Importance of hierarchical interactions in concentrated protein systems

by

Kamal Bhandari

M.S., Tribhuvan University, Nepal, 2010

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Abstract

Protein-protein interactions play important roles in signal transduction, chemical catalysis, cell metabolism, membrane transport, muscle contraction and more. While many of these interactions can be understood as two-body interactions, other important properties can emerge when large numbers of molecules interact together. In this dissertation, we investigate three such systems. We find that a common characteristic in the emergent behavior is interactions occurring on multiple length and energy scales.

Biomolecular condensates appear throughout the cell serving a wide variety of functions. Many condensates form by interactions between multivalent proteins, which produce phase-separated networks with liquidlike properties. We use a model system of poly-SUMO and poly-SIM proteins in which the basic unit of assembly is a zipperlike filament due to the interaction of alternating poly-SUMO and poly-SIM molecules. These filaments have defects of unsatisfied bonds that allow crosslinking for both the formation of a 3D network and the recruitment of additional molecules, called clients. We observe a nonmonotonic client binding response that can be tuned independently by the client valence and binding energy. These results show how the interactions within liquid states can be disordered yet still contain structural features that provide functionality to the condensate.

The nucleolus is a large membrane-less organelle in the nucleus which is responsible for the assembly of ribosomes. We have developed models to understand how interactions among major proteins assist pre-ribosomal assembly and proper rRNA folding in the nucleolus. The first of these models describes the electrostatic interaction between two major nucleolar proteins (NPM1 and SURF6-N) as a driving force for the phase separation in the nucleolus. The second model describes the role of NPM1 as a chaperone in the rRNA folding process. We have developed analytic, numerical, and computational methods which confirm that NPM1 lowers the zipper barrier to unzip misfolded rRNA.

Protein aggregation is a major problem in drug formulation for the pharmaceutical industry. Antibody molecules form elongated complexes due the interaction between their domains. These complexes entangle with each other causing a sharp rise in viscosity of solution. To reduce the production cost, we need to predict the viscous behavior of molecules in the early stages of drug development. We develop a method to use dilute solution measurements to predict the antibody viscosity. We translate the strength of self-association measured by dilute solution experiments (ACSINS, DLS) into the binding affinity of molecules. A theoretical model based on entanglement physics predicts viscosity from these binding affinities. Predictions are in good agreement with the test set of molecules with a few outliers. Using the theoretical model, we predict the physical mechanism of these outliers, and we have proposed solutions to account for these mechanisms in a refined method.

Importance of hierarchical interactions in concentrated protein systems

by

Kamal Bhandari

M.S., Tribhuvan University, Nepal, 2010

_____

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Dr. Jeremy D. Schmit

Copyright

Abstract

Protein-protein interactions play important roles in signal transduction, chemical catalysis, cell metabolism, membrane transport, muscle contraction and more. While many of these interactions can be understood as two-body interactions, other important properties can emerge when large numbers of molecules interact together. In this dissertation, we investigate three such systems. We find that a common characteristic in the emergent behavior is interactions occurring on multiple length and energy scales.

Biomolecular condensates appear throughout the cell serving a wide variety of functions. Many condensates form by interactions between multivalent proteins, which produce phase-separated networks with liquidlike properties. We use a model system of poly-SUMO and poly-SIM proteins in which the basic unit of assembly is a zipperlike filament due to the interaction of alternating poly-SUMO and poly-SIM molecules. These filaments have defects of unsatisfied bonds that allow crosslinking for both the formation of a 3D network and the recruitment of additional molecules, called clients. We observe a nonmonotonic client binding response that can be tuned independently by the client valence and binding energy. These results show how the interactions within liquid states can be disordered yet still contain structural features that provide functionality to the condensate.

The nucleolus is a large membrane-less organelle in the nucleus which is responsible for the assembly of ribosomes. We have developed models to understand how interactions among major proteins assist pre-ribosomal assembly and proper rRNA folding in the nucleolus. The first of these models describes the electrostatic interaction between two major nucleolar proteins (NPM1 and SURF6-N) as a driving force for the phase separation in the nucleolus. The second model describes the role of NPM1 as a chaperone in the rRNA folding process. We have developed analytic, numerical, and computational methods which confirm that NPM1 lowers the zipper barrier to unzip misfolded rRNA.

Protein aggregation is a major problem in drug formulation for the pharmaceutical industry. Antibody molecules form elongated complexes due the interaction between their domains. These complexes entangle with each other causing a sharp rise in viscosity of solution. To reduce the production cost, we need to predict the viscous behavior of molecules in the early stages of drug development. We develop a method to use dilute solution measurements to predict the antibody viscosity. We translate the strength of self-association measured by dilute solution experiments (ACSINS, DLS) into the binding affinity of molecules. A theoretical model based on entanglement physics predicts viscosity from these binding affinities. Predictions are in good agreement with the test set of molecules with a few outliers. Using the theoretical model, we predict the physical mechanism of these outliers, and we have proposed solutions to account for these mechanisms in a refined method.

Table of Contents

List of Figures

xix

List of Tables

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Jeremy D. Schmit. His mentoring has always been the source of inspiration which encouraged me to carry out nicely and finish my PhD in smooth way. I acknowledge all his time, encouragement, guidance and patience during my Ph.D. to make this journey very productive. I am thankful to him for all the tireless efforts he has put on me and stimulating discussions that we had together during my Ph.D.

I would like to extend my sincere thanks to my dissertation committee members: Prof. Christopher M. Sorensen, Prof. Paul E. Smith, Prof. Bret N. Flanders, Prof. Fariba Fateh for valuable suggestions and insightful comments during my research. I would like to remember all the faculties and staff members of physics department for their support and encouragement during my Ph.D.

I would like to extend my gratitude to Prof. Michael K. Rosen, UT Southwestern medical center, for collaborating with us and insightful suggestions in SUMO/SIM projects. Likewise, I am also grateful to Dr. Joon Huh, AMGEN pharmaceutical company, for collaborating with us and thoughtful discussions in antibody project.

I very much appreciate to all my friends who have supported me during my stay at the Kansas State University. Many thanks to all members of Schmit group: Tien, Tam and Nelson for supports and valuable discussions with me during my Ph.D. I would like to thank to my senior Dr. Yuba Raj Dahal, for his advice and support during the starting days of my Ph.D. journey. Brother Tien! you are so kind, caring and supportive. I wish I would get the friends like you forever.

I would like to remember my special friends: Bipin Lamichhane, Megh Nath Jaisi for moral supports and valuable suggestions in every ups and downs.

I owe my deepest gratitude to my parents for their relentless support and guidance through out my life. I am all the time indebted to them for all their sacrifices. I also would

like to remember my brother and sisters for their continuous encouragement and support during the different periods of my life. My heartfelt thanks go to you my dear wife, my success would not have been possible without the selfless support and nurturing of my wife, "Sushmita". Finally, I would like to remember my little sons "Ahwaan" and "Aabhar" for bringing bundles of joy and happiness in my life.

Chapter 1

Introduction

1.1 Introduction

Inter-molecular interactions are crucial for the existence of life because nearly all processes in biology depends on the intermolecular interactions between macromolecules such as protein, DNA, RNA and lipids[1]. These macromolecules interact with each other by a variety of forces (van der Walls forces, electrostatic forces, hydrophobic interactions, dipole-dipole interaction, etc.). The structure of molecules and the location of chemical groups encode binding specificity and catalytic properties. Biomolecules become functional when they interact with binding partners but their functionality may differ depending on their binding partners and binding locations. The same protein molecule may exhibit different functions when it interacts with different biomolecules. Interactions at the two-body level are well studied and their structure-function properties are well understood. However, it is now clear that many-body interactions are also important in biology.

It has been observed from experiments that some cellular processes may require many biomolecules. Many cellular structures are formed by the spontaneous condensation of biomolecules into liquid droplets, often through liquid-liquid phase separation[2,3]. These condensates may contain hundreds of molecules. The molecules that contribute most strongly to phase separation show high interaction valence and high connectivity to other molecules.

Many body interactions can also lead to several unwanted outcomes. Protein phase separation or aggregation as a result of protein-protein interaction can be toxic and cause diseases [4;5] including Alzheimer's disease, Parkinson's disease , Huntington's disease, eye lens cataract, sickle cell anemia, etc. Even in vitro many-body interactions can be important. In drug formulation, delivery and manufacturing constraints can require very high concentration. At high concentration, protein-protein interactions can change the solution properties, for instance, high concentration antibodies will often have very high viscosity. Protein crystals are also an outcome of protein–protein interaction [6]. Protein aggregation is another example of many-body interactions causing problems in drug formulation for the pharmaceutical industry.

This thesis is about many-body interactions in biomolecular systems. In the concentrated regime, biomolecules can interact with each other by different intermolecular interactions. Here, we focus on two types of many-body systems.

(a) Phase separated condensates: Eukaryotic cells organize biological processes using organelles, many of these are enclosed by membranes but there are also several that do not have membranes. Stress granules, nucleoli, P-bodies, cajal bodies, PML nuclear bodies are some examples of membraneless organelles [7–11]. While these membraneless organelles have been known for decades, only recently has it become clear that they have liquid properties and appear to form by liquid-liquid phase separation [2;3]. Possible functions for membraneless organelles includes sequestration of harmful components, storage of useful biomolecules and speeding up biochemical reactions [3;9;12–15].

Membraneless organelles contain tens to hundreds of different biomolecules of which a few, called scaffolds, are essential for formation of the condensates. Scaffolds also recruit the remaining molecules, called clients, with the total composition providing functionality [7;13]. To explore the physics of membraneless organelles, Rosen and co-workers engineered polySUMO and polySIM proteins. In the SUMO/SIM system, there are two deccavalent scaffolds, polySUMO and polySIM, composed of modules connected by short disordeded linkers of twelve amino acids. In addition, there are low valence SUMO and SIM employed as client molecules. Rosen and co-workers found that an interesting non-monotonic client binding

response as a function of scaffold proportions that we are able to explain with our theory.

The nucleolus is one of the largest and most prominent membraneless organelles in the cell nucleus. The nucleolus has become an attractive system in recent years because it is the site of major cellular processes such as DNA transcription, RNA processing, pre-ribosomal assembly[16–18] etc. The nucleolus has three sub-compartments with known function[19]. DNA transcription to produce rRNA occurs at the central layer, rRNA pre-processing and modification occurs in the middle layer, and the outermost layer assembles ribosomal subunits. The major biomolecules involved in pre-ribosomal assembly are NPM1, SURF6 and rRNA. After assembly in the nucleolus, ribosomal subunits are exported to the cytoplasm where they form a mature ribosome. Ribosomes produce proteins in the form of a linear chain or random coil of polypeptides[20].

(b) Formulation of antibody: Therapeutic anitbodies have been developed to treat several human diseases[21;22] and their importance has grown day by day. Subcutaneous administration is the favored method of drug delivery to patients with chronic diseases for which frequent dosing is necessary. Limited volume available in the subcutaneous space requires high concentrations of antibody drugs, which can lead to the high viscosity[23]. The high viscosity not only increases the injection time but also pain at the injection sites[24]. Hence, high viscosity of antibody formulations is a major problem in drug development. There are many factors that affect the viscosity of protein solution, such as concentration, pH, ionic strength, charge, shear rate, temperature, molecular shape, and protein-protein interactions[25–28]. We need to understand the physical mechanisms of high viscosity to provide guidance in drug development. It is difficult to mitigate high viscosity once a drug is developed, therefore, successful prediction of viscosity at early stages of development would reduce manufacturing costs. There are several existing computational and experimental tools to predict antibody viscosity. Existing computational tools have very low resolution because they use machine learning techniques or empirical methods to predict viscosity. Existing experiments measure the strength of interaction in dilute solution (two-body effects) but have difficulty accounting for many-body interactions that drive viscosity. Here, we develop a method that connects the two-body measurements with many-body viscosity.

Despite the major difference in these systems, a common theme has been emerged. We find that in all cases we study, important properties emerge from the combination of two-body interactions and mony-body interactions. Experiments performed at the scale of two-body interaction cannot explain many-body properties and the experiment at the scale of many-body interaction cannot account for two-body effects. We study here the combined effect of these multiple interactions on three different systems. As a result of the multiple interactions, these system have a hierarchical structure,i.e., strong interactions form molecular complexes and weak interactions drive many-body effects.

1.2   Dissertation outline

In chapter two, we study the assembly of SUMO/SIM condensates, a synthetic system designed to explore the properties of membrane-less organelles.

Chapter three explores how the concurrent pre-ribosomal assembly and rRNA folding process occurs in the nucleolus. In this chapter, we briefly explain how a protein (NPM1), drives the phase separation and acts as the chaperone for rRNA folding.

In chapter four, we discuss how dilute solution measurements can be used as an effective way to predict antibody viscosity at early stages of drug development. We have developed an hypothesis that connects dilute solution measurements (two-body effects) to the viscosity (many-body effects) of the molecules.

In chapter five, we summarize each research topic and identify potential future directions.

Chapter 2

Structure-function properties in disordered condensates

2.1 Introduction

The human body is comprised of trillions of cells. Each cell has numerous organelles which perform various functions. Many of these organelles in eukaryotic cell are not enclosed by a membrane, for instance, Stress granules, Nucleoli, P bodies, Cajal bodies, PML nuclear bodies and so on[7–11]. These cellular bodies are micron-sized assemblies of protein and often RNA found in cytoplasm and nucleoplasm[2;29]. The function of membraneless organelles are still unclear but it is hypothesized that cells produce them in order to concentrate biochemical reactions, to sequester harmful components, to store of useful bio-molecules, and to enhance signaling pathways[3;9;12–15] . Many of them display the liquid-like properties, such as, spherical shape, the ability to flow and they coalesce upon contact with each other[3;29].

These properties along with the observation that they form by the spontaneous condensation of bio-molecules, has led to the suggestion that they are a form of liquid-liquid phase transition[2;3]. Liquid-Liquid phase separation (LLPS) is a thermodynamically driven phenomenon in which a homogeneous solution of molecules separates into two distinct liquid phases, with different solute concentrations. In LLPS, molecules can diffuse within the

dense phases or between the dense and dilutes phases because these phases are in equilibrium. LLPS depends mainly on the temperature, type and concentration of molecules, and interaction among solute and solvent molecules. While these condensates may contain hundreds of different molecules, typically only a small number of molecules with high interaction valence and high connectivity to other molecules contribute strongly to phase separation[30–32]. These are said to have "scaffold-like" properties depending on how strongly they promote phase separation. They are main components essential for structural integrity of cellular bodies. In another words, they are building block of cellular bodies. The remaining molecules, which exhibit "client-like" properties, are recruited through interactions with scaffolds[7;13]. Together, the collection of molecules in a condensate determines its functionality.

Since the molecules driving phase separation are multivalent, polymer-like species, many treatments of condensate formation are based on polymer theories[5;33–35]. In particular, scaffold condensation can be understood as the interaction between attractive "stickers" separated by inert "spacers"[35;36]. These efforts explain universal features of condensates, such as how multivalency can amplify the effect of weak interactions to tune the phase coexistence line to lie within physiologically relevant regions of phase space (e.g. physiological concentrations)[5]. However, condensates have been shown to perform diverse functions, and the evolutionary pressure to optimize these specific functions implies that there will also be non-universal features[3;9;12–15]. This begs the question of how the disordered network of interactions within a liquid structure can affect its properties.

The disorder in biomolecular condensates poses a challenge for structural biology to determine what features of the assembly are functionally relevant. In previous work, our research group showed that analytic theory can be used to deduce structural features in these systems[8]. This approach is analogous to conventional methods like NMR and crystallography in that experimental constraints are used to refine a structural model. A major difference is that the model does not specify spatial coordinates, but instead, describes the network connectivity. This is encoded in a free energy model that uses system-specific order parameters to capture structural features as well as regions that remain disordered.

Here we apply this approach to condensates formed by poly-SUMO and poly-SIM, a
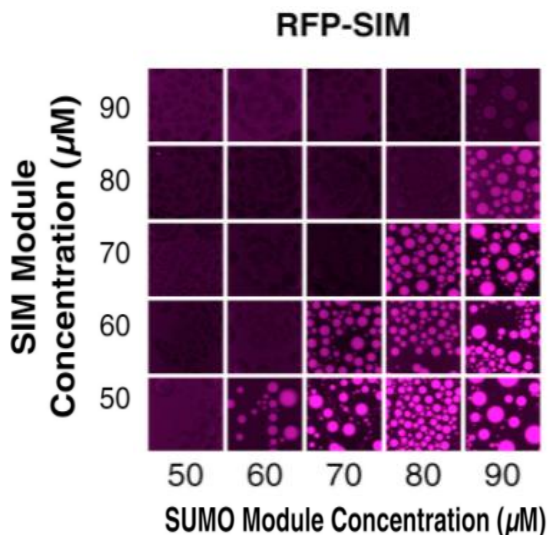
Figure 2.1: *Phase diagram displays client recruitment. The assembly of scaffolds (poly-SUMO/polySIM at different stoichiometric ratios) recruits the client (red fluorescently labeled protein SIM, 100 nM). The client recruitment is observed only on the side of diagonal where its cognate scaffold was in stoichiometric excess in solution. Adapted with permission from Banani et al. Copyright 2016 by the Cell.*

synthetic system composed of repeating units of either SUMO or its binding partner, SIM, that was developed to study the recruitment of client molecules (Fig 2.1)[7]. In SUMO/SIM system, scaffolds of the same kind do not interact and undergo phase separation, so that only mixture of both scaffolds, poly-SUMO and poly-SIM, form condensates[7]. SUMO is the Small Ubiquitin-like Modifier protein which has ~ 100 amino acids and $\sim 12kDa$ molecular weight, and SIM is SUMO Interaction Motif which has ~ 23 amino acids and $\sim 3kDa$ molecular weight. poly-SUMO and poly-SIM both are formed by ten repeats of SUMO protein and SIM protein respectively. In both poly-SUMO and poly-SIM, the modules are separated by a Gly- and Ser-rich short, disordered linker of 12 amino acids. SUMO protein has folded domain while SIM is the intrinsically disordered protein. This system captures the features of the common sticker-and-spacer topology, and the 1:1 binding stoichiometry between SUMO and SIM modules eliminates the ambiguity in identifying the sticker moieties that complicates

7

the study of natural systems. The experimental system also includes constructs containing only 1, 2, or 3 SUMO or SIM repeats, which one considered to be clients of the SIM and SUMO scaffolds respectively.

We find that SUMO/SIM condensates have a zipper-like microstructure that differs from the brush-like microstructure of SPOP/DAXX or the random network of FUS-like proteins[8;35]. This microstructure dictates how client recruitment varies with changes in valence and affinity, and suggests that the composition, and thus function, of a condensate can be changed through modulation of its internal structure independent of its propensity to undergo phase separation. These results show how functionally relevant structures can be embedded within liquid-like disorder.

Based on these results, we propose a general model in which functional properties emerge from hierarchical assembly within a condensate. In this model, strong interactions drive the formation of molecular complexes with functional properties. These complexes then condense into a fluid state via weaker interactions. This combination of strong and weak interactions combines the best features of both interactions modes. Strong interactions provide the structural specificity needed for functional properties to emerge, while weaker interactions allow the formation of a high density state without the kinetic arrest that would accompany assembly driven strictly by strong interactions.

2.2   Methods

2.2.1   Zipper-like filaments allow most modules to form bonds within a sparse network.

Our initial attempts to model SUMO/SIM condensates employed Flory-Huggins theory with three components representing SUMO, SIM, and solvent. Flory-Huggins is a lattice mean-field model representing a maximally disordered system. We refer to Flory-Huggins and similar theories for associative polymers[35–37] as "random network" models, reflecting the fact that each binding site randomly searches the nearby volume for interaction partners,

which results in uncorrelated binding. The Flory-Huggins lattice constant is determined by the dimensions of the macromolecular repeat units. In this case, estimates for the linker radius of gyration and module size yield repeat units on the order of 3 nm. We found that any realistic combination of interaction parameters, which account for excluded volume, binding energies, and spacer entropy, yields lattice occupancies of at least 70%, and usually much greater. With a 3 nm lattice constant, this corresponds to a droplet module concentration on the order of 50 mM, much greater than the observed concentration of ~ 2 mM[7]. To account for the finite valence of the SUMO and SIM modules, we modified our model for SPOP/DAXX assemblies to account for decavalent scaffolds and, again, obtained results consistent with a dense phase module concentration near 50 mM. Electrostatic repulsion was considered as a swelling mechanism and rejected on the grounds that the repeat unit dimensions are greater than the ~ 1 nm Debye screening length, which eliminates long range interactions (short range effects are accounted for in the module binding affinity).
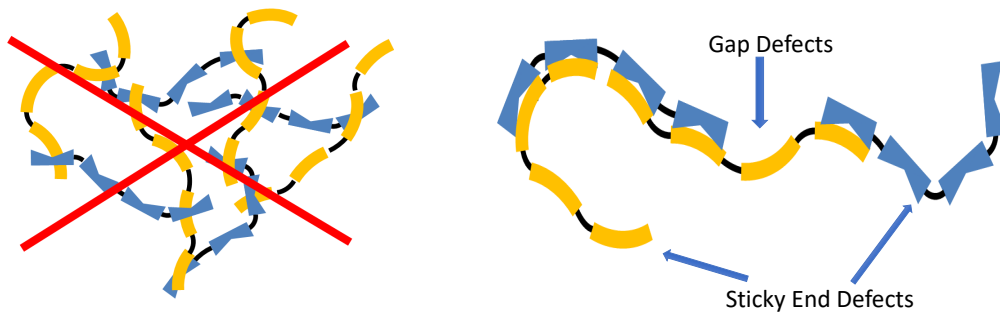
Figure 2.2: *a) Random crosslinked-like structure. b) Zipper-like structure.*

The experimentally measured module concentration implies that the volume per decavalent scaffold is $(2 \text{ nM}/10)^{-1} \simeq (20 \text{ nm})^3$. Since, this is larger than the random walk volume of a scaffold, this separation would allow only a small fraction of intermolecular bonds to be

satisfied if the molecules are randomly oriented. For bond energies strong enough to drive phase separation, on the order of $k_B T$, the penalty for this many unsatisfied bonds is prohibitive. Therefore, we were forced to reject the hypothesis of randomly oriented scaffolds (Fig 2.2a).
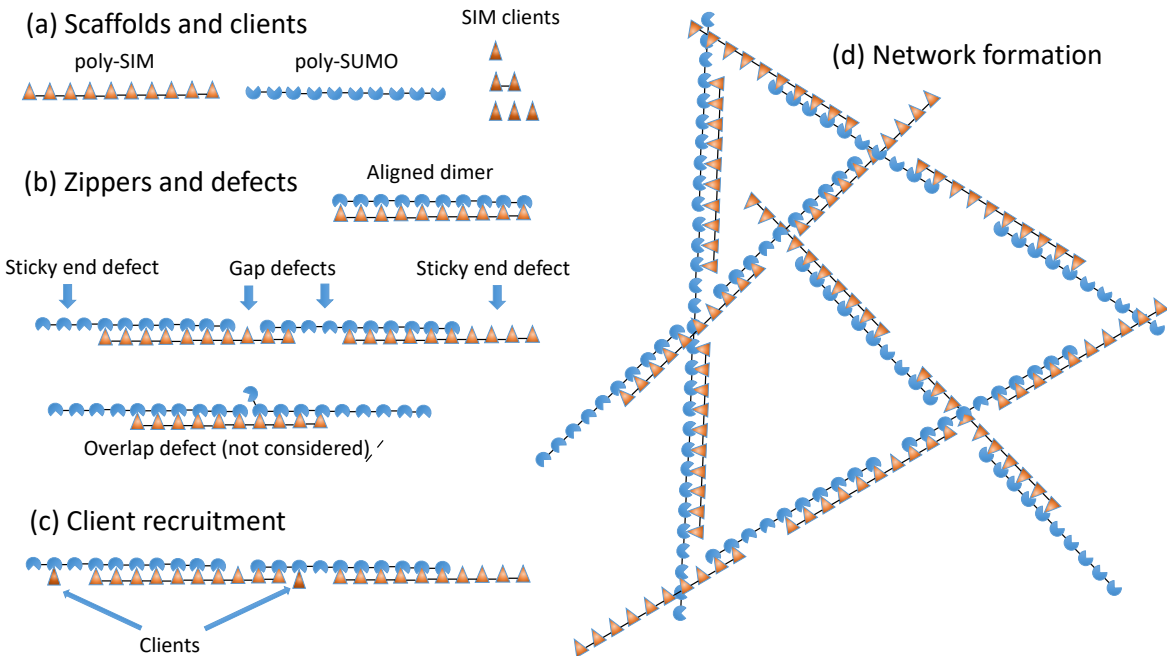
Figure 2.3: *Poly-SUMO and poly-SIM assemble into zipper-like filaments. (a) SUMO/SIM droplets are composed of decavalent scaffolds and clients of valence 1, 2, and 3. (b) Intermolecular bonding is most efficient when the scaffolds align to form zipper-like structures. The zippers have bonding defects including sticky ends and gaps. Overlap defects are also possible but neglected in our calculation. (c) Zipper defects provide binding sites to recruit clients or (d) assemble the zippers into a 3D network. The short linker connecting modules favors consecutive bonds with the same molecule rather than the formation of a random network.*

The alternative to randomly oriented scaffolds is that SUMO/SIM networks satisfy most binding sites by aligning the scaffolds (Fig 2.2b). Therefore, we developed a theory to describe aligned molecules. Fig. 2.3a,b shows a model where the molecules are aligned to form zipper-like structures. By comparing the predictions of this model to the experiments of[7] we can learn the extent to which this "zipper model" describes SUMO/SIM condensates.

