactions enable rapid association and dissociation in high-affinity disordered protein complexes. *Nat Commun* **11**, 5736 (2020).
34. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340-2361 (1977).
35. N. M. Kanaan, C. Hamel, T. Grabinski, B. Combs, Liquid-liquid phase separation induces pathogenic tau conformations in vitro. *Nature Communications* **11**, 2809 (2020).
36. P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, A. A. Deniz, Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angew Chem Int Ed Engl* **56**, 11354-11359 (2017).
37. W. M. Babinchak *et al.*, Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. *Nature Communications* **11**, 5574 (2020).
38. A. N. Milin, A. A. Deniz, Reentrant Phase Transitions and Non-Equilibrium Dynamics in Membraneless Organelles. *Biochemistry* **57**, 2470-2477 (2018).
39. W. M. Babinchak *et al.*, The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J Biol Chem* **294**, 6306-6317 (2019).
40. J. A. Riback *et al.*, Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209-214 (2020).
41. Jacob C. Schwartz, X. Wang, Elaine R. Podell, Thomas R. Cech, RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Reports* **5**, 918-925 (2013).
42. J. A. Morin *et al.*, Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nature Physics* **18**, 271-276 (2022).
43. A. W. Fritsch *et al.*, Local thermodynamics govern formation and dissolution of Caenorhabditis elegans P granule condensates. *Proceedings of the National Academy of Sciences* **118**, e2102772118 (2021).
44. S. Sridharan *et al.*, Systematic discovery of biomolecular condensate-specific protein phosphorylation. *Nature Chemical Biology* **18**, 1104-1114 (2022).
45. G. Krainer *et al.*, Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nature Communications* **12**, 1085 (2021).
46. A. Bremer *et al.*, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat Chem* **14**, 196-207 (2022).
47. J.-M. Choi, F. Dar, R. V. Pappu, LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology* **15**, e1007028 (2019).
48. J. E. Henninger *et al.*, RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224 (2021).

Supplemental Figures:

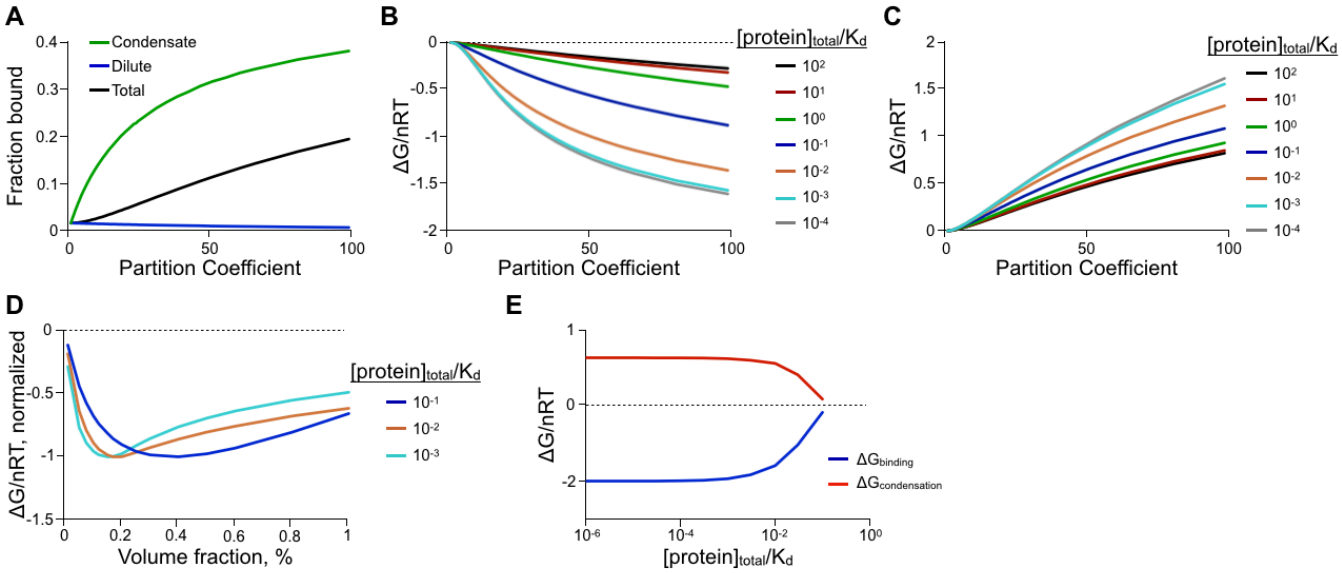

Figure S1: Properties governing condensation in the binding model. A) Fraction of bound protein for homotypic interactions in the condensate, dilute region and total in the system as a function of partition coefficient. **B)** Change in free energy arising from binding for homotypic interactions as a function of the partition coefficient for various $[protein]_{total}/K_d$ values. **C)** Change in free energy arising from entropic cost for homotypic interactions as function of the partition coefficient for various $[protein]_{total}/K_d$ values. **D)** The impact of volume fraction on free energy minimum of condensation for homotypic interactions at different $[protein]_{total}/K_d$ values. **E)** The free energy of binding (blue) and entropic cost (red) for condensation as a function of the $[protein]_{total}/K_d$ ratio. Volume fraction = 1% for all results except (D).