Thermal fluctuations require that SUMO/SIM zippers will have defects in the bonding structure that provide recruitment sites for clients as well as enable the 1D filaments to

assemble into a 3D network (Fig. 2.3c,d). Therefore, the properties of the zippers determine the properties of the larger scale droplets. Notably, the defects will have a spacing on the order of the scaffold dimension, consistent with the $(20 \text{ nm})^{-3}$ scaffold concentration. Fig. 2.3b shows two defects considered in this work, gaps and sticky ends, as well as an overlap defect that we neglect. Overlap defects incur a binding energy penalty identical to a gap, but are further suppressed by the steric clash between the two modules competing for the same binding partner. Therefore, we expect, and our results confirm, that sticky ends and gaps are the dominant defect in the zipper structure.

2.2.2 The filament partition function can be calculated with transfer matrices.

Due to the 1D nature of the zippers, their properties can be conveniently calculated using a transfer matrix formalism[38;39]. To illustrate this approach, consider a tetravalent ($v = 4$) system (in subsequent calculations we use $v = 10$ to compare to the experiments of[7]) and define $Q_N(j)$ as the partition function of a filament containing N molecules that terminates with a sticky end of j unpaired sites. Next, we construct a $v - 1$ dimensional vector, $\mathbf{V}(N)$, whose elements are the statistical weights $Q_N(j)$ for sticky ends of each length, j. This vector obeys a recursion relation $\mathbf{V}(N + 1) = \mathbf{M}_i \mathbf{V}(N)$, where the transfer matrix $\mathbf{M}_i$ generates the possible states upon addition of a molecule to the filament

$$
\begin{pmatrix} Q_{N+1}(1) \\ Q_{N+1}(2) \\ Q_{N+1}(3) \end{pmatrix} = \begin{pmatrix} 0 & 0 & z_i e^{-3\epsilon} \\ 0 & z_i e^{-2\epsilon} & z_i e^{-2\epsilon} g_i(1) \\ z_i e^{-\epsilon} & z_i e^{-\epsilon} g_i(1) & z_i e^{-\epsilon} g_i(2) \end{pmatrix} \begin{pmatrix} Q_N(1) \\ Q_N(2) \\ Q_N(3) \end{pmatrix} \tag{2.1}
$$

where ϵ is the SUMO-SIM binding free energy (all energies are expressed in units of $k_B T$) and z_i is the fugacity of the added molecule. Matrix elements below the antidiagonal leave gap defects which can bind clients or form crosslinks with other filaments. This is accounted for by the gap partition function $g_i(m)$, where m is the size of the gap. Since we have

two scaffold molecules, the grand partition function for a filament of $N + 1$ poly-SUMO and $N + 1$ poly-SIM molecules can be generated by alternately applying the matrices for the two species and applying vectors $\mathbf{V}_{L/R}$ to collapse the matrix product into a scalar polynomial (see Appendix). This operation is given by $Q_{N+1} = \mathbf{V}_L(\mathbf{M}_U\mathbf{M}_I)^N\mathbf{V}_R$, where the U/I subscripts denote SUMO/SIM.
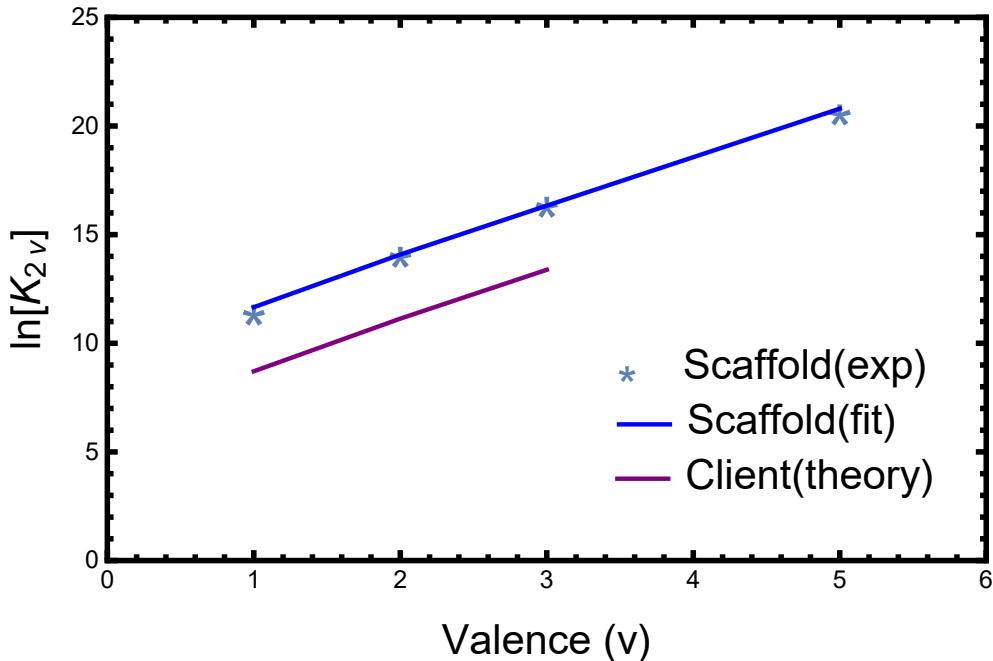
Figure 2.4: *Model parameters are obtained from poly-SUMO/poly-SIM dimerization experiments. Plot of the dimer association constant, K_{2v}, vs. the valence, v. The module binding free energy, ϵ, is obtained from the slope of $\ln K_{2v}$ (blue line), while the intercept provides the reference concentration c_0. Clients (mono-, bi-, or trivalent SIM) have a lower binding affinity than scaffolds in the droplet phase, which we attribute to steric interactions between the network and the fluorescent labels. We account for this with a free energy offset, f_{RFP}, for clients in the dense phase (purple line).*

The most important parameter in Eq. 2.3 is the module binding affinity ϵ, which can be obtained from dilute solution measurements[7] as follows. Consider the dimer association constant between scaffolds of valence v, $K_{2v} = c_{2v}/(c_{SUMO_v}c_{SIM_v}) = c_0^{-1}e^{-\Delta F_2}$ (See Appendix), where c_{2v} is the concentration of scaffold dimer. Here the reference concentration c_0 gives the equilibrium constant appropriate units and enters the matrix formalism through the fugacity $z_i = c_i/c_0$ (valid for dilute solution when the monomer free energy is set to zero). The dimerization free energy is given by the dimer partition function

$e^{-\Delta F_2} = e^{-v\epsilon}(1 + 2e^{\epsilon} + \ldots + 2e^{(v-1)\epsilon})$, where the terms represent the perfectly aligned state and successively larger mis-alignments. Plotting $\ln K_{2v}$ vs. v (Fig. 2.4), we can extract $\epsilon = -2.23\ k_B T$ from the slope and $c_0 = 64\ \mu$M from the intercept. Note that ϵ accounts for both the binding affinity between modules, as well as the entropic cost of constraining the connecting spacers.

To compute solution properties we construct the grand partition function

$$\mathcal{Q} = c_{\mathrm{U}} + c_{\mathrm{I}} + c_{\mathrm{U}}c_{\mathrm{I}}e^{-n\epsilon} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{UL}}(\mathbf{M}_{\mathrm{I}}\mathbf{M}_{\mathrm{U}})^N \mathbf{V}_{\mathrm{UR}} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{IL}}(\mathbf{M}_{\mathrm{U}}\mathbf{M}_{\mathrm{I}})^N \mathbf{V}_{\mathrm{IR}} \qquad (2.2)$$

where the first two terms represent the monomers, the third term represents the perfectly aligned dimer, and the two sums represent filaments starting with SUMO and SIM respectively. The matrices $\mathbf{M}_i$ and end vectors $\mathbf{V}$ are given in the Appendix.

2.2.3 Transfer matrix formalism enumerates all the ways to assemble a 1D filament of alternating poly-SUMO and poly-SIM molecules.

Transfer matrices are the technique to compute the interaction between nearby locations in 1D systems. In Quantum Mechanics, transfer matrices are used to compute 1D wavefront. Here, we apply them to compute the partition function of scaffold chains. In another words, when a scaffold is added in scaffold complexes of order n, transfer matrix accounts all possible ways of n+1 order scaffold assembly formation. Transfer matrix conveys all information about scaffold-scaffold binding, defect sites and scaffold-client binding.

A graphical representation of transfer matrix for poly SUMO (four repeats of SUMO)/poly SIM (four repeats of SIM) is shown in Fig 2.5.

The column vector in the right hand side of Fig.2.5 represents the end state of a scaffold filament which has three possible state overlaps. The red overhang allows further growth. The next blue scaffold can bind to the red overhang in up to three possible ways (shown in last column of transfer matrix Fig.2.5), depending on the size of the red overhang. These

13
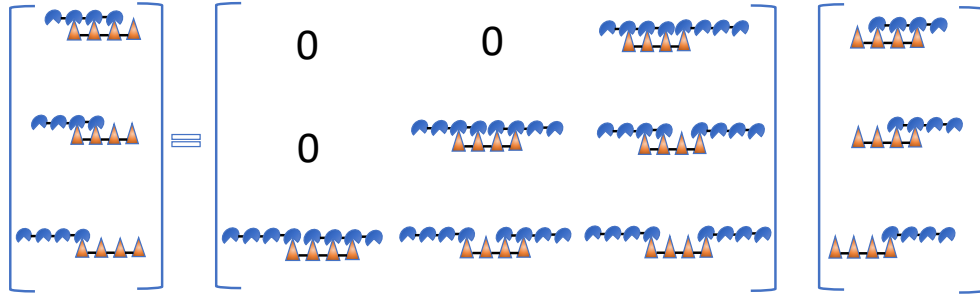
Figure 2.5: *The graphical representation of transfer matrix: Transfer matrix generates the possible states upon the addition of a molecules to the filaments.*

possibilities are depicted in the matrix elements. After the blue molecules is added, there are again three possible states for the ends of the filament, which are depicted by the cartoons in the left vectors.

Figure 2.6: *First Possibility: The added blue scaffold may occupy all three end defects on the red scaffolds.*

To determine the statistical weight of each molecules addition event, we consider each of the outcomes for a starting overhangs of three units. The first possibility is that each of three sticky end defects are occupied by blue scaffold (Fig.2.6). This does not create gap defect and blue scaffold now has only one new sticky end defect (see Fig.2.6). The corresponding matrix elements is given by $z_i \exp^{-3\epsilon}$.

Figure 2.7: *Second Possibility: The added blue scaffold may occupy two of three end defects on the red scaffolds.*

In the second possibility, just two of three red sticky ends are occupied by blue scaffold (Fig.2.7). Here, assembly create one gap defect and blue scaffold currently has two new sticky end defects (see Fig2.7). The corresponding matrix element is given by $z_i \exp^{-2\epsilon} g_i(1)$ where $g_i(1)$ is the gap partition function for a single defect in the filament which we discuss in next section.

Figure 2.8: *Third Possibility: The added blue scaffold may occupy only a defect on the red scaffolds.*

The final possibility, only one of the three red sticky end defects are occupied by the blue scaffold (Fig.2.8). This outcome creates two gap defects and the blue scaffold now has

three new sticky end defects (see Fig.2.8). The corresponding matrix element is given by $z_i \exp^{-\epsilon} g_i(2)$ where $g_i(2)$ is the gap partition function for two defects in the filament (see next section).

These three cases give last column components of transfer matrix. Similarly, we can write all components of transfer matrix. The complete transfer matrix for $\nu = 4$ is given by

$$
\begin{pmatrix} Q_{N+1}(1) \\ Q_{N+1}(2) \\ Q_{N+1}(3) \end{pmatrix} = \begin{pmatrix} 0 & 0 & z_i e^{-3\epsilon} \\ 0 & z_i e^{-2\epsilon} & z_i e^{-2\epsilon} g_i(1) \\ z_i e^{-\epsilon} & z_i e^{-\epsilon} g_i(1) & z_i e^{-\epsilon} g_i(2) \end{pmatrix} \begin{pmatrix} Q_N(1) \\ Q_N(2) \\ Q_N(3) \end{pmatrix} \tag{2.3}
$$

Since we are examining the client binding behavior with an arrangement of polySUMO and polySIM, both have ten repeats of scaffolds, we need transfer matrix for poly SUMO (ten repeats of SUMO)/poly SIM (ten repeats of SIM) system which is shown in Appendix.

2.2.4 Clients can bind to filament defects.

We next examined how the structural properties of scaffold assemblies in the droplet and bulk phases influence partitioning of mono-, di-, and trivalent SIM molecules employed as clients by Banani et al.[7]. In particular, we sought to understand the curious observation that high-valence clients show non-monotonic partitioning behavior as the SUMO/SIM scaffold ratio changes[7]. To account for the recruitment of clients to the droplet we add client binding to the gap partition function. The form of the gap partition function depends on the valence and the concentration, c_{cl}, of the clients as follows.

a) **Monovalent Client**: Monovalent clients can occupy any defect in the filament. We consider each monovalent client can occupy these defects independently. In case there are m defects in the filament, total gap partition function for monovalent clients is given by

$$
g(m) = (1 + (c_{cl}/c_0)e^{-(\epsilon + f_{RFP})})^m \tag{2.4}
$$
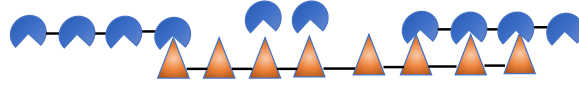
Figure 2.9: *Monovalent Clients occupy the defects independently.*

First terms in gap partition function stands for a defect not occupied by a monovalent client and the second term stands for the deformity occupied by monovalent client. While the binding energies in Eq. 2.4 is same as those for the scaffolds due to the same underlying SUMO-SIM interaction, the binding affinity of clients is lower in the droplet than in the bulk[7]. We attribute this to the entropy cost of confining the bulky RFP fluorescent tag within the droplet network. This steric penalty is included in our model with the factor f_{RFP}. A least squares fit of the client partitioning yields $f_{\mathrm{RFP}} = +3.15\ k_B T$ (Fig. 2.4). f_{RFP} is the only free parameter in our model.

b) **Divalent Client**: Divalent client occupy two defects. If there is only one defect in a scaffold, it can not be occupied by divalent clients. But if there are two or more defects in row, none of them might be occupied or the defects sites could bind any number of divalent client up to the half of the defects. If a scaffold has m defects, they can bind 0, 1, 2...m/2 divalent clients. For a given number of m defects, gap partition function will be

$$g(m) = \sum_{N_{\mathrm{c}}=0}^{m/2} (c_{\mathrm{cl}}/c_0)^{N_{\mathrm{c}}} \binom{m - N_{\mathrm{c}}}{m - 2N_{\mathrm{c}}} e^{-N_{\mathrm{c}}(2\epsilon + f_{\mathrm{RFP}})} \tag{2.5}$$

where N_c is number of divalent clients that occupies defects and $\binom{m-N_c}{m-2N_c}$ is the total number of ways N_c divalent client could be bound with m defects. This degeneracy factor is com-

puted below.

Figure 2.10: *Divalent Client: The binding entropy of the clients in a gap arises from the degeneracy of arranging the clients and unbound sites.*

c) **Trivalent Client**: Trivalent clients require three or more defects in row. If a scaffold has m defects, they can bind 0, 1, 2...m/3 trivalent clients. For a given number of m defects, the gap partition function will be

$$g(m) = \sum_{N_c=0}^{m/3} (c_{cl}/c_0)^{N_c} \binom{m - 2N_c}{m - 3N_c} e^{-N_c(3\epsilon + f_{RFP})} \tag{2.6}$$

where, N_c is number of trivalent clients that occupies defects and $\binom{m-N_c}{m-2N_c}$ is the total number of ways N_c trivalent client could be bound with m defects.

In conclusion, each gap site is independent in the case of monovalent clients, which gives a simple binomial with the two terms representing the empty and client-bound states. For multivalent clients we have the more involved process of counting the number of ways to add N_c clients to m sites and summing over the number of clients that will fit in a gap. The binding entropy of clients in a gap arises from the degeneracy of arranging the clients and unbound sites. We can treat the clients and vacancies as two particle types. When N_c clients of valence v bind to a gap of m sites, there are $N_v = m - vN_c$ unoccupied sites. So
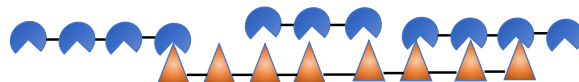
Figure 2.11: *Trivalent Client: As client valence increases the binding entropy of clients in a gap decreases.*

the total number of particles is $N_t = N_c + N_v = m - (v-1)N_c$. The number of permutations is, therefore, the number of ways to select N_v particles from N_t positions. This degeneracy gives the binomial coefficients appearing in Eqs. 2.5 and 2.6.

2.3 Results

2.3.1 The condensed phase has structural features emerging on different length scales.

On length scales larger than the scaffold dimension the droplet resembles a random network of zippers (Fig. 2.3d). On this scale the only structural feature is the mesh size, which is determined by the crosslink density. Since the zipper motif fully satisfies the SUMO and SIM binding sites, the crosslinks can only form at defect sites. This limits the affinity of crosslink bonds, so the crosslinks are relatively weak and rapidly break and reform due to thermal fluctuations. This rapid reconfiguration is responsible for the liquid properties of the dense phase. In contrast, the cooperative binding within the zipper gives them a persistent structure that provides crucial functional characteristics. The zippers have considerable

disorder, due to the many ways that the scaffolds can be aligned, but the essential structural features can be described by two quantities.

Defect density

The first is the defect density, which describes the capacity of the filaments to bind clients or crosslink into a 3D network. Using the substitution $g_i(m) = g^m$ so that g is the statistical weight of an unbound module, the density of unpaired sites per scaffold molecule is given by $\rho_g = N_{tot}^{-1} g \partial \ln \mathcal{Q} / \partial g$. This gives $\rho_g \simeq 0.6$ unbound sites per decavalent scaffold at equimolar scaffold mixtures (Fig. 2.12).
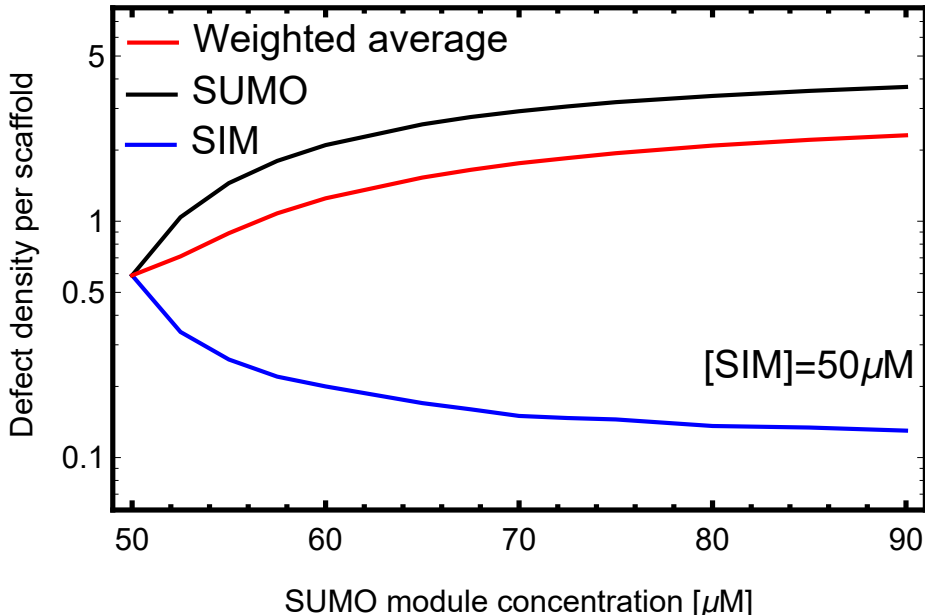
Figure 2.12: *Filament length and defect density depend on stoichiometry. (a) At equal stoichiometry most scaffolds are fully bound resulting in few defects. As the stoichiometric imbalance increases, the number of unbound SUMO modules increases while the number of unbound SIM sites decreases.*

Using this defect density, the zippers can be described in the sticker/spacer framework where the defects are stickers and the spacers are zippers with a length of $10/\rho_g \simeq 17$ modules. The expected radius of gyration of these spacers is $\sim (3 \text{ nm})\sqrt{17} \sim 12\text{nm}$, which is consistent with, but somewhat smaller than, the 20 nm average molecular spacing. If we divide total number of modules between two stickers (~ 17) by the volume corresponding

	SIM Module Concentration(50μM)				
SUMO	0.59	2.10	2.92	3.39	3.70
SIM	0.59	0.2	0.15	0.136	0.13
Weighted	0.59	1.25	1.76	2.09	2.31
	50	**60**	**70**	**80**	**90**

Defect Density

SUMO Module Concentration(μM)

Table 2.1: *Average SUMO/SIM defect density*

to radius of gyration ($\sim$ 12nm), it gives the expected droplet density ($\sim$ 4500μm). The experimental droplet density is about $2500 - 2900\mu$m. We attribute this discrepancy to the fact that not every defect will be involved in a crosslink and zippers will be less flexible than individual scaffolds due to the double linkage and increased charge density, so the spacer radius of gyration will be larger than our 12 nm estimate.

Total average length of zipper

The second useful quantity to describe the zipper structure is the average number of scaffolds in a zipper. Here, we calculate the length of chain/assembly as the function of scaffold module concentration. Since related exploratory work[7] has done in certain scaffold concentration range, polySUMO and polySIM sfaffold module vary from $50\mu M$ to $90\mu M$ in our calculations.

Calculation: Total grand partition function for zipper like structure is:

$$\mathcal{Q} = c_{\mathrm{U}} + c_{\mathrm{I}} + c_{\mathrm{U}}c_{\mathrm{I}}e^{-n\epsilon} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{UL}}(\mathbf{M}_{\mathrm{I}}\mathbf{M}_{\mathrm{U}})^N\mathbf{V}_{\mathrm{UR}} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{IL}}(\mathbf{M}_{\mathrm{U}}\mathbf{M}_{\mathrm{I}})^N\mathbf{V}_{\mathrm{IR}} \quad (2.7)$$

The average number of SUMO molecules per assembly is given by

$$N_U = C_U \frac{\partial \ln Q}{\partial C_U} \quad (2.8)$$

The average number of SIM molecules per assembly is given by

$$N_I = C_I \frac{\partial \ln Q}{\partial C_I} \tag{2.9}$$

Total assembly size is given by $N_{\text{tot}} = \langle N_{\text{I}} + N_{\text{U}} \rangle$.
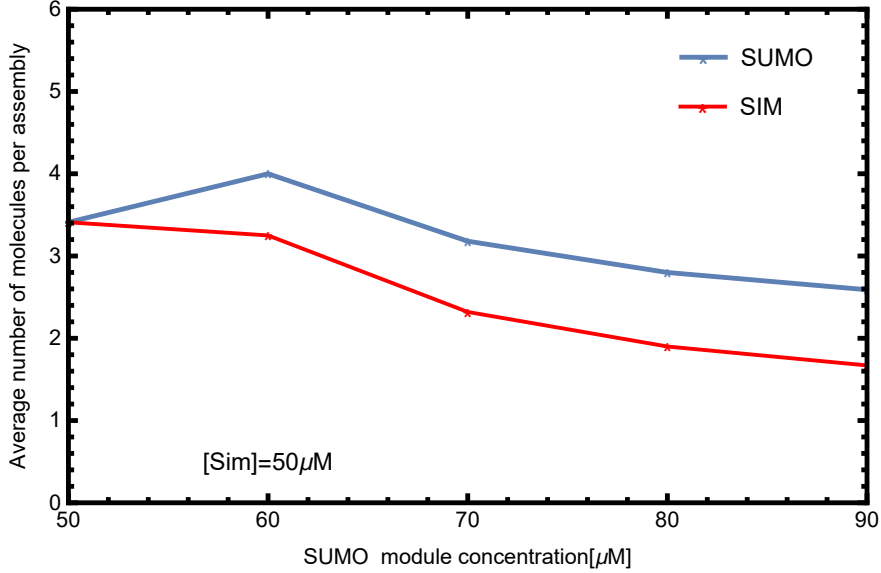
Figure 2.13: *Scaffold stoichiometry determines the average number of SUMO/SIM scaffolds in the assembly.*

Fig.2.13 and Fig.2.14 give average number of SUMO/SIM molecules and total average length of an assembly for fixed SIM module concentration 50 μM and SUMO module concentration ranging from 50 μM-90 μM. For equal module concentration(50 μM SUMO/50 μM SIM), average number of molecules for both scaffold per assembly are equal and high i.e. chain length is long because both scaffold are sufficient to bind each other. There is no restriction to make chain through scaffold proportion. They make assembly longer up to the point where binding interaction is sufficient to pay the entropy cost of recruiting additional scaffolds. But as we go off slightly from equal module mixing (increase SUMO module concentration keeping SIM module fixed), average number of SUMO molecules per assembly increases because excess module quickly finds one of the sticky ends but average number of SIM molecules per assembly decreases. At higher scaffolds inequality, lower concentration SIM scaffolds have to bind more highly concentrated SUMO scaffolds which is possible only
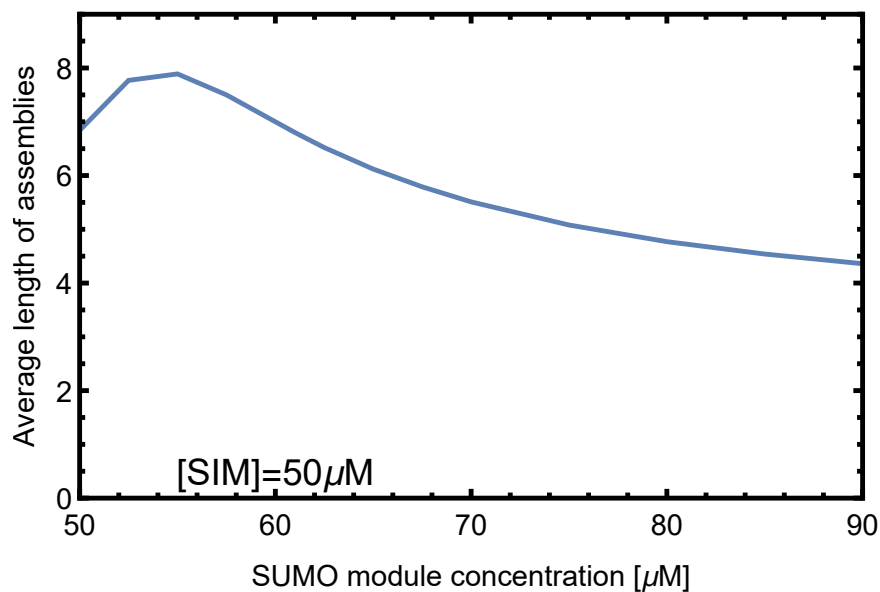
Figure 2.14:   *The average length of filaments is a non-monotonic function of scaffold stoichiometry. A small excess of one scaffold leads to an increase in the filament length because unpaired scaffolds are available to stabilize sticky ends resulting from mis-aligned states. Larger stoichiometric mismatches result in a decline in filament length, which provides more sticky ends to bind the excess scaffold.*

when the chains become more shorter. Our calculation shows that low concentration scaffolds can bind more of the high concentration scaffolds when chains become shorter. For example, four red scaffolds can bind five blue scaffolds in a single chain but same number of red scaffolds can bind eight blue scaffolds when forming many short chain (Fig.2.15).
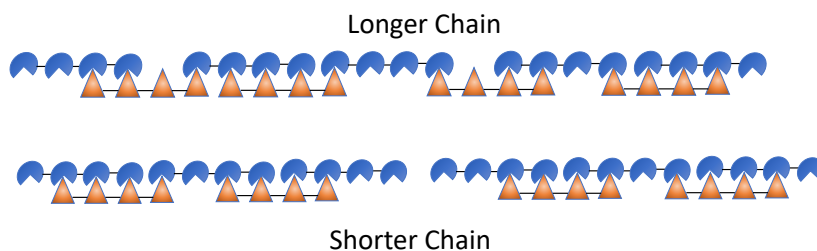
Figure 2.15: *Larger stoichiometric mismatches prefers the shorter chain.*

In conclusion, the most energetically favored state is the perfectly aligned dimer which allows all binding sites to be satisfied (Fig. 2.3b). This competes with the misaligned states,

which have a higher entropy due to the many possible alignments. This is particularly true upon the formation of filaments with $N_{\text{tot}} \geq 3$ because the unpaired sites can be distributed throughout the filament rather than localized at the ends. At sufficiently high concentration the entropic gain of misalignment can overcome the energetic cost of defects allowing the formation of large assemblies.

2.3.2 Asymmetric scaffold stoichiometries promote filaments over perfectly aligned dimers.

Monomer concentration variation

SUMO and SIM scaffolds interact and form filament assemblies. These filaments forms droplet of higher density and leave other scaffolds in the dilute phase. Free monomer concentration varies with different ratios of scaffold mixing. Here we discuss the monomer concentration variation for different scaffold module concentration ratios.

Calculation: The grand partition function for azipper is:

$$\mathcal{Q} = c_{\text{U}} + c_{\text{I}} + c_{\text{U}}c_{\text{I}}e^{-n\epsilon} + \sum_{N=0}^{\infty} \mathbf{V}_{\text{UL}}(\mathbf{M}_{\text{I}}\mathbf{M}_{\text{U}})^N \mathbf{V}_{\text{UR}} + \sum_{N=0}^{\infty} \mathbf{V}_{\text{IL}}(\mathbf{M}_{\text{U}}\mathbf{M}_{\text{I}})^N \mathbf{V}_{\text{IR}} \qquad (2.10)$$

The total module concentraion of SUMO and SIM can be written in the following form from which we can calculate monomer concentration:

$$C_{SUM0}^{Total} = C_U \frac{\partial Q}{\partial C_U} \qquad (2.11)$$

$$C_{SIM}^{Total} = C_I \frac{\partial Q}{\partial C_I} \qquad (2.12)$$

For given value of discrete SUMO/SIM total module concentration, we have calculated the monomer concentration for both SUMO/SIM scaffold (Fig.2.16). We observe from the figure that

(a) If SUMO and SIM have equal module concentration, their monomer concentrations are

equal but it is low ($\sim 10^{-11}$) in comparison to module concentration($\sim 10^{-6}$).

(b) If there is asymmetry in Scaffold mixing, monomer concentration for the lower concentration scaffold decreases ($\sim 10^{-15}$) whereas the higher concentration scaffold's monomer concentration increases ($\sim 10^{-6}$).
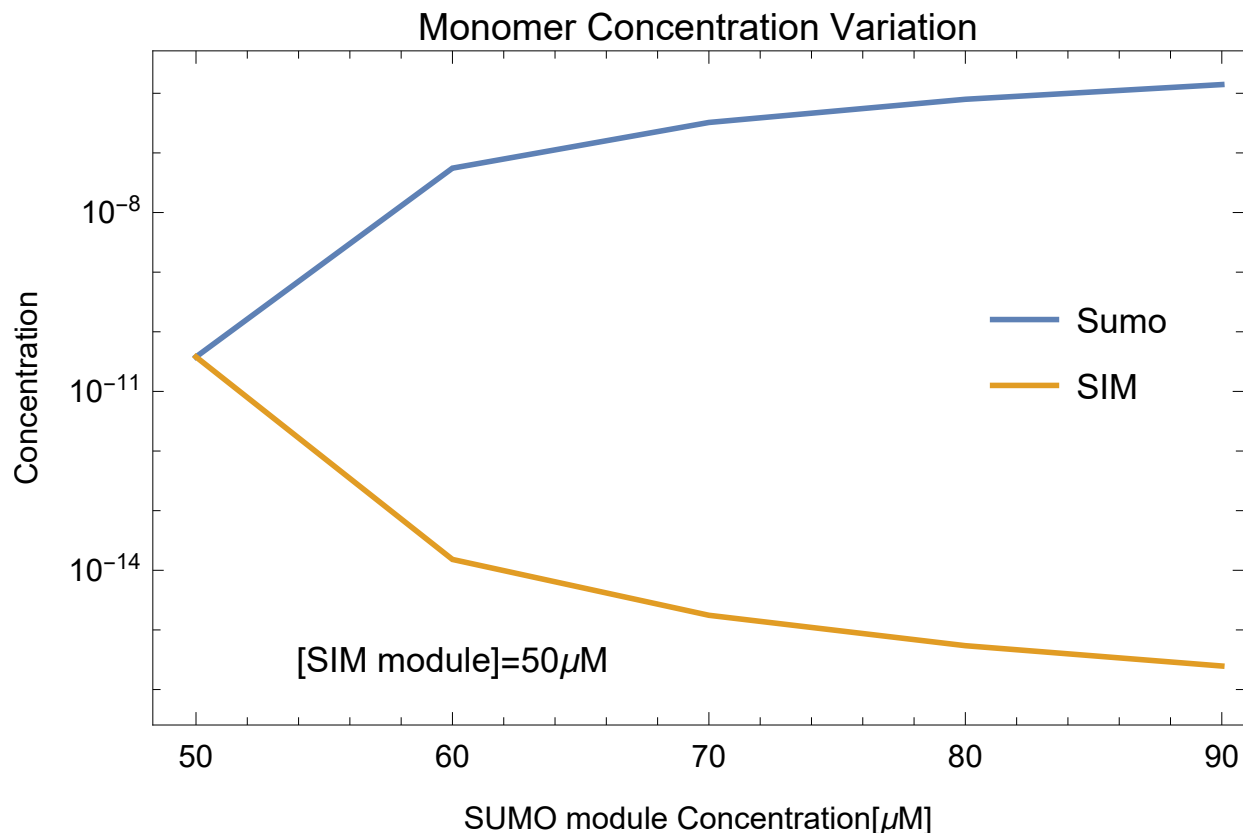
Figure 2.16: *Monomer concentration variation: Excess SUMO scaffold accumulates in dilute phase and depletes the concentration of monomeric SIM scaffolds*

It is observed from Fig.2.17 that chain length is long for equal module concentration. In case there is a slight disparity in scaffold proportion, the abundance monomer quickly binds with the assembly so that the chain length increases but if we increase further inequality in scaffold mixing, lower concentration scaffold tries to bind as much as higher concentration scaffold making chain length shorter and shorter but if there is large inequality in mixing, lower concentrated scaffold are not adequate to bind more higher concentrated scaffold even in the form of shortest assembly so that unbound higher concentration scaffold remains in the solution.
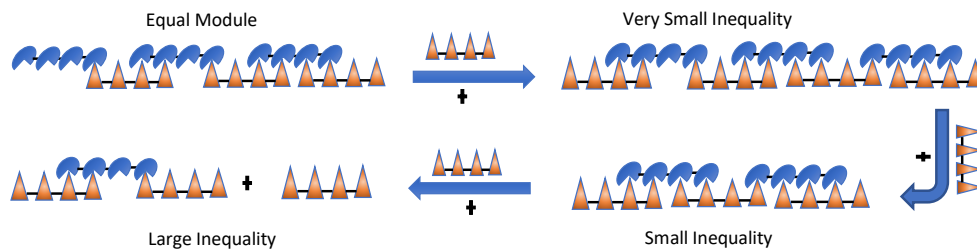
Figure 2.17: *Larger stoichiometric mismatches not only shortens the assembly length but also leaves the large portion of free monomer in the solution.*

Perfectly aligned dimer concentration variation

While writing up the grand partition function of our model, we have referenced that that there is plausible of both scaffold to be perfectly aligned. Numerically it is addressed by $\exp^{-n\epsilon} c_I c_U$.

For a fixed SIM module concentration $[50\mu M]$ and SUMO module ranging from $50\mu M - 90\mu M$, the perfectly aligned dimer concentration variation is shown in Fig.2.18. It is observed that this perfectly aligned dimer concentration is high for equal scaffold mixing but that concentration decreases in case of scaffold inequality. This is because at equal module concentration even N-mers are more favorable than odd N-mers. Among all even N-mers, perfectly aligned dimer is most favorable. But at an unequal module concentration, odd N-mers are more favorable than even N-mers.

Higher N-mers concentration variation

The assemblies of scaffolds contain higher N-mer terms (imperfectly aligned dimer, trimer, tetramer, pentamer, hexamer, heptamer etc). Here we have studied their concentration variation in both cases of equal and unequal scaffold mixing.

Fig. 2.19 shows the higher N-mers concentration for fixed SIM concentration $[50\mu M]$
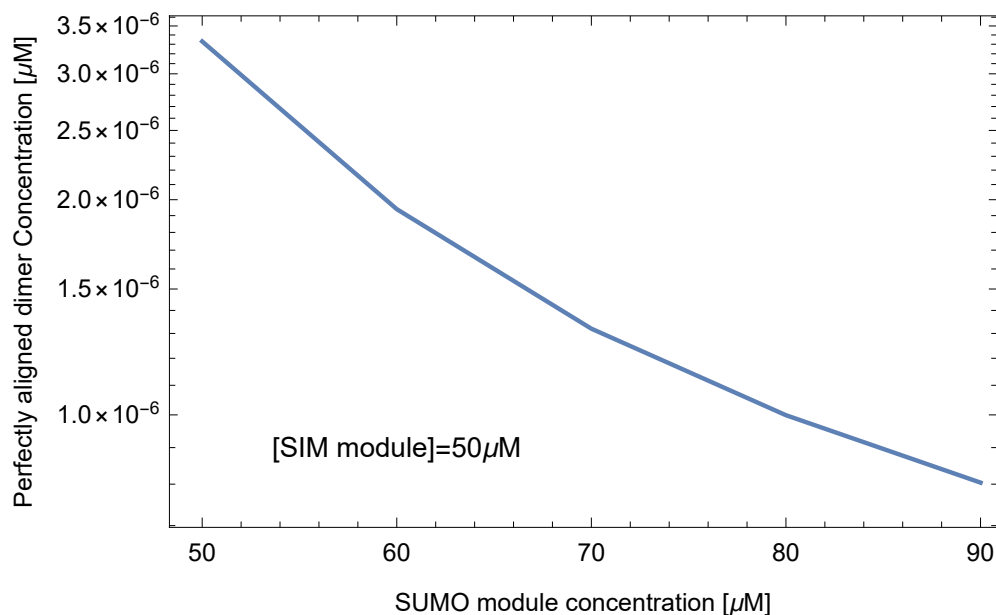
Figure 2.18: *Perfectly aligned dimer concentration variation: equal module mixing favors the perfectly aligned dimer whereas stoichiometric mismatches depletes their concentration.*
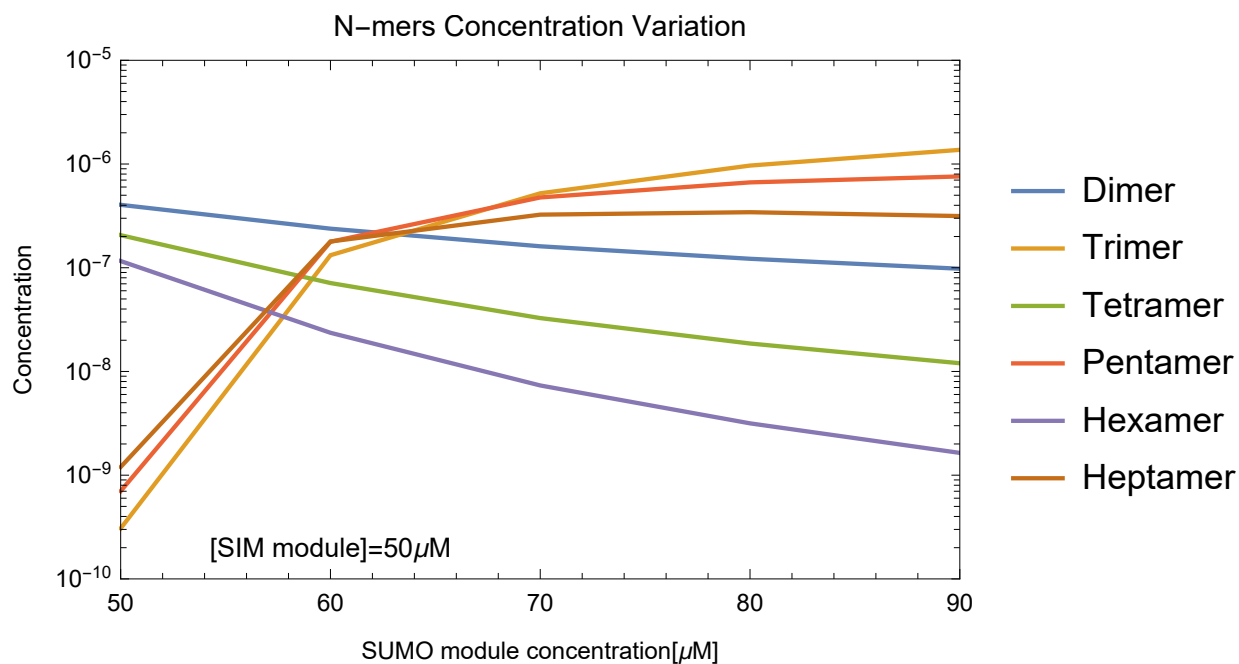
Figure 2.19: *Nmers concentration variation: filament with equal number of SUMO and SIM scaffolds are favored at symmetric mixing but unequal stoichiometries favors odd length filaments.*

and variable SUMO concentraion (ranging from $50\mu M - 90\mu M$). For equal module mixing [SUMO[$50\mu M$] and SIM[$50\mu M$]], even N-mers are more favorable than odd N-mers in light of the fact that more binding interaction is achieved. We have to keep in mind that total free energy of system not only depends on the free energy due to the binding interaction but also on free energy due to entropy. Among even N-mers, perfectly aligned dimer is most favorable because maximum binding interaction is accomplished. The hierarchy of remaining even N-mers depends of their length. If length increases, they are less favorable because they have to pay more entropy cost to sustain their length.

At unequal module concentration, odd N-mers are more favorable than even N-mers because odd N-mers get additional free energy due to entropic arrangement than even N-mers. There are excess monomers in the solution at unequal module concentration so that excess monomers bind with assembly without paying any entropy cost. The hierarchy of odd N-mers relies upon scaffold module proportion, i.e., scaffold module proportion determines the favorable odd N-mers size with which free energy of the system is least. It is observed that small sized odd n-mers are more favored as we go to large scaffold inequality.

Studying all these concentration mentioned above, it is concluded that the combined behavior of the defect density and the filament length can be understood by looking at the populations of different filament lengths (Fig. 2.19). At equal stoichiometries, the solution is dominated by perfectly aligned dimers resulting in a small average complex size. This is similar to the "magic number" effect seen in the simulations of EPYC1 and Rubisco in which phase separation is inhibited when the concentration and valence of scaffolds are matched[40;41]. At unequal stoichiometries the energetic penalty for misalignments is reduced because excess scaffolds are available to bind one of the sticky ends, which facilitates the formation of longer complexes. As the stoichiometry mismatch increases, the system becomes dominated by odd-length complexes which are initially large before trimer (2:1 stoichiometry) states dominate at large stiochiometric mismatches. The decrease in filament size at large stoichiometric asymmetry can be understood as an abundance of filaments that have two sticky ends of the same type, which prevents them from joining into larger assemblies. The shortening of filaments causes an increase in sticky ends and, therefore, an increase in the

28

concentration of defect sites (Fig. 2.12a). The defect density increases rapidly at first, then slows at SUMO module concentrations above 70 μM as the 2:1 trimer begins to dominate.

2.3.3 Asymmetry in scaffold stoichiometry can only be accommodated at the filament ends.

Our matrix formalism does not allow for direct calculation of the assembly of filaments into a 3D network within the droplet. However, we can compute the ratio of SUMO to SIM scaffold in the droplet by assuming that the monomer and perfectly aligned dimer remain in the bulk solution, $\mathcal{Q}_{\text{dilute}} = c_U + c_I + c_U c_I e^{-v\epsilon}$, while all other complexes are in the droplet, $\mathcal{Q}_{\text{droplet}} = \mathcal{Q} - \mathcal{Q}_{\text{dilute}}$. The calculated stoichiometry is compared to the experimentally measured droplet composition in Fig. 2.20a. The droplet SUMO/SIM ratio has downward curvature indicating that the droplet is unable to accommodate the stoichiometry mismatch of the bulk solution. This is a consequence of the filamentous structure because the interior of the filament is constrained to a 1:1 stoichiometry, so stoichiometric imbalance can only be accommodated at the filament ends. The filament calculation has somewhat more curvature than the experimental ratio (Fig. 2.20a, solid line) suggesting that the approximation of purely 1D assembly breaks down when the SUMO/SIM ratio is greater than ~ 1.4. The cause of the failure emerges upon comparing the monomer scaffold concentration (Fig. 2.16) to the binding affinity of the modules, $K_a^1 = 10^5$ M^{-1}[7]. This comparison reveals that when the SUMO concentration exceeds 80 μM the monomer concentration is sufficiently high that scaffold binding to gap defects cannot be neglected. A simple correction, allowing monovalent binding to the gap defects (see Appendix), resolves the discrepancy (Fig. 2.20a, dashed line). This calculation provides a valuable consistency check for our model. It shows that below 80:50 SUMO/SIM ratios the 1D zipper model provides an excellent representation of the droplet. However, above this ratio the 1D model will underestimate the SUMO scaffold content.