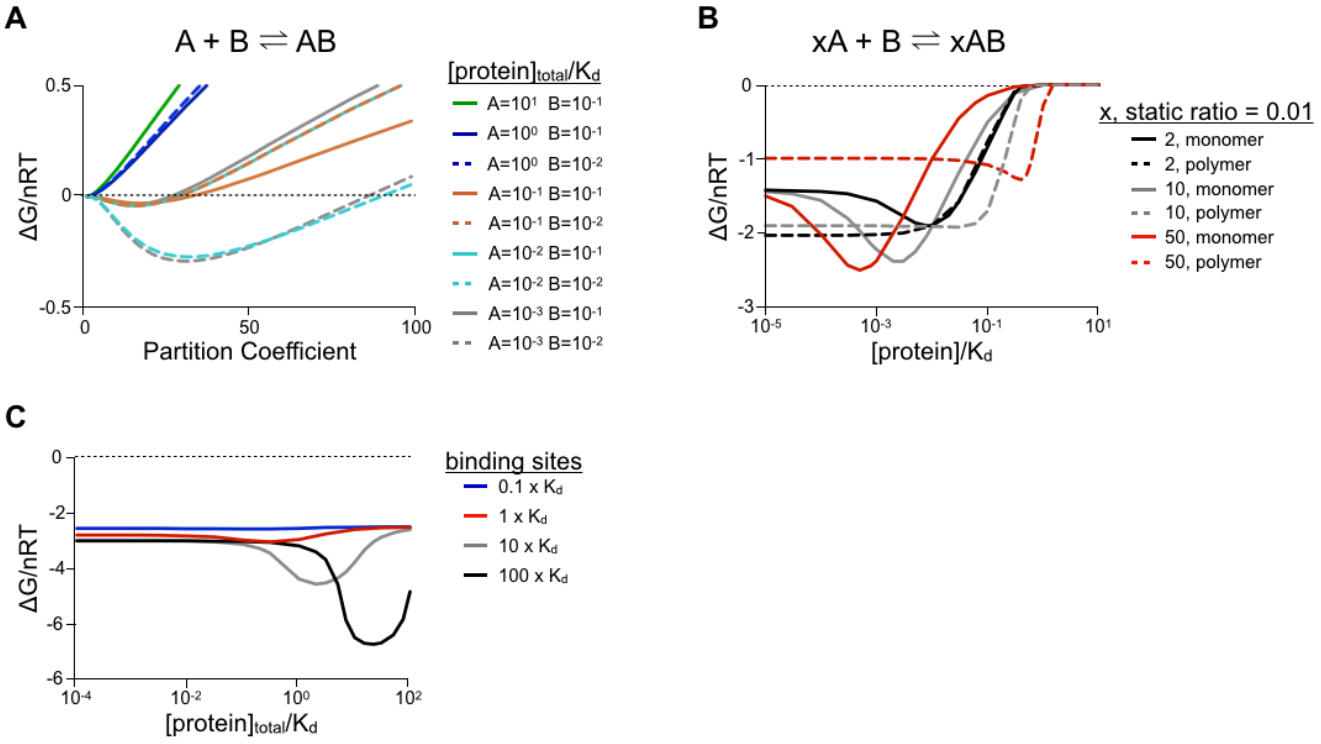

Figure S2: Results from the binding model for heterotypic interactions. A) Total energy as a function of partition coefficient for a heterotypic interaction at multiple $[protein]_{total}/K_d$ ratios for both binding molecules, A and B. Partition coefficient is calculated from the most abundant molecule, here binding affinity remained constant. If either species exists above a ratio of 10^{-1} condensation is not energetically favorable. **B)** Total energy as a function of $[protein]_{total}/K_d$ ratio for heterotypic binding where multiple A molecules can bind a single B molecule. Shown are binding model results for multiple binding ratios. The static $[protein]_{total}/K_d$ ratio remained at 0.01, while the dynamic ratio is plotted - for 2, monomer (A), the plot is a function total energy vs. polymer (B) $[protein]_{total}/K_d$ values. **C)** Total energy as a function of client $[protein]_{total}/K_d$ ratio at a partition coefficient = 20 for heterotypic interactions where one component (scaffold) is not subject to entropic cost of condensation, e.g. it is condensed due to interactions with other molecules at other binding sites. Shown are plots for different amounts of scaffold concentrations in the condensate relative to the dissociation ratio of the client binding interaction. Volume fraction = 1% for all results.