A related calculation is to compare the scaffold partition coefficient (PC), defined as the concentration in the dense phase to the concentration in the dilute phase. The frac-
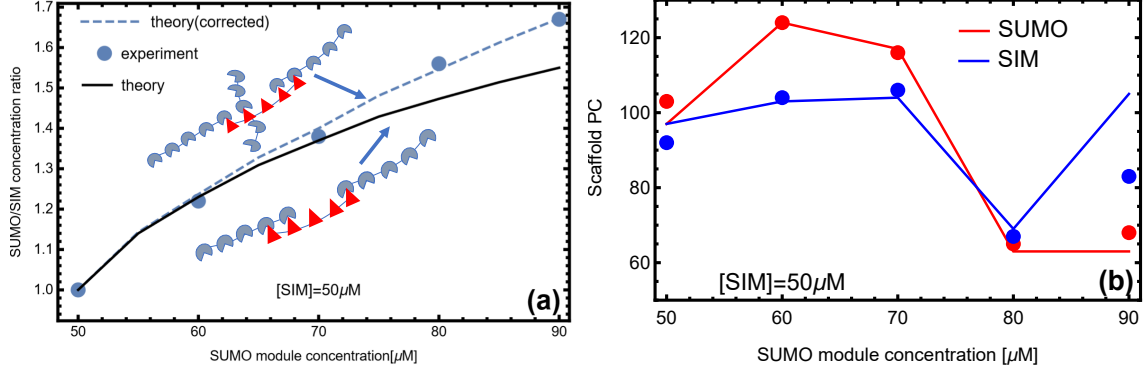
Figure 2.20: *Model captures the droplet scaffold composition observed in experiment. (a) The ratio of poly-SUMO to poly-SIM (N_U/N_I) calculated from our theory agrees well with the droplet stoichiometry in experiments. At high stoichiometric mismatches the approximation of purely 1D filaments breaks down because the concentration of free scaffolds is high enough to allow binding in the gaps. A correction accounting for monovalent scaffold-gap binding (inset cartoon) resolves the discrepancy with experiment (dotted line). (b) Small stoichiometric mismatches promote increased scaffold accumulation in the droplet phase, but the trend reverses as the average filament size drops. The discrepancy at 90 μM can be explained by a breakdown of the 1D approximation as depicted in the cartoons of panel (a). In all panels lines indicate theory and circles denote the experiments of Banani et al.*

tional composition in each phase is $N_i/(N_U + N_I)$, where $i =$ U or I. In order to compare with experiments we weight this quantity by the experimentally determined total module concentration in each phase. Fig. 2.20b plots the quantities

$$\mathrm{PC}_i = \frac{(\frac{N_i}{N_U+N_I})_{\text{droplet}} * [C_{\text{total}}^{\text{droplet}}]}{(\frac{N_i}{N_U+N_I})_{\text{bulk}} * [C_{\text{total}}^{\text{bulk}}]} \tag{2.13}$$

where $[C_{\text{total}}^{\text{bulk}}]$ and $[C_{\text{total}}^{\text{droplet}}]$ are the total module concentration in bulk and droplet phase respectively[7]. The initial rise in the scaffold partitioning is due to the increase in filament lengths, but the trend reverses with the accumulation of smaller filaments at large stoichiometric asymmetry. The excellent agreement between theory and experiment *without free parameters* provides strong evidence supporting the zipper structural model. Notably, the partition coefficients in Fig. 2.20b underestimate SUMO recruitment at 90 μM, which is consistent with the breakdown of the 1D model observed in Fig. 2.20a,b.

2.3.4 Filament defects compete with soluble scaffold molecules for client binding.

The condensate formation due to the assembly of scaffolds might have two sorts of deformities. These defects are either end defects or gap defects. Clients are bound in these defects. Not all clients bind in defects, it depends on their binding affinity. In this segment we will sort out the extent of clients dividing into condensate and bulk.

Client partition coefficient is defined as the ratio of the concentration of client into the droplets to the concentration of client in bulk.

Mathematically it is represented as

$$\text{PC} = \frac{\text{Concentration of client in droplets}}{\text{Concentration of client in bulk}} \tag{2.14}$$

The schematic representation of PC is
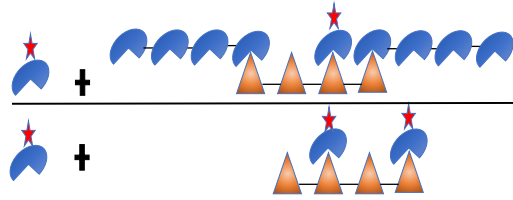
Figure 2.21: *Schematic represenation of PC*

Total concentration of client in condensates is the sum of concentration of free clients in

condensate and concentration of client bound in the defects. Similarly, total concentrations of clients in the bulk is the sum of concentration of free client in bulk and the concentration of clients bound with scaffolds remaining in the bulk.

$$\text{PC}_{\text{cl}} = \frac{[C_{\text{client}}^{\text{free}}]_{\text{droplet}} + [C_{\text{client}}^{\text{bound}}]_{\text{droplet}}}{[C_{\text{client}}^{\text{free}}]_{\text{bulk}} + [C_{\text{client}}^{\text{bound}}]_{\text{bulk}}} \tag{2.15}$$

where $[C_{\text{client}}^{\text{free}}]_{\text{droplet}} = [C_{\text{client}}^{\text{free}}]_{\text{bulk}}$ are the free client concentrations in the droplet and bulk, respectively, and $[C_{\text{client}}^{\text{bound}}]_{\text{bulk}}$ and $[C_{\text{client}}^{\text{bound}}]_{\text{droplet}}$ are the bound client concentration in the bulk and droplet phase, respectively. Here, we have assumed that the monomer and perfectly aligned dimer remain in solution/bulk state whereas the imperfectly aligned dimer and all other longer complexes are in the droplets.

The total grand partition function of bulk/solution state is given by

$$Q_{solution} = Q_{mono} + Q_{dimer}$$

where first and second term are partition function for monomer state and perfectly aligned dimer state respectively.

Similarly, grand partition function for assembly state is total grand partition function minus grand partition function of solution i.e.

$$Q_{assembly} = Q_{total} - Q_{solution}$$

The average number of bound clients per scaffold is computed as $\rho_{\text{c}} = N_{\text{tot}}^{-1} c_{\text{cl}} \partial \ln \mathcal{Q}_{\text{droplet}} / \partial c_{\text{cl}}$. The bound client concentration is obtained by multiplying ρ_{c} times the experimentally determined scaffold concentration[7] which is shown in Table 2.2. Similarly, the bound client concentration in the dilute phase $[C_{\text{client}}^{\text{bound}}]_{\text{bulk}}$ can be calculated from the total number of clients per scaffold in the dilute phase multiplied by the corresponding scaffold monomer concentration in the bulk phase.

Note that the concentrations c_{U}, c_{I}, and c_{cl} appearing in the partition functions all

SIM Module Concentration(50μM)				
2479.6	2720.2	2885.1	2856.5	2889
50	**60**	**70**	**80**	**90**

SUMO Module Concentration(μM)

Table 2.2: *Experimental SUMO module concentration in droplets*

represent the concentration of molecules that have not formed intermolecular bonds. These values must be determined by numerically solving the equations $C_{\text{tot}} = c_i \partial \mathcal{Q} / \partial c_i$ for scaffolds and $C_{\text{tot}} = c_{\text{cl}} + \phi C_{\text{droplet}}^{\text{bound}} + (1 - \phi) C_{\text{bulk}}^{\text{bound}}$ for the client. Here C_{tot} is the total (bound and unbound) concentration of the relevant molecule and ϕ is the volume fraction occupied by the droplet phase. The droplet volume fraction can be determined from the experimental values[7] of the dilute and droplet protein concentrations using $C_{\text{tot}} = \phi C_{\text{droplet}} + (1 - \phi) C_{\text{dilute}}$.

The client partitioning provides further evidence for our structural model (Fig. 2.25). The 1D calculation captures the relative peak heights and locations, neither of which are sensitive to the free parameter. With the aid of a free parameter, the calculation quantitatively agrees with experiments for SUMO concentrations less than 80 μM. Above this concentration, the calculation systematically underestimates client recruitment, which again is consistent with increased SUMO scaffold content as the 1D approximation fails.

Monovalent client

We have calculated the monovalent client (SIM_1) partition coefficient for different scaffold concentrations shown in Fig. 2.22. It has been observed from Fig. 2.22 that,

1) PC is low for equimolar mixtures of scaffolds.

2) PC increases for small excess of the complementary scaffold and then begins to diminish gradually at higher scaffold inequality.

Our calculation closely agrees with experimental PC[7]. While analyzing PC curve (Fig. 2.22), we need to see the role of free client and bound client in the condensate and solution. At equal module concentration, PC is low since most chains are even N-mers, they do not have
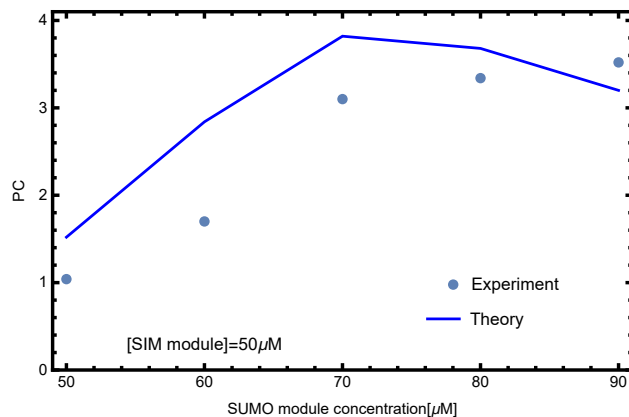
Figure 2.22: *Monovalent SIM client partition coefficient: Non-monotonic client partitioning appears even though weak binding affinity of monovalent SIM clients results in slow PC change.*

enough sticky end defects in the assembly. Clients are mainly bound in gap defects but their number are also tiny due to low monovalent client binding affinity. The number of clients that are bound with scaffolds in solution are also small due to small binding affinity and low free scaffold concentration. Here, the free client concentration are dominant in the PC calculation which are equal for both the condensate and bulk. That is why PC is low (~ 1). In the case of small inequality of scaffold, chain length increases which opens up more sticky end defects where clients can bind so the PC increases. If we go to very large inequality, we will observe that PC starts to diminish. In this case, client density in the droplet turns out to be practically steady in light of the fact that shortest possible chain length has been reached. Therefore, they can not open more sticky end defects but scaffold concentration at solution increases consistently.

Divalent client

We have calculated divalent SIM client PC for a fixed SIM module concentration [$50\mu M$] and SUMO module concentration ranging from $50\mu M$-$90\mu M$ and compared with experimental PC[7], they look close to each other [Fig. 2.23]. At equal module concentration, PC is low likewise in monovalent client. In case of small inequality, odd N-mers opens up more sticky end defects so that the PC increases. If we observe beyond intermediate inequality, PC starts
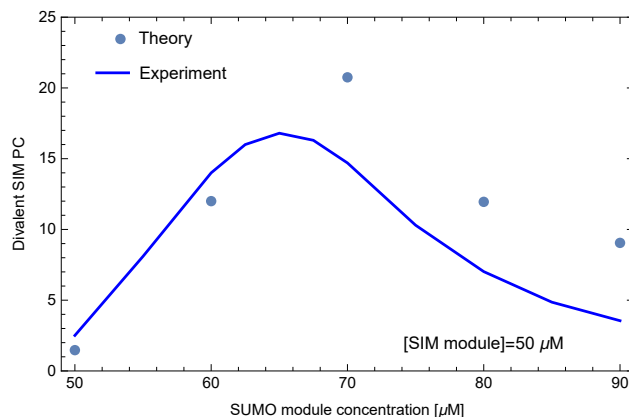
Figure 2.23: *Divalent SIM client partition coefficient: The magnitude of client PC elevates with valence (binding affinity) of clients.*

to decrease even if there are a large number of sticky end defects. In this case, there are a large volume of scaffold monomers in solution so that client not only binds with scaffold assembly but also with monomers in solution. The non-monotonic patterns begins at an earlier concentrations for divalent client compared to monovalent clients.

Trivalent client

We have calculated trivalent client (SIM_3) partition coefficient for different scaffold concentrations shown in Fig. 2.24.
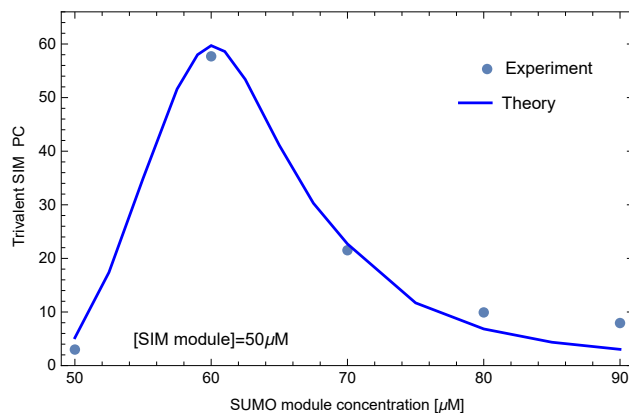
Figure 2.24: *Trivalent SIM client partition coefficient: Increase in client valence not only increases PC magnitude but also the location of PC peak.*

Likewise in monovalent and divalent client, we observe the non-monotonic client parti-

tioning behavior but the non-monotonic trend starts at an earlier concentration for trivalent client than that for divalent and monovalent clients. We find that the maximum value of PC increases with client valence because client binding affinity increases with client valence.
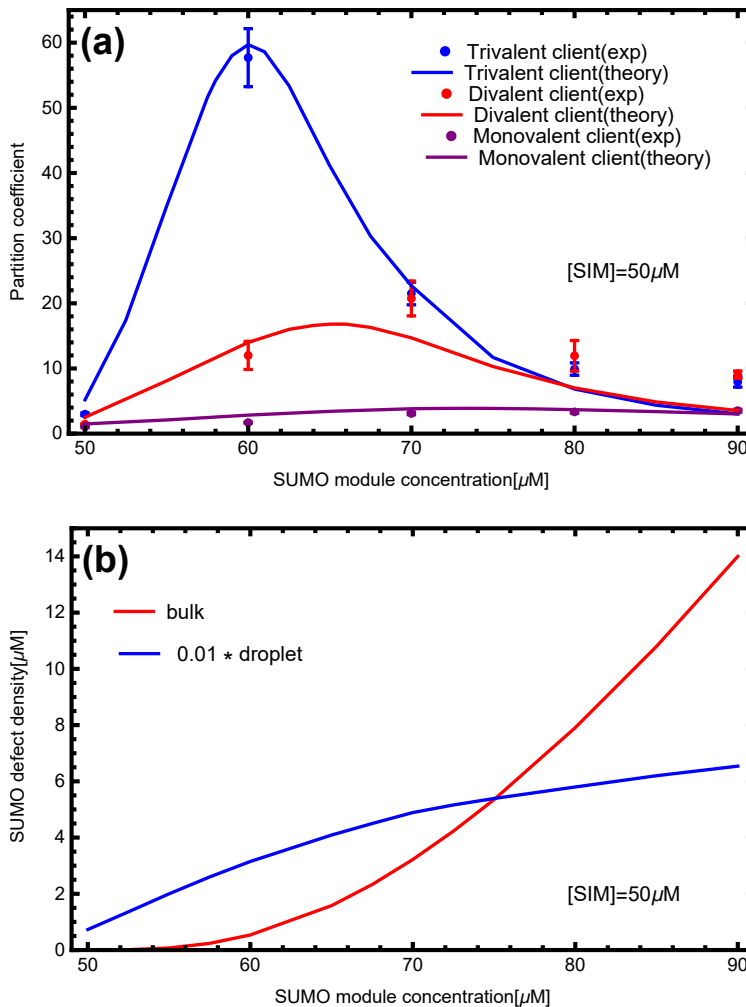
Figure 2.25: *The scaffold composition that optimizes client recruitment depends on the client valence. (a) The transfer matrix theory (lines) captures the shift in the experimental (dots) partition coefficient peak as the client valence increases. The high affinity of trivalent clients is more sensitive to both the appearance of defects in the droplet and the presence of free scaffold in the bulk. Experimental data from Banani et al. (b) Concentration of unbound SUMO modules in the dense and dilute phases. Excess scaffolds, and associated defect sites, are initially bound in the droplet as shown by the blue curve (given by ρ_g times the droplet poly-SUMO concentration). However, above 70 μM additional poly-SUMO accumulates primarily in the bulk phase, which has a defect density of $10 c_U$ (red). The scaling factor applied to the droplet curve is approximately equal to the ~ 0.01 volume fraction of the droplets. Therefore, these curves approximately represent the number of defects sites in each phase.*

If we analyze these three client PCs together, the most striking feature of the PC is the non-monotonic dependence on the poly-SUMO concentration (Fig. 2.25a). This can now be understood as follows. Small excesses of poly-SUMO scaffold are readily incorporated in the droplet, which increases the PC by providing more sites for SIM clients to bind. However, as previously discussed, the 1D filaments are limited in their ability to accommodate unequal stoichiometry. As the filament ends become saturated with the abundant scaffold species, excess scaffold monomers are forced to accumulate in the dilute phase (Fig. 2.16). At SUMO concentrations above 75 μM the number of unpaired SUMO modules in the dilute phase exceeds those in the dense phase (Fig. 2.25b). These free scaffolds compete for clients, which increases client concentration in the dilute phase and lowers the PC.

The client binding response is sensitive to the valence and binding affinity of the clients. Increasing the client valence increases the maximum value of the PC and shifts the peak toward equal scaffold stoichiometry. The increased affinity of clients through higher valence allows them to bind more readily to the sticky ends that appear with a small excess of complementary scaffold. However, the larger valence of these clients means that there is comparatively lower configuration entropy when binding to small defects. Therefore, when free scaffolds appear in the dilute phase, the favorable entropy provided by 10 consecutive binding sites overwhelms gap binding. Conversely, low valence clients have lower affinity, requiring a higher defect density for appreciable recruitment. But, they do not feel a confinement entropy in small defects, so they are sensitive only to the total number of available binding sites in the two phases (Fig. 2.25).

2.3.5   Network structure can provide binding specificity.

Up until this point, we have studied the behavior of clients partitioning into the droplets for different valence clients having their fixed binding affinities. In this section, we have considered variable client binding affinity to study how the scaffold network acts with different valence clients having variable binding affinities. Not surprisingly, increasing the client's specific affinity enhances client recruitment and increases the PC (Fig. 2.26a,b). Fig. 2.26c

37
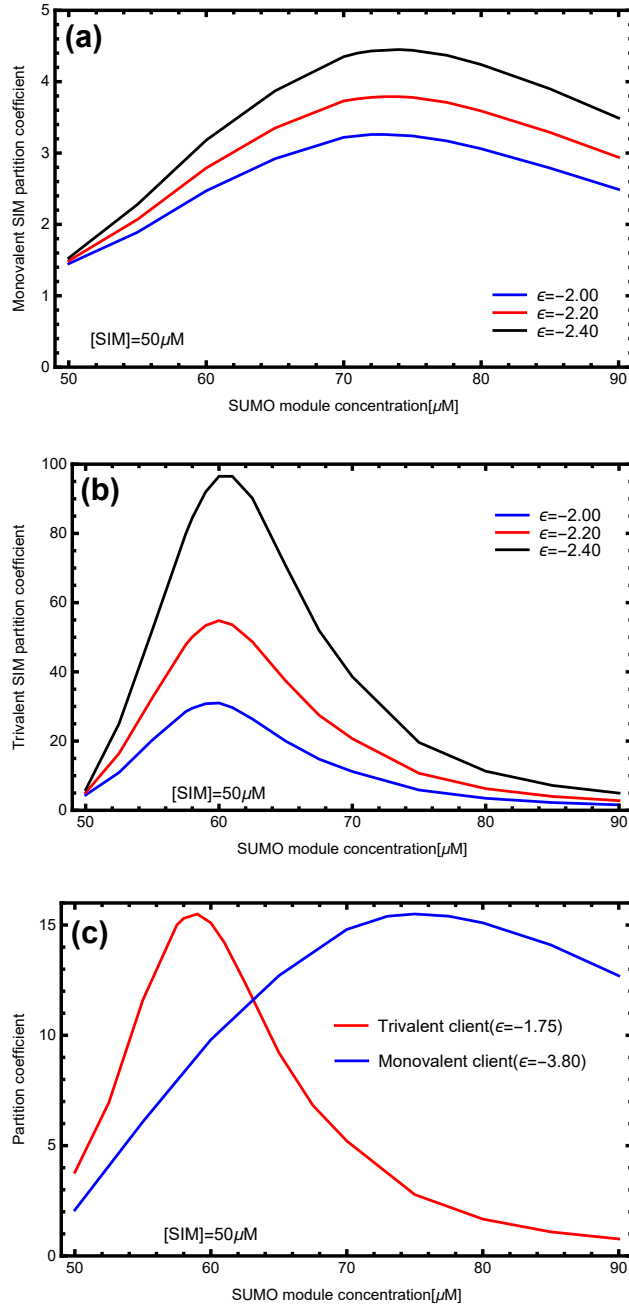
Figure 2.26: *Increasing the client module binding affinity enhances client recruitment with minimal effect on the location of the peak. This provides separate mechanisms to tune the location and magnitude of client recruitment. (a) Monovalent SIM and (b) Trivalent SIM at different module binding affinities. The different effects of client valence and affinity on the PC curve allow the network to switch between the recruitment of different clients. (c) With tuned client affinities, it is possible for the network to selectively recruit either the monovalent or trivalent client in different regimes of parameter space.*

shows PC curves for a monovalent and trivalent client where the binding affinities have been tuned to have similar peak recruitment. However, because the peaks in the PC curves occur at different scaffold stoichiometries, the network will predominantly recruit either one client or the other at different scaffold ratios. Therefore, the zipper microstructure of the droplet provides independent mechanisms to tune the magnitude (via the affinity) and concentration (via the valence) of the maximial PC. These orthogonal methods of tuning client recruitment provide a mechanism by which the structure of the network influences the recruitment of clients, and thus function of the droplets. This means that, despite the liquid characteristics of the condensate, there is enough structure to specifically select between two closely related clients. This is because the zipper structure produces correlations in the location of available binding sites that biases the ensemble of binding states for multivalent clients. These structural correlations are not captured by random network models or the original mass action model of Banani et al.[7]

2.4   Discussion

Interactions stabilizing liquid states impart functionally relevant condensate properties.

Comparing these results to the recently determined microstructure of SPOP/DAXX condensates[8], an interesting trend emerges. Both the SUMO/SIM and SPOP/DAXX systems can be described as sticker and spacer motifs[42,43], yet the underlying networks have very different connectivity (Fig. 2.27). Despite the superficial differences, both systems assemble in a hierarchical manner; strong interactions form molecular complexes and weaker interactions drive the condensation of the complexes. In the SPOP/DAXX system, strong SPOP-DAXX interactions result in the formation of brush-like assemblies that condense into a liquid phase via weaker DAXX-DAXX interactions. However, at higher SPOP-DAXX ratios, DAXX crosslinks SPOP into rigid bundles that associate into a gel state. These distinct states allow the scaffold stoichiometry to serve as a kinetic switch between fluid and arrested states[8]. In contrast, in the poly-SUMO/poly-SIM system the same SUMO-SIM interactions drive
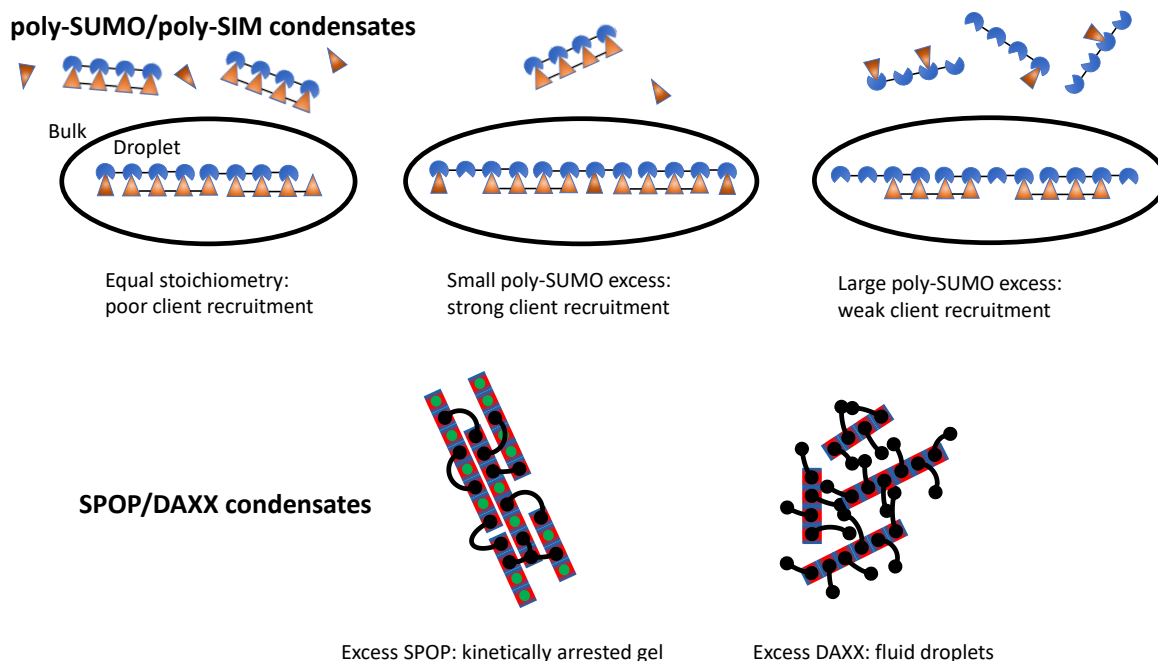
Figure 2.27: *The microscopic connectivity of biomolecule condensates imparts specific properties. While both poly-SUMO/poly-SIM and SPOP/DAXX condensates form by the association of multivalent molecules, the assemblies have very different properties. Poly-SUMO/poly-SIM condensates are composed of linear filaments that provide a client binding response that is sensitive to the scaffold stoichiometry. SPOP/DAXX assemblies contain a "kinetic switch" that allows the system to convert between gel states with arrested dynamics and fluid droplets. Here the black lines represent bivalent DAXX molecules, while the rectangles represent polymerized SPOP rods. Adapted with permission from Schmit et al. Copyright 2020 by the American Chemical Society.*

the strong zipper formation and the weaker network crosslinks, with the presence or lack of cooperative binding providing the two different energy scales[44;45]. Also, the filamentous structure, which is favored by a spacer that is much shorter than the mean intermolecular spacing, allows for a sensitive control over client recruitment. In both systems the assemblies are disordered at the length scales of both the strong and weak interactions. However, the hierarchy gives structure to the fluid in a way that provides functional properties.

We conclude that even within a conserved sticker and spacer framework, differences in the scaffold valence, linker length, rigidity etc. can result in networks with widely different properties. It is inevitable that evolution will exploit these differences to optimize each condensate for its specific biological function. This inspires the question of how to identify

functionally relevant structure within networks that are visually disordered. Here we have shown that theoretical modeling is a powerful tool for this task because parameters like N_{tot} and ρ_{g} emerge naturally to describe structural features in a way that connects to droplet function.

Chapter 3

The concurrent pre-ribosomal assembly and RNA folding process in nucleolus

3.1 Introduction

Nucleus, also called the control center of the cell, regulates cell function. The Nucleus contains the genome, controls the gene expression and performs several functions which are organized into sub-compartments.

There are various membraneless organelles inside the nucleoplasm[46–55], including the nucleolus, nuclear speckles, paraspeckles, PML body, cajal body etc. They are formed through condensation (liquid-liquid phase separation) of molecules into liquid droplets[56–58]. Membraneless organelles play a vital role in cellular process such DNA transcription, RNA processing, protein synthesis etc[59]. The cell may use these organelles to sequester harmful components, to store useful biomolecules, to amplify signal pathway, and to concentrate biomolecules to accelerate the chemical reaction[59–66].

There are several condensates in nucleoplasm among which nucleolus is the biggest. Nucleolus is the site of rDNA transcription, rRNA processing and modification, and pre-

ribosomal assembly[16–18]. It contains DNA, RNA and proteins[19]. The nucleolus has three distinct sub-compartments[14;17;18;46;67] (Fig. 3.1a). The peripheral layer, the Granular component (GC), surrounds the Dense Fibrillar component (DFC) and the Fibrillar center (FC). Different sets of biomolecules inside the nucleus undergo liquid-liquid phase separation into droplets. These droplets of distinct nucleolar proteins have different biophysical properties. In particular, the droplet surface tension is a property that can control the spatial organization of droplets. When two droplets of different surface tension coalesce, the high surface tension phase embeds within the low surface tension phase. This behaviour is responsible for the multilayered structure of the nucleolus[14]. The transcription of rDNA to rRNA happens at the boundary between FC and DFC region[68;69]. The processing of rRNA and their covalent modifications occurs at DFC region. The rRNA folding and pre-ribosomal assembly of ribosomal proteins and rRNA occur in the GC region[17;19;70]. The nucleolus contains many chaperones and proteins which are imported from the cytoplasm. The ribosomal subunits are then exported to cytoplasm for ribosome biogenesis. These mutual exchanges occur continuously between the nucleolus and the cytoplasm.

There are several proteins in granular component of nucleolus but two primary proteins that have important role in cellular processes are Nucleoplasmin1 (NPM1)[46;71] and surfeit locus protein 6 (SURF6)[72]. NPM1 (nucleophosmin) is an essential protein for GC assembly[73;74] and acts as a chaperone for rRNA folding. It has several different roles in biological processes such as DNA repair, chromatin remodeling, embryogenesis[75;76] etc. A schematic diagram for NPM1 is shown in Fig. 3.1c. NPM1 has three domains; two folded domains (oligomerization domain: OD and RNA binding domain: RBD domain) are separated by a disorder linker (DL)[46]. The N-terminal oligomerization domain is responsible for the formation of NPM1 pentamer, which looks like a "starfish"[77] (Fig. 3.1b). The core of a NPM1 pentamer is comprised of the OD and each arm ends with a RBD. The RBD region has net positive charge through which rRNA can crosslink NPM1 pentamers. The DL resembles a block co-polymer with alternate blocks of positive and negative charged amino acids[77;78]. There are two blocks of positive amino acids (B1 and B2) and two blocks of negative amino acids (A1 and A2), with a net negative charge (Fig. 3.1c).
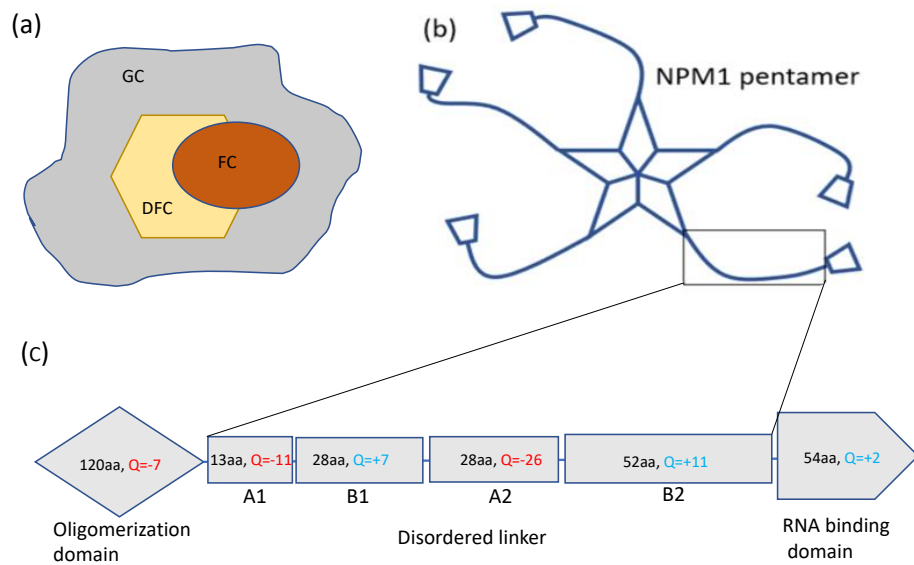
Figure 3.1: *(a) Nucleolus has three sub-compartments. (b) Oligomerization domain drives pentameric NPM1. (c) The block representation of NPM1.*

Surfeit locus protein 6 (SURF6) is a non-ribosomal protein and another one significant for the ribosomal biogenesis. The N-terminal segment of SURF6 (SURF6-N) has 182 residues and drives NPM1 to phase separate[46]. It has net positive charge, with two large positive blocks separated by a small negative charge block so that we can consider SURF6-N as two distinct positively charged blocks (Fig. 3.2).
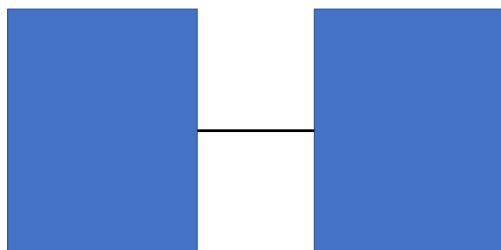
Figure 3.2: *The charge distribution of SURF6-N shows mostly positively charged residues with few negatively charged residues. This charge distribution suggests that we can model it as a pair of positively charged blocks.*

In this chapter, we discuss the GC assembly process and RNA folding process in two different sections.

3.2   Granular component assembly in nucleolus

In vitro experiments have shown that NPM1 undergoes phase separation due to at least three different types of interactions.

1) SURF6-N, a positively charged protein interacts with the negative charge of disordered region of NPM1 molecules[46;67;79].

2) rRNA crosslinks pentameric NPM1 through the RBD regions of NPM1[80].

3) Crowding agents allow NPM1 arms to expand and form intermolecular interaction between positively charged blocks and negatively charged blocks of two distinct NPM1 pentamers[46;67].

3.2.1 Tails of NPM1 pentamer collapse due to interaction between charged blocks

The disordered tails of NPM1 consists of alternating blocks of positive and negative charges. This results in an electrostatic force between alternate blocks which causes contraction (shown in Fig 3.3 A). In the absence of salt, the electrostatic force between alternating charge blocks causes them to contract whereas the high salt screens the attractive force so that they tend to extend[46]. Here, we discuss the free energy of collapse between the alternating blocks.

First, consider the A2 negative charge block and B2 positively charged block (see Fig. 3.1C). The A2 negative charge block has a very high linear charge density (26 negatively charged residue out 28 amino acids) which favors an extended state whereas the B2 positive charge block has a low linear charge density which allows it to adopt random coil states (Fig 3.3B). We have modeled these blocks such that

(i) There are almost no neutral molecules between negatively charged amino acids in A2 block (Fig 3.3 B).

(ii) There are on an average four neutral amino acid between positively charged amino acids in B2 block (Fig 3.3 B).

(iii) Consider a pair of positive and negative charge binding as shown in Fig 3.3C. There are three different possibilities for the binding of the next pair of amino acids. First, the next positive charge could bind with the adjacent negative charge (loop1). Second, the next positive charge may leave the adjacent negative charge unbound and could bind with third negative charge (loop2). Third, the next positive charge and second negative charge may remain unbound so that the next binding event occurs between the third positive charge and third negative charge (loop3). We have neglected other possibilities of binding.

For each of these three cases, we have calculated the electrostatic attractive energy for the bonding between negative and positive charge, electrostatic repulsive energy between positives charges and the repulsive entropy due to constraining the polymer to form a loop. The attractive and repulsive energy are given by

46
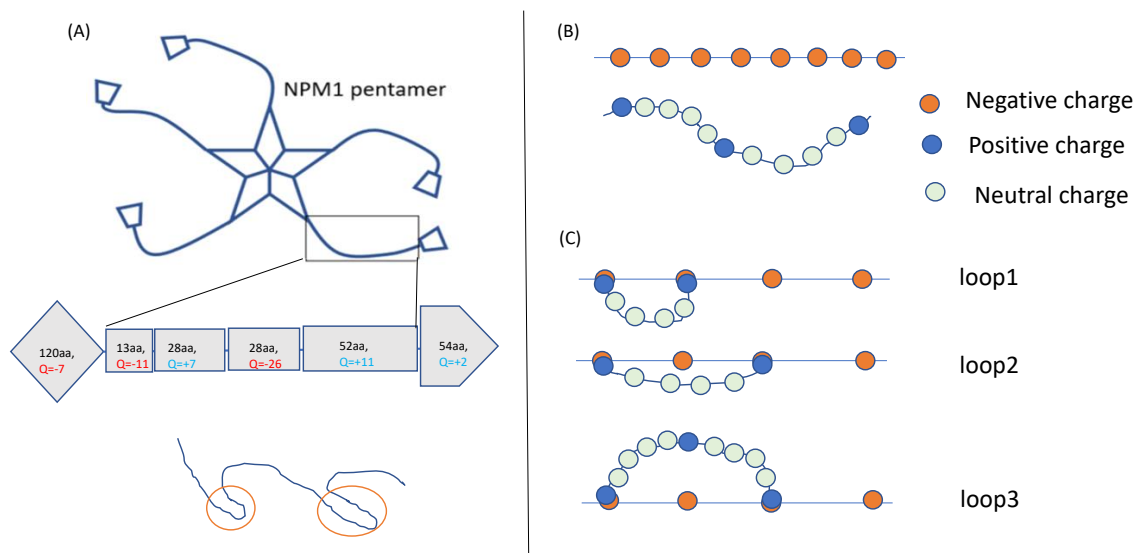
Figure 3.3: *(A) NPM1 tail experiences electrostatic force. (B) NPM1 A2 block resembles like linear charges while B2 block looks like random coil. (C) A2 and B2 block interact in different binding fashion.*

$$F_{A/R} = \mp \frac{e^2}{\epsilon a} = \mp \frac{l_b}{a}$$

where l_b= 0.7 nm is Bjerrum length and a is the distance between charged residues.

For attractive energy, a= 0.2 nm

For repulsive energy, a= 0.45 nm (for loop 1) and a= 0.7 nm (for loop 2 and 3)

The loop formation entropy is given by $\log L$, where L is total number of residues in the loop. The calculation of free energy is shown in Table 3.1.

	Loop1	Loop2	Loop3
Entropy ($K_{\mathrm{B}}T$)	$\log 6 = 1.79$	$\log 6 = 1.79$	$\log 11 = 2.39$
Attractive Energy ($K_{\mathrm{B}}T$)	-3.5	-3.5	-3.5
Repulsive Energy ($K_{\mathrm{B}}T$)	1.5	1	1
Total Free Energy ($K_{\mathrm{B}}T$)	-0.2	-0.7	-0.1
Probability	28 %	47 %	25 %

Table 3.1: *The free energy for each of the states shown in Fig. 3.4C representing the interaction between adjacent charged blocks on NPM1 tail.*

Now, we write the partition function as,

$$Q = \sum_{loop} \exp^{-F_{total}} \tag{3.1}$$

Probability of each loop state is given by,

$$P = \frac{\exp^{-F_{total}}}{\sum_{loop} \exp^{-F_{total}}} \tag{3.2}$$

In loop1 and loop3, there are equal number of positive and negative charges. But loop2 has an excess unbound negative charge and the probability of loop2 is 50%. Based on these probabilities, we expect that the total eleven positive residues in B2 block will cover around sixteen of the negative residues in the A2 block. The total free energy of binding between the blocks is given by: $< E >= -N \log Q = -17 K_{\mathrm{B}}T$. Hence, the average energy per positive charge is $\sim -1.5 K_{\mathrm{B}}T$.

A similar calculation can be done for SURF6-N molecules interacting with the A2 block of NPM1 shown in Table 3.2 are the energies of the loop states . The free energy of binding between NPM1 and SURF6-N is given by: $< E >= -N \log Q = -21 K_\mathrm{B} T$ and, hence, the average energy per charge binding between NPM1 and SURF6-N is $\sim -1 K_\mathrm{B} T$. This value $\sim -1 K_\mathrm{B} T$ is lower than average energy per binding $\sim -1.5 K_\mathrm{B} T$ in NPM1 tail collapse which is due to the fact that SURF6-N has lower charge density than the B2 positive block.

	Loop1	Loop2	Loop3
Entropy $(K_\mathrm{B}T)$	$\log 10 = 2.3$	$\log 10 = 2.3$	$\log 19 = 2.94$
Attractive Energy $(K_\mathrm{B}T)$	-3.5	-3.5	-3.5
Repulsive Energy $(K_\mathrm{B}T)$	1.5	1	1
Total Free Energy $(K_\mathrm{B}T)$	0.3	-0.2	0.44
Probability	28 %	47 %	25 %

Table 3.2: *Calculation of different energies existing in each loop during NPM1 and SURF6-N interaction.*

As we discussed earlier, the eleven positive charges in B2 block will cover only sixteen of the negatively charged residues in the A2 block. This allows the positively charged SURF6-N to interact with the remaining ~ 10 negative charges in the A2 block in such a way that SURF6-N makes a bridge between tails of distinct NPM1 pentamers as shown in Fig. 3.4. Since the charge distribution in two halves of SURF6-N is roughly symmetric, the first half can interact with one pentamer whereas the second half interacts with the tail of distinct pentamer. This forms the crosslinks between NPM1 that drives the phase separation.

3.2.2   NPM1 and SURF6-N drive phase separation

In our model, the NPM1 pentamers are treated as a lattice gas occupying a volume fraction ϕ. The change in the free energy due to the interaction between NPM1 and SURF6-N is given by:

$$\frac{F * V_0}{V} = \phi \log \phi + (1 - \phi) \log (1 - \phi) - \chi \phi^2 \qquad (3.3)$$

where V_0 is the volume of a NPM1 pentamer. The first two terms represent the entropy of mixing. χ is the Flory interaction parameter which is used to characterize the difference of
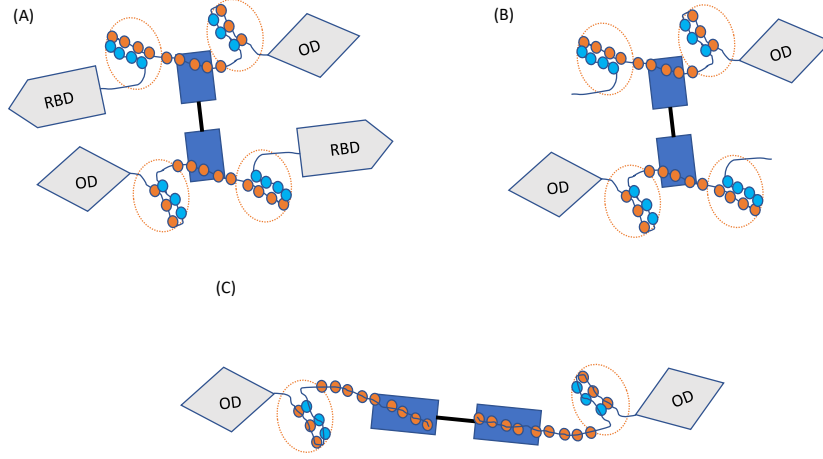
Figure 3.4: *In vitro experiment provides the phase diagram for three different cases. SURF6-N crosslinks NPM1 pentamer in either cases. (A) Wild type NPM1 (B) NPM1 with out RBD region (C) NPM1 with out both RBD region and B2 block.*

interaction energy due to mixing which is given by,

$$\chi \approx \frac{z}{2}(F_{dense} - 2F_{dilute})$$

where $F_{dense} = -\log Q_{dense} = -\log\left(1 + c_{surf}e^{-E_2}\right)$ and $F_{dilute} = -\log Q_{dilute} = -\log\left(1 + c_{surf}e^{-E_1}\right)$ are dense and dilute phase free energy respectively. Finally, we get

$$\chi = -\frac{z}{2}\log\frac{1 + c_{surf}e^{-E_2}}{(1 + c_{surf}e^{-E_1})^2} \tag{3.4}$$

where $z = 5$ is the co-ordination number of NPM1 pentamer, the log term describes the free energy of pair of NPM1 arms. In the soluble state (denominator of log term) the arms are independent, so they can be unbound or bind to a single SURF6-N molecule with energy E_1. In the aggregated state (numerator of log term) the arms can be unbound or a SURF6-N

binds with two arms of distinct pentamer with energy E_2. Now Eq. 3.3 becomes,

$$\frac{F * V_0}{V} = \phi \log \phi + (1 - \phi) \log (1 - \phi) + \phi^2 \frac{z}{2} \log \frac{1 + c_{surf} e^{-E_2}}{(1 + c_{surf} e^{-E_1})^2} \tag{3.5}$$

Now, we need to find the monomer concentration of SURF6-N necessary to drive the phase transition by equalizing the coefficient of the $\phi^2/2$ term to 4 which is obtained by solving $F'' = F''' = 0$ (This is a required condition for phase separation). On solving, we get the critical SURF6-N monomer concentration as

$$c_{surf} \rightarrow \frac{\exp^{-E_2} + 2 \exp^{(-E_1 + \frac{4}{z})} + \sqrt{\exp^{-2E_2} + 4 \exp^{-2E_1 + \frac{4}{z}} - 4 \exp^{-E_1 - E_2 + \frac{4}{z}}}}{\exp^{(2E_1 - \frac{4}{z})}} \tag{3.6}$$

Using Mathematica we can numerically solve the double tangent construction for phase separation, which gives the dilute phase and dense phase NPM1 concentration. The free energy of phase separated system has two minima or we can draw a common tangent line which touches two point on free energy curves. The corresponding concentrations for two minima or at which common tangent line touches the free energy curve give the dilute and dense phase concentration. These concentrations depend on E_1 (binding interaction between a NPM1 arm and a SURF6-N in solution state), E_2 (binding interaction due to a SURF6-N crosslinking two arms of distinct NPM1 in condensed state).

Here we would like to hightlight that we are splitting the electrostatic interaction into two parts: (a) Short range force for NPM1 tail collapse (Table 3.1) and NPM1-SURF6-N interaction (Table 3.2) (b) Long range Debye-Huckel potential for repulsive force between unbound negative charges in NPM1 arms.

The electrostatic interaction energy is very complicated for disordered polymers like this. We start by making the simplest possible approximation in which the polymers are treated as charged spheres in the Debye-Huckel approximation. The Debye-Huckel potential is given by

$$V = \frac{q_1 q_2 l_b \exp^{-\kappa(r-a)}}{r(1 + \kappa a)} \tag{3.7}$$

where q_1 and q_2 are the two charges, l_b is the Bjerrum length, κ is the screening parameter, a is the radius of the charge, and r is the distance between the charges. We will neglect the screening term for the moments (valid if r = a). The screening length at salt concentration $C_{salt} = 150mM$ is given by

$$\kappa = \sqrt{\frac{2 * N_A * q_1 * q_2 * C_{salt}}{\epsilon_r \epsilon_0 K_B T}} \simeq 1.25 * 10^9 nm^{-1}$$

where N_A is Avogadro's number, K_B is the Boltzmann constant, ϵ_r is dielectric constant, ϵ_0 is permittivity of vaccum. Here, we have considered three NPM1 variants studied by[46].

(a) **Full NPM1 (WT)**:

In the WT case, there is only enough unbound negative charge on NPM1 to bind one half of SURF6-N due to the interaction between the B2 region and the A2 region. In the condensed phase, the other half of SURF6-N can also find an adjacent NPM1 tail (Fig. 3.4A). Using Table 3.1, Table 3.2, and Eq. 3.7, we get $E_1 = -10 K_B T$, $E_2 = -15 K_B T$.

b) **NPM1 without RBD domain**:

This case is similar to case first except we do not have the RBD region (Fig. 3.4B) so that the net negative charge in the NPM1 tail increases which causes an increase in the repulsive energy. Hence, using Table 3.1, Table 3.2, and Eq. 3.7 we get, $E_1 = -10 K_B T$, $E_2 = -14.5 K_B T$.

c) **NPM1 without RBD domain and B2 positive block**:

In this case, the electrostatic attraction exist SURF6-N and the A2 negative block of NPM1 tail in solution state while in the condensed state, the SURF6N interacts with A2 block of two distinct NPM1 arms (Fig. 3.4C). Hence, using Table 3.1, Table 3.2, and Eq. 3.7 we get, $E_1 = -17 K_B T$, $E_2 = -21 K_B T$. .

Fig 3.5 shows the calculated phase separation regime which is qualitatively similar to the experimental result[46].

In case of NPM1^{N188} (NPM1 variant which has no RBD region and B2 block), there is no short range polymer collapse due to absence of the positive B2 block and the RBD region. The long range Debye-Huckel interaction between the NPM1^{N188} A2 block and the positively
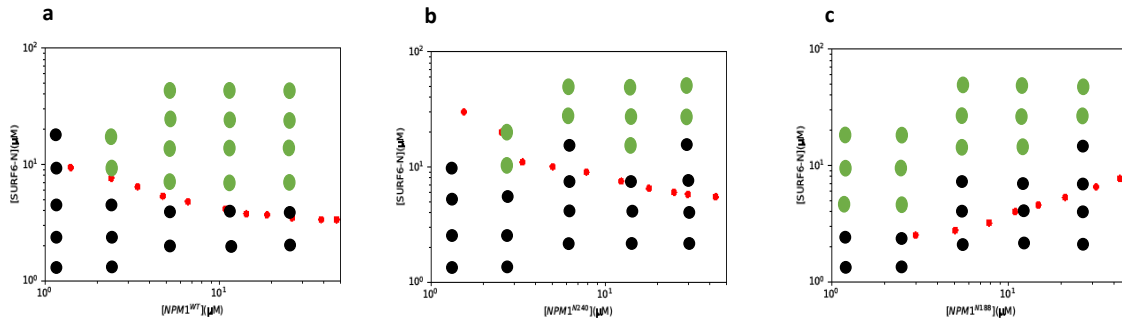
Figure 3.5: *It compares our theoretical phase diagram with the in vitro experimental phase diagram for three NPM1 mutants (a) NPM1^{WT} (b) NPM1^{N240} and (c) NPM1^{N188}. Green dots and black dots represent experimental demixed and mixed phase respectively, and red dashed line represents our theoretical phase separating lines. Qualitatively, our theoretical predictions are in good agreement with experimental findings.*

charged SURF6-N drives phase separation. We observe that the critical SURF6-N concentration increases as NPM1 concentration increases. This is because there is more repulsion between NPM1^{N188} due to the high charge density of the A2 block so that more positively charged SURF6-N are needed to oppose the repulsive force. But in case of NPM1^{WT} and NPM1^{N240} (which has no RBD region), the tail collapses due to short range forces between the A2 and B2 blocks. The tail now has low net negative charge. If we increase NPM1 concentration, NPM1 are easily accessible and hence, small numbers of SURF6-N are sufficient to drive phase separation but for very high NPM1 concentration (not shown in figure), if we increase NPM1 concentration more SURF6-N are required in relative proportion to drive the phase separation. Qualitatively, Phase separating lines for NPM1^{WT} and NPM1^{N240} are similar but there is a quantitative difference due to the fact that NPM1^{N240} has no RBD domain (i.e., it has more negative charge than WT).

3.3 NPM1 protein acts as a chaperone in the rRNA folding process

3.3.1 Introduction

It is thought that the phase separated environment of the nucleolus facilitates the maturation of ribosome[46]. Ribosomes are responsible for reading encoded mRNA to synthesize proteins from amino acid carried by tRNA[20]. Newly synthesized proteins are a random coil of amino acids which need to fold into a stable 3D structure.

Many proteins do not fold themselves. Sometimes, they exist in misfolded state or non-native form[81]. They may lead to pathological aggregates or may be degraded due to interaction with proteases. In many cases, even native proteins become denatured as a result of cellular stresses[81;82]. Proteins might need assistance from other proteins to stop further degradation[83] or the formation of pathological aggregates of unfolded, misfolded and non-native forms[82]. Those proteins that facilitate correct protein folding are called chaperones[84;85][Fig. 3.6]. Chaperone can break apart aggregates, stop the denaturation of native forms by stress, and aid the formation of native structure from unfolded, misfolded, and non-native forms[81;82;84;86]. There are many chaperones inside the cell and they are necessary for life[87;88].

Similar to how protein chaperone assists protein folding, other chaperones aid RNA folding by preventing or resolving misfolding states[89]. RNA has two types of folding problems (a) kinetic trapping in an alternative conformation (b) difficulty in existing the thermodynamically preferred tertiary structure[90]. RNA binding proteins can help to resolve either of RNA folding problem. Non-specifically binding proteins can acts as RNA chaperones by helping the RNA escape kinetic trap whereas specifically binding proteins may act as RNA chaperones by stabilizing specific tertiary structure[91].

As we discussed earlier, NPM1 undergoes phase separation either due to crowding agents or due to interactions with rRNA or positively charged proteins. These modes of interactions have led hypothesis that NPM1 drives ribosome assembly by stabilizing unfolded rRNA
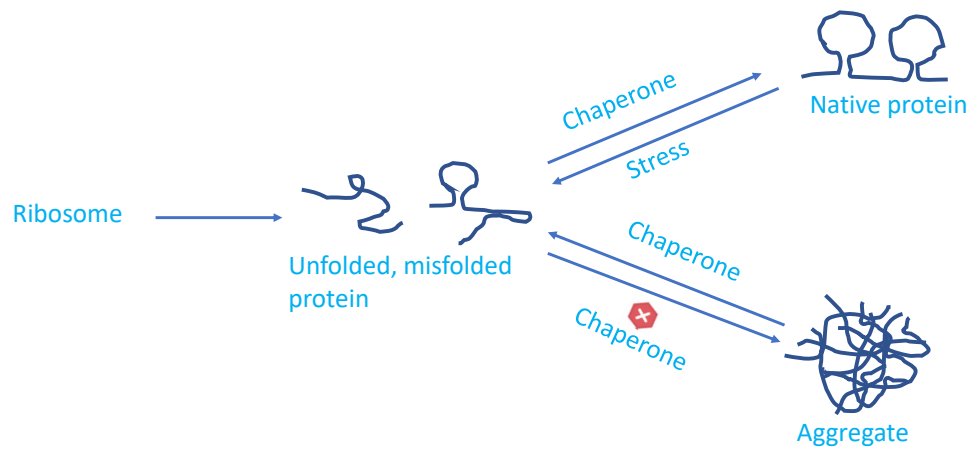
Figure 3.6: *The schematic diagram for chaperone represenataion. Ribosomes produce unfolded proteins. Stress may denature the native protein into unfolded or misfolded state. Chaperone assists unfolded and misfolded protein to get their native form. Some chaperone prevents to occur protein aggregation and some chaperones break apart the aggregation.*

at the GC-DFC boundary, then exchanging NPM1-rRNA interactions for NPM1-SURF6-N interactions as rRNA folds[92]. When rRNA misfolds it is thought that NPM1 in the nucleolus acts as a RNA chaperone by assisting unfolding rRNA so that rRNA can search for the proper configuration in the next attempt. In the following, we will present a model describing how NPM1 can facilitates the folding process.

3.3.2 Methods

When rRNA folds, it forms Watson-Crick helicos that we refer to as "zipper". We assume that the rRNA can form many different zippers, most of which are incorrect or misfolded. The chaperone NPM1 will help unzip these misfolded regions. We call the free energy difference between the misfolded state and unfolded state the zipper height or zipper barrier. In order to compare the effect of chaperone folding rRNA, we consider two cases (i) rRNA with out chaperone: In this case we assume, rRNA unfolds itself due to thermal motion. Here, rRNA may exist in two states (Fig. 3.7A) either the folded state or the unfolded state. (ii) rRNA with chaperone: In this case rRNA unfolds due to both thermal motion and chaperone assistance. Here, system msy exist in four states (Fig. 3.7B). They are, A: folded rRNA, B: partially opened rRNA due to thermal motion, C: NPM1 interacting with a partially unfolded rRNA, D: NPM1 interacting with fully unfolded rRNA. In this section we have introduced three different methods in order to check the effect of chaperone (NPM1) binding in converting between the misfolded and unfolded states.

Matrix method

In matix method, we calculate the probability of unzipping the misfolded rRNA with and without the presence of chaperone (NPM1). In order to see in detail how matrix method works, lets start with a general case and then, apply it to our rRNA/NPM1 case.

Consider, we have a set of states A, B, C, D,...that evolve according to the master
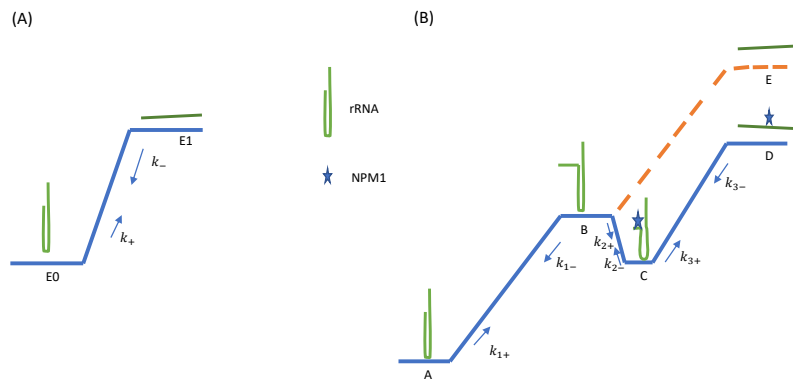
Figure 3.7: *(A) Misfolded rRNA unzips itself in the absence of chaperone. It has two states: E0 represents folded state and E1 represents unfolded state. (B) NPM1 assists in unzipping misfolded rRNA. There are four possible states: A represents folded rRNA, B represents partially opened rRNA, C represents NPM1 (indicated by star) interacting with partially opened rRNA, D represents NPM1 interacting with fully unfold rRNA. The orange dashed line represents the zipper barrier in the absence of chaperone.*

equations

$$\frac{\partial P_A}{\partial t} = -\sum_{i \neq A} k_{Ai} P_A + \sum_{i \neq A} k_{iA} P_i$$

$$\frac{\partial P_B}{\partial t} = -\sum_{i \neq B} k_{Ai} P_B + \sum_{i \neq B} k_{iB} P_i \tag{3.8}$$

and so on.

where, k_{ij} is the rate of $i \rightarrow j$ transition. Since the probabilities are time dependent i.e. $P_A(t), P_B(t), P_C(t)....,$ let us convert the partial derivatives into discrete time.

$$\frac{\partial P}{\partial t} \rightarrow \frac{P(t + \Delta t) - P(t)}{\Delta t} \tag{3.9}$$

Then Eqn. 3.8 becomes,

$$P_A(t + \Delta t) - P(t) = -\sum_{i \neq A} k'_{Ai} P_A + \sum_{i \neq A} k'_{iA} P_i$$

where $k' = k\Delta t$ is the transition probability in time Δt.

Similarly,

$$P_A(t + \Delta t) - P(t) = -\sum_{i \neq A} k'_{Ai} P_A + \sum_{i \neq A} k'_{iA} P_i$$

$$P_B(t + \Delta t) - P(t) = -\sum_{i \neq B} k'_{Ai} P_B + \sum_{i \neq B} k'_{iB} P_i \tag{3.10}$$

$$P_C(t + \Delta t) - P(t) = -\sum_{i \neq C} k'_{Ai} P_C + \sum_{i \neq C} k'_{iC} P_i$$

and so on.

Now, we want Eqn 3.10 in matrix form, which gives the change in probability after time Δt.

$$\vec{P}(t + \Delta t) = M\vec{P}(t)$$

where,

$$\tilde{\mathbf{P}}(\mathbf{t}) = \begin{bmatrix} P_A(t) \\ P_B(t) \\ P_C(t) \\ -- \\ -- \end{bmatrix} \tag{3.11}$$

$$\tilde{\mathbf{P}}(\mathbf{t} + \mathbf{\Delta t}) = \begin{bmatrix} P_A(t + \Delta t) \\ P_B(t + \Delta t) \\ P_C(t + \Delta t) \\ --- \\ --- \end{bmatrix} \tag{3.12}$$

and

$$M = \begin{pmatrix} 1 - \sum K'_{Ai} & K'_{BA} & K'_{CA} & K'_{DA} & K'_{EA} \\ K'_{AB} & 1 - \sum K'_{Bi} & K'_{CB} & K'_{DB} & K'_{EB} \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \\ -- & -- & -- & -- & -- \end{pmatrix} \tag{3.13}$$

For four states A,B,C and D;

$$M = \begin{pmatrix} 1 - \sum K'_{Ai} & K'_{BA} & K'_{CA} & K'_{DA} \\ K'_{AB} & 1 - \sum K'_{Bi} & K'_{CB} & K'_{DB} \\ K'_{AC} & K'_{BC} & 1 - \sum K'_{Ci} & K'_{DC} \\ K'_{AD} & K'_{BD} & K'_{CD} & 1 - \sum K'_{Di} \end{pmatrix} \qquad (3.14)$$

Let us consider the rRNA case, where A is our initial state and D is the final state (Fig. 3.7B). We have assumed that once system reach D state, it remains there (zipper is fully unfolded). Our model (Fig. 3.7) only allows transitions between adjacent states, for example, system at state B can move to state A or state C but can not jump to state D directly. With these restrictions the matrix M becomes,

$$M = \begin{pmatrix} 1 - K'_{AB} & K'_{BA} & 0 & 0 \\ K'_{AB} & 1 - K'_{BA} - K'_{BC} & K'_{CB} & 0 \\ 0 & K'_{BC} & 1 - K'_{CB} - K'_{CD} & 0 \\ 0 & 0 & K'_{CB} & 1 \end{pmatrix} \qquad (3.15)$$

The probability of each state after N steps is given by,

$$\vec{P}(t + N\Delta t) = M^N \vec{P}(t)$$

$$= U^{-1}(UMU^{-1})^N U\vec{P}(t) \qquad (3.16)$$

$$= U^{-1}(D)^N U\vec{P}(t)$$

In Eq. 3.16, $D = UMU^{-1}$ is a diagonal matrix where the diagonal elements $\lambda_1, \lambda_2, \lambda_3, \lambda_4....$, λ's are the eigenvalues of matix M. The largest eigenvalue is $\lambda_1 = 1$ and has corresponding eigenvectors representing the equilibrium distribution of states. We take second largest eigen value of M which represents slowest transition, in this case from state A to D. The probability

of transitions from initial state A to final state D is given by

$$P = 1 - \lambda_2$$

where λ_2 is second largest eigen value of the matrix M.

Now lets calculate the transition probability in the two different cases shown in Fig. 3.7:

(a) **System without chaperone**: Consider a two state system with an energy difference ΔE (Fig. 3.7A) and the ratio of transition rates satisfies the detailed balance relationship $\frac{K_-}{K_+} = e^{\Delta E}$. The transfer matrix will be

$$\begin{pmatrix} 1 - K_- e^{-\Delta E} & 0 \\ K_- e^{-\Delta E} & 1 \end{pmatrix}$$

Finding the second largest eigenvalue from above matrix and subtracting from 1, we get the probability to unzip the misfolded state will be: $P = K_- e^{-\Delta E}$.

(b) **System with chaperone**: Consider a misfolded rRNA has free energy ΔE. In Fig. 3.7B, state A represents fully misfolded rRNA, state B represents the partially opened rRNA zipper, state C represents NPM1 interacting with a partially unfolded states and state D represents NPM1 interacting with fully unfolded rRNA. Now, we have two questions.

(i) Does the chaperone help to increase probability of unfolding the zipper?

(ii) If chaperone helps, what is the best position of NPM1 in the zipper?

In order to answer these question we have considered all positions of NPM1 in the rRNA zipper(Fig. 3.8).

After writing matrix elements for each position and evaluating the probability, it is observed that the chaperone increases the probability of transition to the unfolded state and the effect is maximum when chaperone binds halfway through the unfolding process (Fig. 3.8). The ratio of maximum probability with chaperone to without chaperone is approximately equal to the exponent of the NPM1 interaction energy, therefore, the beneficial effect in all cases arises from lowering the energy of the unfolded state. However, the location effect
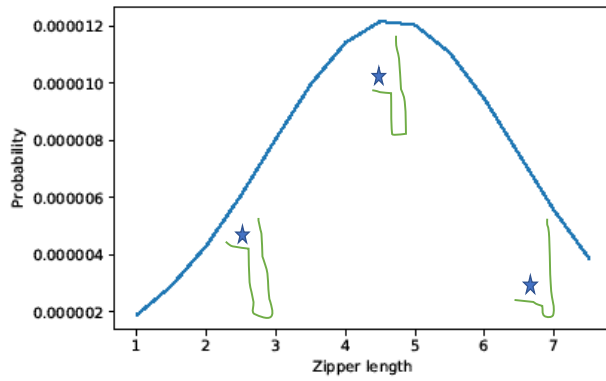
Figure 3.8: *Figure shows the probability of unfolding rRNA as the function of chaperone position in the zipper. rRNA has the total zipper length or barrier of $10K_{\beta}T$. We have taken the several different position of NPM1 in zipper length. Cartoons indicate NPM1 binding to the early, middle, and late stage unzipping.*

cannot be captured from this single free energy argument.

Analytic method

In this section, we calculate the average time required to unfold the misfolded rRNA with and without the presence of chaperone (NPM1).

(a) **Single step model**

This model is analogous to the misfolded rRNA without chaperone in which there are two states shown in Fig 3.9 with energy difference ΔE so that the ratio of transition rate between two states is given by the relation $\frac{K_-}{K_+} = e^{\Delta E}$.
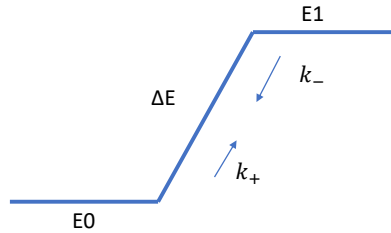
Figure 3.9: *Two state zipper system similar to Fig. 3.7A*

The total time to reach the E_1 state from E_0 state is given by the relation,

$$\text{time} = \frac{1}{K_+} = \frac{1}{K_-}e^{\Delta E} \tag{3.17}$$

The average required time to unfold the rRNA zipper $= e^{\Delta E}/K_-$.

(b) **Three step model**

This model represents the system with chaperone (rRNA in the presence of NPM1) as in Fig. 3.7B. First, we calculate the general expression of average time required to reach the final state 3 from initial inital state 0 (Fig. 3.10) and then, apply the result into the our system of rRNA in the presence of chaperone NPM1.
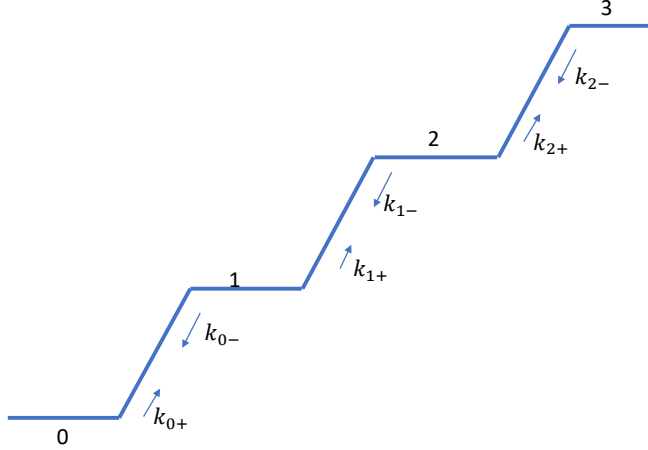
Figure 3.10: *Three step model is a general version of a system rRNA along with presence of chaperone NPM1 (Fig. 3.7B). Fig. 3.7B is also a three step model but the step 2 has lower enery than the step 1.*

Let us consider a three step model which has four states (0,1,2,3) and the transition rates between states shown in Fig 3.10. We are going to calculate the average time required to travel from initial state 0 to the final state 3. If system reaches state 3, we call it a successful walk but if system returns again to the initial state 0, we call it failed walk. Suppose there are N walks among which the number of successful walks is $P_s N$ and the number of failed walks is $P_f N$. The total elapsed time for these N walks is $P_s N t_s + P_f N t_f$.

Hence time per success is,

$$(\frac{P_s N}{P_s N t_s + P_f N t_f})^{-1} = (\frac{P_s}{P_s t_s + P_f t_f})^{-1} \tag{3.18}$$

Fig. 3.11 shows the ways to get the successful and failed trajectories. It is clear from Fig. 3.11A,B that system can move any number of times between states 1 and 2 but success or failure depends on whether it reaches to state 3 or state 0 respectively.

(A) Successful trajectories in the three steps model

	Elapsed time	Probability of each step
	$t_0\ t_1\ t_2$	$P_{1+}\ P_{2+}$
	$t_0\ t_1\ t_2\ t_1\ t_2$	$P_{1+}\ P_{1-}P_{1+}\ P_{2+}$
	$t_0\ t_1\ t_2\ t_1\ t_2\ t_1\ t_2$	$P_{1+}\ P_{1-}P_{1+}\ P_{1-}P_{1+}\ P_{2+}$

(B) Failed trajectories in the three steps model

	Elapsed time	Probability of each step
	$t_0\ t_1$	$P_{0+}\ P_{0-}$
	$t_0\ t_1\ t_0 t_1$	$P_{0+}P_{0-}\ P_{0+}P_{0-}$
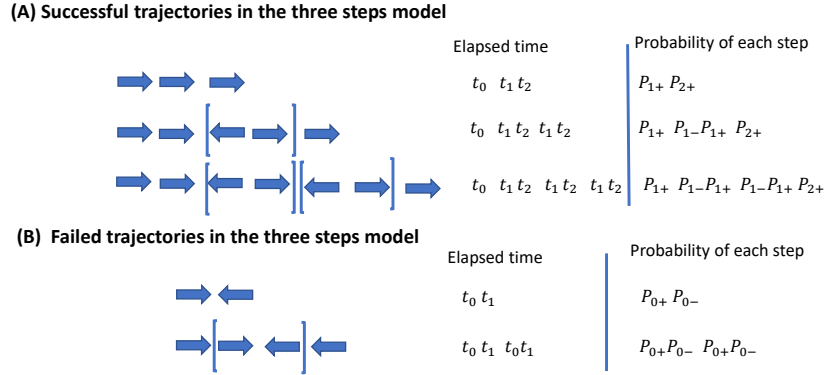
Figure 3.11: *The schematic representation of (A) successful trajectories and (B) failed trajectories in three steps model. Here, the probability of transition from 0→1, 1→0, 1→2, 2→1, 2 →3 are $P_{0+}, P_{0-}, P_{1+}, P_{1-}, P_{2+}$ respectively. Similarly, the average time to travel from 0→1, 1→2 or 1→0, 2→3 or 2→1 are t_0, t_1, t_2 respectively.*

The probability of transition from 0→1, 1→0, 1→2, 2→1, 2 →3 are $P_{0+}, P_{0-}, P_{1+}, P_{1-}, P_{2+}$ respectively. Similarly, the average time to travel from 0→1, 1→2 or 1→0, 2→3 or 2→1 are t_0, t_1, t_2 respectively. Now, average time for a successful trajectories is given by,

$$
\begin{aligned}
\text{Success time} &= \sum_{n=0}^{\infty}[(t_0 + t_1 + t_2) + n(t_1 + t_2)]P_{1+}P_{2+}(P_{1-}P_{1+})^n \\
&= \frac{(t_0 + t_1 + t_2)P_{1+}P_{2+}}{1 - P_{1-}P_{1+}} + \frac{(t_1 + t_2)P_{1+}P_{2+}}{(1 - P_{1-}P_{1+})^2}
\end{aligned}
\tag{3.19}
$$

In Eq. 3.19, the first term $(t_0 + t_1 + t_2)$ gives the time for shortest walk consisting of three forward steps. The second term $n(t_1 + t_2)$ accounts for additional time hopping between states 1 and 2, n times.

Similarly, the average time for a failed trajectories is given by,

$$
\begin{aligned}
\text{Failure time} &= \sum_{n=0}^{\infty} [(t_0 + t_1) + n(t_1 + t_2)] P_{0-}(P_{1-}P_{1+})^n \\
&= \frac{(t_0 + t_1)P_{0-}}{1 - P_{1-}P_{1+}} + \frac{(t_1 + t_2)P_{0-}}{(1 - P_{1-}P_{1+})^2}
\end{aligned}
\tag{3.20}
$$

In Eq. 3.20, the first term $(t_0 + t_1)$ represents the shortest failed trajectory and again, the second term $n(t_1 + t_2)$ represents additional time hopping between states 1 and 2, n times.

The total success probability is given by

$$
\begin{aligned}
P_{tot}^{Success} &= (P_{0+}P_{1+}P_{2+}) \sum_{n=0}^{\infty} (P_{1-}P_{1+})^n \\
&= \frac{P_{1+}P_{2+}}{1 - P_{1-}P_{1+}} \qquad \text{where, } P_{0+} = 1
\end{aligned}
\tag{3.21}
$$

In Eq. 3.21, the term $P_{0+}P_{1+}P_{2+}$ represents the minimal set of moves $0 \rightarrow 1 \rightarrow 2 \rightarrow 3$ in every trajectory while the sum accounts for the different number of hopping between states 1 and 2.

Similarly, total failure probability is,

$$
\begin{aligned}
P_{tot}^{Fail} &= (P_{0+}P_{0-}) \sum_{n=0}^{\infty} (P_{1-}P_{1+})^n \\
&= \frac{P_{0-}}{1 - P_{1-}P_{1+}} \qquad \text{where, } P_{0+} = 1
\end{aligned}
\tag{3.22}
$$

where $P_{0+}P_{0-}$ is the probability of the shortest walk $0 \rightarrow 1 \rightarrow 0$.

Hence, total probability must be 1.

$$
P_{tot}^{Success} + P_{tot}^{fail} = \frac{P_{1+}P_{2+}}{1 - P_{1-}P_{1+}} + \frac{P_{0-}}{1 - P_{1-}P_{1+}} = \frac{P_{1+}(1 - P_{1-})}{1 - P_{1-}P_{1+}} + \frac{1 - P_{1+}}{1 - P_{1-}P_{1+}} = 1
$$

The normalized success time is given by,

$$t_{success} = \frac{\text{Success time}}{P_{tot}^{Success}}$$

$$= (t_0 + t_1 + t_2) + \frac{t_1 + t_2}{1 - P_{1-}P_{1+}}$$

(3.23)

And, the normalized failure time is given by

$$t_{failure} = \frac{\text{Failure time}}{P_{tot}^{Failure}}$$

$$= (t_0 + t_1) + \frac{t_1 + t_2}{1 - P_{1-}P_{1+}}$$

(3.24)

Hence average time per successive move is given by (Eq. 3.18).

$$\text{Average time per success} = \frac{P_s t_s + P_f t_f}{P_s}$$

$$= t_s + \frac{P_f}{P_S} t_f$$

$$= (t_0 + t_1 + t_2) + (\frac{t_1 + t_2}{1 - P_{1-}P_{1+}}) + \frac{P_{0-}}{P_{1+}P_{2+}}[(t_0 + t_1) + (\frac{t_1 + t_2}{1 - P_{1-}P_{1+}})]$$

(3.25)

Application of three steps model to NPM1 and rRNA

Let's apply the results of three step model to a system in which NPM1 (chaperone) interacts with the misfolded rRNA so that energy of E2 state is lower than E1 state. In this case E0 represents a zipper of misfolded rRNA, E1 represents partially opened zipper, E2 represents the chaperone interacting with partially opened zipper and E3 represents the unfolded rRNA with NPM1 still attached as in Fig 3.12.

In our case, we set the downhill processes to have the same free energy, $K_{0-} = K_{1+} = K_{2-} = K_-$, $K_{0+} = K_{0-}e^{-(E_1-E_0)} = K_-e^{-(E_1-E_0)}$, $K_{1-} = K_{1+}e^{-(E_1-E_2)} = K_-e^{-(E_1-E_2)}$, $K_{2+} = K_{2-}e^{-(E_3-E_2)} = K_-e^{-(E_3-E_2)}$

The time taken for first step move (E0 to E1) is,

$$t_0 = \frac{1}{K_{0+}} = \frac{e^{\Delta E}}{K_{0-}} = \frac{e^{E_1-E_0}}{K_-}$$
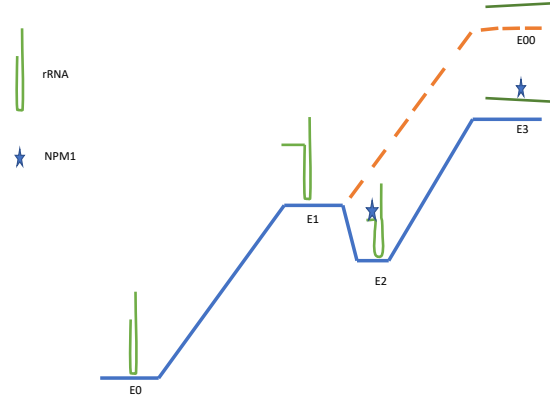
Figure 3.12: *NPM1 acts a chaperone to unfold the misfolded rRNA. State E0 represents a zipper of misfolded rRNA, state E1 represents partially opened zipper, state E2 represents the chaperone interacting with partially opened zipper and state E3 represents the unfolded rRNA with NPM1 still attached.*

The time taken for second step move (E1 to E2) is,

$$t_1 = \frac{1}{K_{0-} + K_{1+}} = \frac{1}{K_- + K_-} = \frac{1}{2K_-}$$

The time taken for third step move (E2 to E3) is,

$$t_2 = \frac{1}{K_{1-} + K_{2+}} = \frac{1}{K_-} \frac{1}{e^{-(E_1 - E_2)} + e^{-(E_3 - E_2)}}$$

The probability that the system moves from E1 to E0 is given by,

$$P_{0-} = \frac{K_{1+}}{K_{1+} + K_{0-}} = \frac{1}{2}$$

The probability that the system moves from E1 to E2 is given by,

$$P_{1+} = 1 - P_{0-} = \frac{1}{2}$$

The probability that the system moves from E2 to E3 is given by,

$$P_{2+} = \frac{K_{2+}}{K_{1-} + K_{2+}} = \frac{e^{-(E_3-E_2)}}{e^{-(E_3-E_2)} + e^{-(E_1-E_2)}} = \frac{1}{e^{(E_3-E_1)} + 1}$$

Similarly, the probability that the system moves from E2 to E1 is given by,

$$P_{1-} = 1 - P_{2+} = \frac{1}{e^{(E_1-E_3)} + 1}$$

Substituting all these values in Eqn. 3.25, we get the result as,

$$\text{Average time per success} = t_0 A + t_2(1 + B + C) \tag{3.26}$$

where, $A = (2 + e^{E_3-E_1})$, $B = \frac{2e^{E1-E3}+1}{2(e^{E1-E3}+1)}$ and $C = (\frac{1}{2}e^{E1-E3} + 1)$
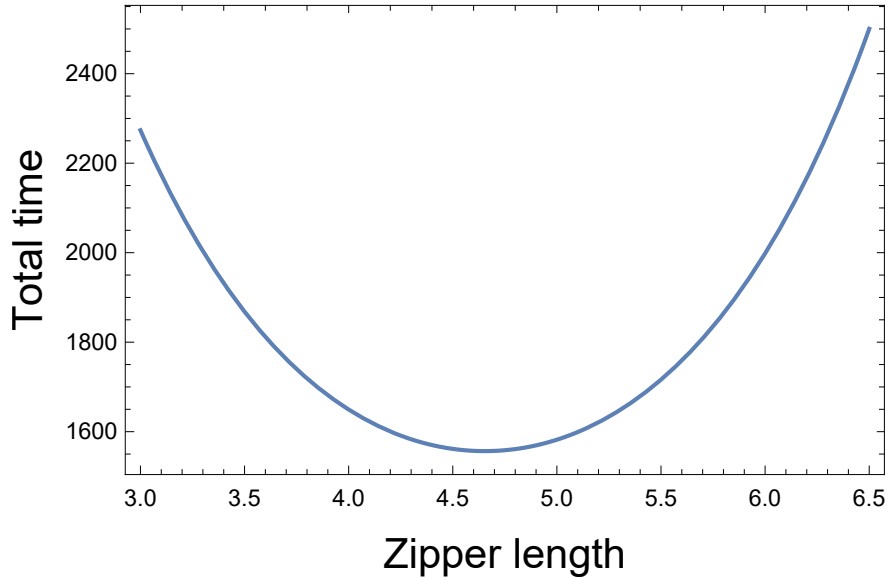
Figure 3.13: *Figure shows the total average time taken to unfold rRNA zipper as the function of chaperone position in the zipper. rRNA has the total zipper length of $10K_BT$. Chaperone lowers the total time about by e^{E_h} (E_h= NPM1 interaction energy). We have assumed that $E_0 = -10K_BT, E_1 - E_2 = E_h = 3K_BT$, $E_3 = E_3' - E_2 = -E_h$.*

Eq. 3.26 is plotted in Fig. 3.13. It is seen that the average time per successful trajectory

is lower in the presence of chaperone (Fig. 3.13),i.e., the chaperone helps to lower the barrier. We observe the minimum time occurs when the chaperone binds at the middle of the zipper. The ratio of minimum time with out chaperone to with chaperone is approximately equal to the exponent of the NPM1 interaction energy.

This calculation for the unzipping time is consistent with the matrix calculation of the transition probability. In both cases the enhancement due to the chaperone is approximately equal to the exponent of the NPM1 interaction energy. These two methods confirm the conclusions that chaperone lowers the zipper barrier and the middle of the zipper is the most effective position.

Random walk method
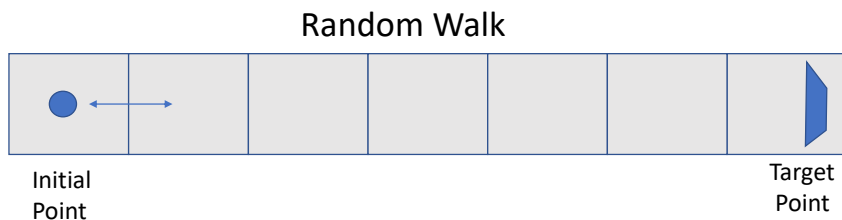
Figure 3.14: *Random walk in one dimension*

We have attempted to understand how chaperone (NPM1) assists the unfolding of misfolded rRNA using a simple random walk. The total number of base pairs bound in the misfolded rRNA zipper is considered as the total distance (grid) to travel by random walk. Consider each base pair in the zipper as a grid, unfolding the zipper is a random walk that

ends on the right side and folding the zipper means a random walk that ends on the left side as shown in the Fig. 3.14.

In an unbiased random walk, the probabilities to move right (unzip an unit) and to move left (zip an unit) are equal (i.e. $P_{zip} = P_{unzip} = 1/2$.) but in a biased random walk, the probability of zipping an unit of zipper is larger than that of unzipping, for example, $P_{zip} = 4/5$ & $P_{unzip} = 1/5$. In each time step, random walk goes to either the left (zip) or the right (unzip). We have accounted the average (average of 10 repeats) total time step spent to unzip barrier (zipper) of different length in both types of random walk which is shown in Table 3.3.

Zipper size	time steps in unbiased random walk	time steps in biased random walk
2	5	30
3	11	200
4	18	1000
5	30	7000
6	37	20000
7	46	100000
8	66	

Table 3.3: *We calculate the total time steps taken to unzip the zipper barrier as the function of zipper length in both unbiased and biased random walk.*

It is observed fron Table 3.3 that total time steps taken is proportional to square of zipper length for an unbiased random walk[93].

In this method, we have assumed that the presence of a chaperone breaks a longer zipper into the two smaller pieces. If the chaperone breaks a longer zipper into two smaller zippers, total time steps taken to unzip the whole barrier is the sum of total time steps taken to unzip these two smaller barriers. Now we are going to figure out that which one is easier to unzip either a full zipper or the two smaller zipper of same total length. From Table 3.3, it is seen that for both biased and unbiased random walk,

(1) The total time steps to fully unzip single barrier is greater than the sum of total time steps taken to unzip two smaller barriers of same total length.

(2) The total time steps to unzip two smaller zippers of equal length is always smaller than that of two zippers of unequal length in which total length of zippers in both cases are same. Mathematically,

$$T_L > T_{\frac{L}{2}-N,\frac{L}{2}+N} > T_{\frac{L}{2},\frac{L}{2}}$$

where, L= total length of zipper, N=1,2...$\frac{L}{2}-1$, T_L = total time steps to unzip the full zipper, $T_{\frac{L}{2},\frac{L}{2}}$=total time steps to unzip two small barrier of same length, $T_{\frac{L}{2}-N,\frac{L}{2}+N}$=total time steps to unzip two small barrier of unequal length.

Let us compare this model with our system of misfolded rRNA in the presence of chaperone molecules (NPM1). The binding of NPM1 at any position of misfolded rRNA zipper splits a zipper into two smaller zippers. We have a result that total time steps to unzip two smaller zipper is always smaller than that to unzip a single zipper of same total length, i.e., NPM1 lowers the zipper barrier. Also, we have an another result that total time steps to unzip two equal zipper is always smaller than that to unzip two unequal zipper of same total length, i.e., the middle of the rRNA zipper is the best placement for the chaperone NPM1 to be most effective.

Hence, all three methods indicate that the chaperone lowers the barrier height and the middle of zipper is the most effective chaperone position to lower the barrier.

3.4 Discussion

Nucleolus is very important cellular bodies for the existence of life. It involves in major cellular processes such as DNA transcription, rRNA processing and modification, ribosome biogenesis etc. We have applied our modeling technique to the nucleolus to determine its internal structure and learn how this contribution to function. It has been observed from our study that preribosomal subunit construction is the result of phase separation of nucleoplasmin (NPM1) molecules due to electrostatic interaction with SURF6-N protein. The interaction between NPM1 and SURF6-N not only drives phase separation but also detaches the rRNA and NPM1 interaction as a result rRNA folds. NPM1 also acts as chaperone

moleculecule to unzip the misfolded rRNA and the middle of the zipper is the most effective position of the chaperone to unfold the rRNA.

Chapter 4

The prediction of antibody viscosity from dilute solution measurements

4.1 Introduction

Antibodies are method of treatment of many diseases[21;22]. It can take several hours to deliver a therapeutic dose of antibodies to a patience if prepared at low concentration. On the other hand, high concentration solutions can be administered in a single injection, but may have a high viscosity that creates problems in production and subcutaneous delivery[27;28;94–103]. Unfortunately, the viscosity issue only shows up late in the development pipeline. There are not adequate strategies and guidelines to test antibody viscosity in early development so that Pharmaceutical companies can focus on low viscosity candidate molecules or optimize possible candidates for lower viscosity. A better approach would be to predict viscosity from dilute solution experiments[104–106]. Since viscosity is a many body problem whereas the dilute solution measurements only reveal two body effects, we need a hypothesis that connects low concentration measurements to viscosity.

Antibodies have sharp rise in viscosity at low volume fraction compared to colloids. The high viscosity of antibody when compared with rigid body[28] is not only caused by excluded volume effects but also inter-molecular interaction[107;108] due to entanglement similar to that
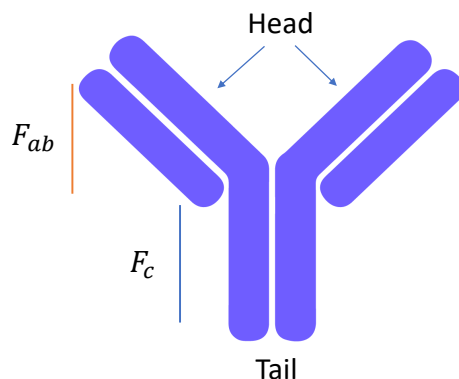
Figure 4.1: *Schematic represenation of antibody*

of long, flexible polymers. It has been observed from experiments that antibody possesses mainly two type of self-association[25;26;109;110] (a) binding between antigen domains ($F_{ab} - F_{ab}$ interaction, which we call Head-Head, HH binding). (b) binding between the antigen domain and the constant domain ($F_{ab} - F_c$ interaction, also called Head-Tail, HT binding). Both of these result in elongated complexes. These extended but flexible complexes entangle with each other resulting the sharp rise in solution viscosity. To address impact of inter-molecular interaction, we have introduced an entanglement model[25;26] to predict antibody viscosity. The entanglement model provides the connection between two-body effects (binding affinity of molecules) and many-body effects (solution viscosity). In this chapter, we clarify how this model enables the prediction of antibody viscosity from low concentration measurements.

There have been previous attempts to predict viscosity in the early stage of antibody development but they do not yet have satisfactory reliability. Dilute solution measurements such as ACSINS and DLS (we have explained about ACSINS and DLS in method section) are one promising approach. Our entanglement model has revealed some limitations in these dilute solution measurements and provided insights into how they can be refined. One

limitation is that some experiments are not sensitive to certain interactions, for instance, ACSINS experiment can not detect HT interactions, only HH binding. Other limitations are that dilute solution experiments can not detect any many body effects or identify molecules prone to forming branched complexes.

The correlation between dilute solution measurements has helped us to identify and filter outliers to make these experiments more reliable in predicting viscosity. Our entanglement model acts as a bridge between dilute solution experiments and viscosity so that we can make quantitative predictions of antibody viscosity as a function of concentration.

4.2 Methods

4.2.1 Entanglement model
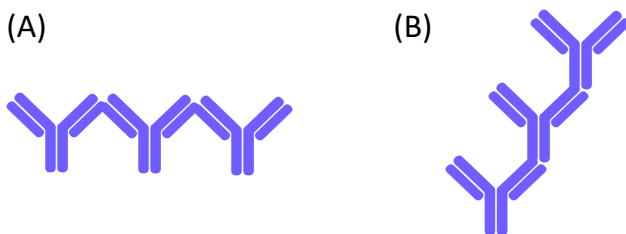
Figure 4.2: *Different types of antibody binding (a) Head-Head (HH) binding (b) Head-Tail (HT) binding*

Affinity parameter gives the distribution of complex sizes.

Here, I have summarized previous work in our group to determine the size of antibody complexes[25;26].

There are mainly two formss of association between antibody molecules: HH and HT binding.

(a) Head-Head (HH) binding: Head-Head (HH) binding is the binding between antigen binding domains of antibody molecules, as shown in Fig. 4.2A. The equilibrium constant for HH binding between antibody molecules is given by

$$K = \frac{C_2}{C^2} \tag{4.1}$$

where C_2 is the dimer concentration and C is monomer concentration.

Similarly, the concentration of longer complexes is defined by $K = \frac{C_i}{C_{i-1}C}$. So,

$$C_i = K^{i-1}C^i \tag{4.2}$$

Since the same interaction site is responsible for the addition of each molecules, K is same each molecular addition. The grand partition function of the system is the sum of the concentration of monomers and all longer complexes.

$$Q = C + KC^2 + K^2C^3 + K^3C^4 + ...$$
$$= C(1 + KC + K^2C^2 + K^3C^3 + ...) \tag{4.3}$$

This power series can be summed yielding,

$$Q = \frac{C}{1 - KC} \tag{4.4}$$

The average size of complexes is given by

$$< n >_{HH}= C\frac{\partial \log Q}{\partial C} = \frac{2KC}{\sqrt{1+4KC}-1} \tag{4.5}$$

(b) Head-Tail (HT) binding: Head-Tail binding is the binding between antigen binding domain and the constant domains of an antibody. While writing partition function, we have to include the fact that HT binding could form multiple distinguishable complexes from the same number of monomers. The calculation accounting for this is given by Ramallo et al[25]. The average aggregate size for HT binding is

$$< n >_{HT}= \frac{\sqrt{2}SC}{\sqrt{1+4SC+S^2C^2-(1+SC)\sqrt{1+6SC+S^2C^2}}} \tag{4.6}$$

where S is equilibrium constant for HT binding.

It is clearly seen from Eqs. 4.5 and 4.6 that the equilibrium constant or binding affinity parameter controls the size of complexes.

Reptation gives dynamical properties (kinetics)

Antibody complexes can be modeled as a polymer. At certain concentration of polymer, the random coil volume of polymer covers the whole volume of solution and above this concentration, the polymers are entangled. This threshold concentration is called the overlap concentration. Above the overlap concentration, entangled antibodies can be treated as a semidilute polymer solution[26]. In semidilute regime, the viscosity is proportional to the product of the elastic modulus G and the longest relaxation time τ[111].

$$n \propto G\tau \tag{4.7}$$

The movement of an entangled polymer in semi-dilute regime looks like the motion of a polymer in a tube[111]. Encircling and entwining polymer molecules form such a virtual tube.

Now, The diffusion path of polymer within the tube resembles as a creeping snake[112]. The term reptation explains the diffusion path in an entangled polymer.

In the semidilute regime, the relaxation time $(\tau)^{112;113}$ and shear modulus $(G)^{111}$ are given by

$$\tau_{rep} \sim \frac{n_s L^{3\nu} b^3}{K_\beta T} \left(\frac{C}{C^*}\right)^{\frac{(3-3\nu)}{3\nu-1}} \tag{4.8}$$

$$G \sim \frac{C}{L} K_\beta T \left(\frac{C}{C^*}\right)^{\frac{1}{3\nu-1}} \tag{4.9}$$

where n_s is solvent viscosity, $\nu \sim \frac{3}{5}$ is flory exponent and $C^* \sim L^{1-3\nu} b^{-3}$ is the overlap concentration[111].

This reptation model is valid if the polymer length is much larger than the monomer size. However, this model provides reasonable agreement with experiments, even though the average length of complexes, between 3 to 10 molecules, is too short to satisfy the criteria[26].

Using Eqn. 4.8 and Eqn. 4.9, Eqn. 4.7 becomes,

$$n \sim C^{\frac{3}{3\nu-1}} L^3 \tag{4.10}$$

Using the average length for HH binding, above equation becomes,

$$n = A C^{\frac{3}{3\nu-1}} \left(\frac{2KC}{\sqrt{1+4KC}-1}\right)^3 \tag{4.11}$$

where A is a proportionality constant with the value $5.4 * 10^{-8} cp \left(\frac{mg}{ml}\right)^{3.75}$ obtained by fitting this model to experiments[26]. It is seen that the viscosity of antibodies in the semidilute regime not only depends on concentration but also depends on the binding affinity.

In conclusion, our entanglement model connects the binding affinity of molecules with viscosity. So, the next step is to predict the binding affinity from dilute solution measurements.

4.2.2   Methods

AMGEN provided us with a data set containing viscosity, ACSINS measurements and DLS measurements of sixty three antibodies. We used the ACSINS diffusion value, ACSINS plasmon shift, DLS kD values, the viscosity around 70 mg/ml and viscosity around 150 mg/ml in our analysis. In this section, we briefly discuss about the experiments and their measurements.

ACSINS Method

ACSINS (Affinity-Capture Self Interaction Nanoparticle Spectroscopy) is a new approach to measure monoclonal antibodies (mABs) self-association[101;104–106;114;115]. In this method, gold nanoparticles are coated with capture antibodies and non-capture antibodies (Fig. 4.3). Capture antibodies bind human mABs to be tested from dilute solution whereas non-capture antibodies do not interact with the test antibodies but they prevent test mABs from interacting directly with the gold nanoparticle. These coated nanoparticles create a crowded environment for low concentration mABs to self-associate through multi-body inter-molecular interaction. Antibody self-associations can be measured in two ways; the apparent diffusion of nano-particles and the plasmon wavelength λ_p which is the maximum absorbance wavelength in the visible spectrum. Change in the plasmon wavelength upon self-association of antibodies reflects the strength of self-association i.e. the large change in plasmon wavelength($\Delta\lambda_p$) indicates stronger self-interaction between mAbs. The ratio of capture to non-capture antibody also controls the association of nanoparticles by controlling the density of test mABS.

 (a) **ACSINS amplifies weak interaction.**
ACSINS not only creates the crowded environment for human mAbs to self-associate but also amplifies weak interactions between mAbs. This amplification depends on the ratio of capture to non-capture antibody which are used to coat the gold nano-particle. The higher the ratio, more test antibodies captured on a gold nanoparticle and hence, the more weak interactions between nanoparticles (Fig 4.3). We can tune the strength of interactions between nano-particles by changing density of capture antibody.
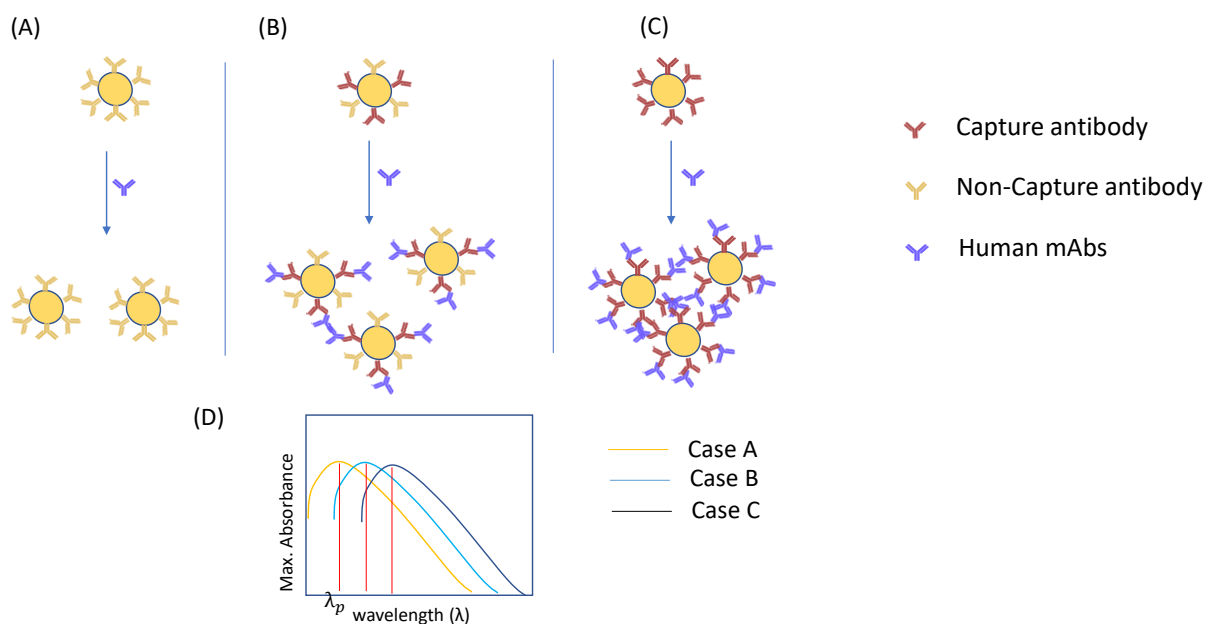
Figure 4.3: *The schematic diagram of ACSINS method (a) gold nanoparticle coated with only non-captured antibody doesnot assist further assembly. (b) the ratio of capture to non-capture antibody control the self-association of antibody (c) gold nanoparticle coated only with capture antibody amplifies weak interaction to a large extent (d) large shift in plasmon wavelength is the indicator of strong self association.*

(b) **ACSINS does not see HT intercations.**

Since the capture antibody is designated to bind the tail of human mAbs, ACSINS allows the self-interaction of human mAbs only through head-head interaction i.e. ACSINS method is blind to HT interactions (Fig 4.3).

(c) **ACSINS diffusion (Nanoparticle diffusion) is not sensitive to strong interactions.**

It has been observed that ACSINS diffusion (Nano-particle diffusion) quickly decays with binding affinity, but is not sensitive to strong interaction (Fig. 4.4A). If capture antibody density is high, it amplifies the weak interaction so that it forms the longer complexes even if binding affinity is very low. For intermediate or strong interaction, they may form fractal structure.
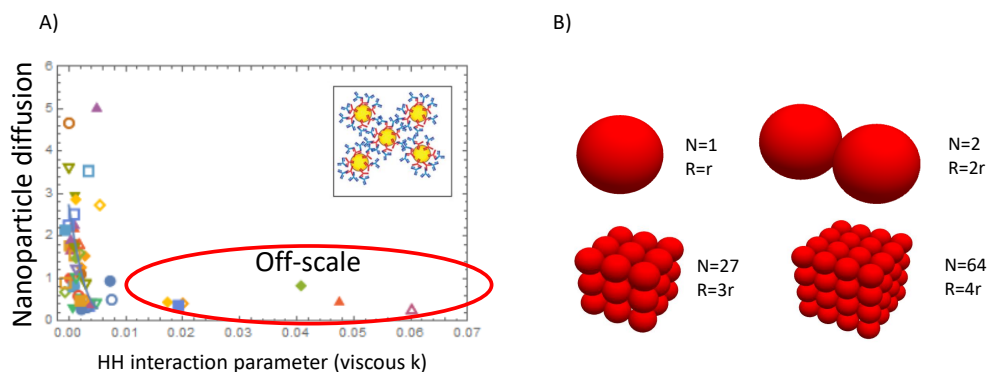
Figure 4.4: *The ratio of capture to non-capture antibody in coated gold nano-particle determines ACSINS sensitivity range. (a) Strong interaction with high capture antibody may result fractal form (b) Size of fractal structure is almost same for large assembly*

The size of fractal structure is given by $R \propto N^{\frac{1}{d_f}}$, where d_f is fractal dimension and N is total molecules in assembly. The diffusion coefficient is given by Stokes-Einsteins equation

as:

$$D = \frac{k_B T}{6\pi\eta R} \propto \frac{1}{R}$$

For high binding affinity (strong interaction), each cluster has large number of nanoparticles. It is seen from Fig. 4.4B that cluster size changes significantly even for small growth but for large cluster, R varies weakly with N so that the diffusion coefficient is sensitive to small clusters (weak interaction) but it is almost constant for large cluster (strong interaction).

(d) **ACSINS lambda shift is not sensitive to weak interaction.**

In ACSINS, antibody self association can also be assessed by the change in plasmon wavelength. Plasmon shift depends on the volume of the nanoparticle cluster. As shown in Fig. 4.5A, the volume for smaller clusters is low where as that volume is high for larger complexes. Therefore weak interactions that form small sized complexes (Fig. 4.5B) have a small plasmon shift. As a result weak interaction are not distinguishable because the interior volumes of small sized clusters are almost same. But from Fig. 4.5B, it is seen that nano-particle plasmon is sensitive to large clusters because the volume increases significantly in large size clusters.

Since both ACSINS and DLS are not sensitive to strong interaction, plasmon shift could be the most appropriate way to measure strong particle affinity but there are still some outliners for strong interaction (Fig. 4.5B). This could be because the strong interactions are prone to kinetic traping so that strongly interacting systems are sensitive to waiting time.

Dynamic light scattering (DLS)

DLS is a light scattering experiment in which the particles are illuminated by laser light and scattered light is studied by a detector. The information imprinted in the scattered light can be used to determine the particle size and mutual diffusion coefficient[27;104;116–118]. The mutual diffusion coefficient (D_m) is a measure of rate of a non-equilibrium mass transfer process.[119]. The plot of the mutual diffusion coefficient (D_m) as a function of antibody concentration(C) (Eqn. 4.12) gives self diffusion coefficient (D_o) as a y-intercept value and diffusion interaction parameter (k_D) as a slope. The self diffusion coefficient (D_O) gives the
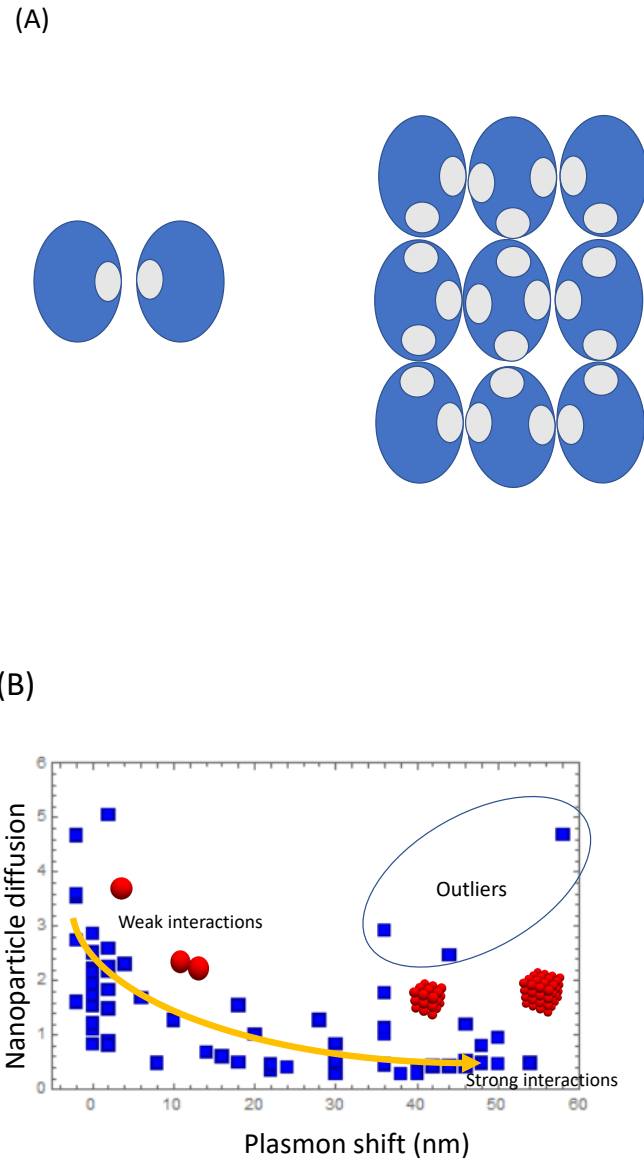
Figure 4.5: *Plasmon shift measures the strength of mAbs self association. (a) Volume that tends to form cluster determines the plasmon shift (b) Plasmon shift is sensitive to strong interaction but not for weak interaction. The outliers may be the result of kinetic trapping in the strongly interacting system.*

measurement of molecular movement or displacement due to the thermal motion.

$$D_M = D_O(1 + k_D C) \qquad (4.12)$$

Eq. 4.12 shows the first order concentration dependence of the diffusion coefficient. The diffusion interaction parameter (k_D) is very useful to study the interaction of molecules[120]. If k_D is negative, the diffusion coefficient decreases with concentration, which indicates that the molecules are attractive. Positive k_D values are indicative of repulsive inter-molecular interactions for which diffusion coefficient increases with concentration. The value of k_D depends on the solution environment such as PH, ionic strength.
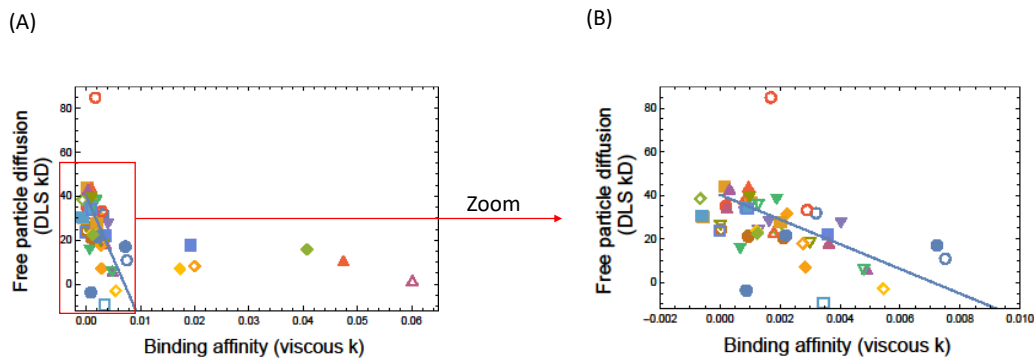
Figure 4.6: *This graphs shows free particle diffusion for different binding affinity range. Binding affinity (K) represents viscous behavior of of molecules. (A) DLS is not sensitive to strong interaction (B) Zoom in of free particle diffusion only for low viscous molecules. Free particle diffusion decays linearly with binding affinities of low viscous molecules.*

Fig. 4.6A shows that strong interaction are only outliers, i.e., DLS sees HT interaction. Therefore, DLS addresses the inability of ACSINS to detect HT interaction but it also has difficulty with strongly interacting molecules.

4.2.3 Proposed method

In previous section we have discussed some drawbacks in dilute solution measurements so that they are not successful to predict about all molecules. The following are the proposed solutions for these drawbacks.
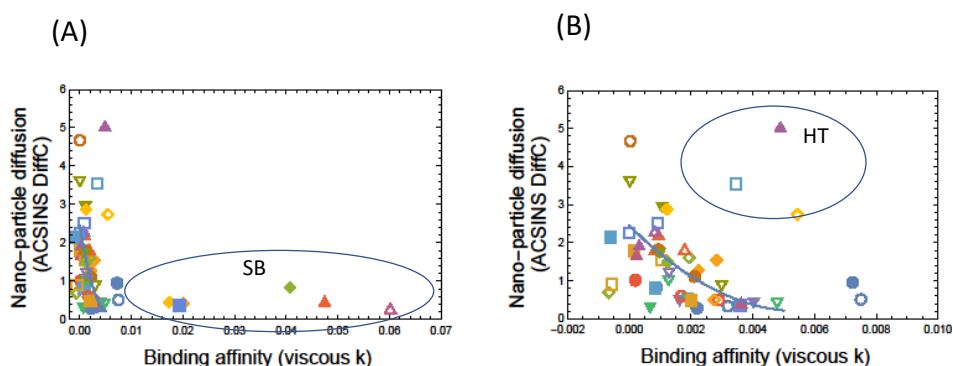
Figure 4.7: *Nono-particle diffusion quickly decays with binding affinity. Plot (a) and (b) are same but are plotted in different affinity range.(a) Nano-particle diffusion is not sensitive to strong binder (Generally, molecules with binding affinity larger than 0.01 are strong binders (SB).) (b) Nano-particle diffusion is blind to HT interaction. Three molecules at the center of right side figure which have large ACSINS diffusion but binding affinity is not low. This discrepancy between the viscosity and the ACSINS diffusion can be explained if the nanoparticle capture antibodies block the interaction that drives viscosity. Therefore, we predict these are HT outliers.*

(a) **Reduce capture antibody density**: As we discuss earlier, ACSINS diffusion is not sensitive to strong interaction (flat region in Fig. 4.7A) but if we reduce the capture antibody density, we can reduce the number of antibody-antibody interaction between nanoparticles. This can shift the ACSINS diffusion back into the sensitive range. In case of reduced capture antibody density weak interaction may not form longer complexes (flat region in Fig. 4.8) and the strong interaction comes in sensitive range.
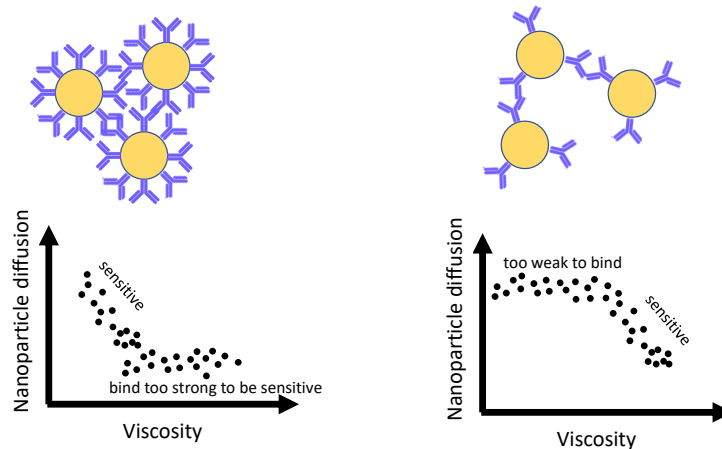
Figure 4.8: *Adjusting the density of capture antibodies whether ACSINS is sensitive to strong or weak binding.*

(b) **Filter HT molecules**: ACSINS does not see HT interaction (Fig. 4.7B) but DLS does. If we plot ACSINS diffusion (ACSINS DiffC) and DLS diffusion (kD) together, we expect they will show a linear relationship as long as the interaction sites are accessible in both experiments. Therefore, we predict the outliers in the plot are HT molecules. This provides a method to filter these molecules in our prediction algorithm. Alternatively, if we choose two different types capture antibody which could bind either the head or tail of the test mAbs, we can separately test HT interactions by ACSINS (Fig. 4.9).

(c) **Avoid kinetic trap**: The plasmon shift is sensitive to strong interaction and is a promising approach to assess the strong interactions but there are still some outliers. These outliers may be the result of kinetic trapping. Kinetic trapping, an outcome in which proteins are restrained to non-equilibrium conformations because of an energy barrier that cannot be quick overcome with out external force or extended timescales. The strong interaction regime where the plasmon shift is sensitive is particularly prone to kinetic trap. The equilibrium time for strong interactions is long due to the time required for bonds to break and re-form. Therefore, the waiting time (time between sample preparation and measurement) may effect

Standard AC-SINS   Proposed solution
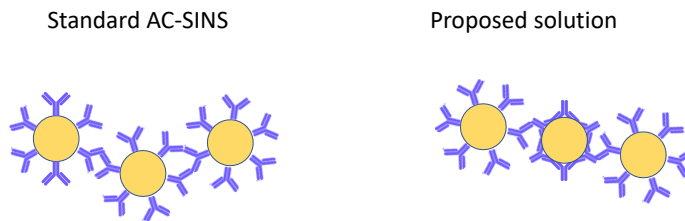
Figure 4.9: *Inverted ACSINS mechanism could solve HT problem.*

the experimental result. To avoid being kinetically trapped, we can reduce the strength of weak interaction by reducing the capture antibody density (Fig. 4.8).

(d) **Find best fit line**: Based on the sample data set, we can predict the viscosity from ACSINS diffusion values or DLS kD values which is described in the following steps. (i) Sample data provides the viscosity at two different concentration points. Using our entanglement model, we can determine the binding affinity from these two viscosity values for each molecules. (ii) Then, we plot "Nano-particle diffusion vs binding affinity" and "Free particle diffusion vs binding affinity" for all sample molecules. Next, we find the best fit line. (iii) Using the best fit line, we can estimate the binding affinity for given ACSINS and DLS value for the test set molecules. (iv) This best fit line can be used to calculate the viscosity of new molecules using the entanglement theory.

The method described in (a)-(d) is shown as the flow chart in Fig. 4.10.

4.3 Results

We have sample data of about 60 molecules, which provides ACSINS diffusion, DLS kD value, plasmon shift($\Delta\lambda$), viscosity at 70 mg/mL, viscosity at 150 mg/mL etc. Using these
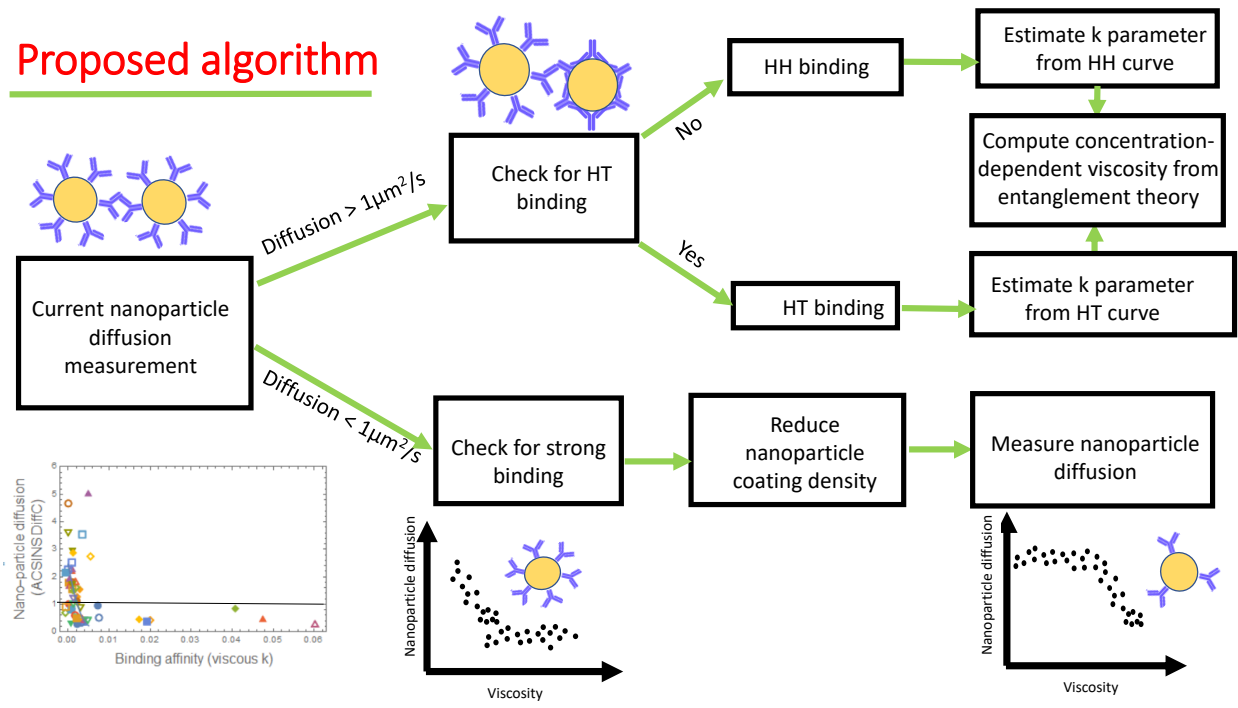
Figure 4.10: *Flow chart of the proposed method for predicting antibody viscosity.*

two viscosity value and our entanglement model, we calculate binding affinity parameter, K, for each molecule. Then we plot ACSINS diffusion vs binding affinity for sample data. The following are the steps we have taken.

4.3.1   Classification of molecules

Comparing ACSINS to the viscosity, we identify four groups (Fig. 4.11): 1) **HT group**: These molecules have very high ACSINS diffusion values, but also have a high viscosity. 2) **Strong Binders (SB) group**: These molecules have low ACSINS diffusion values and very high binding affinity. But ACSINS diffusion is not as low as the regression line would predict. 3) **False Positive (FP) group**: These molecules have both low ACSINS DiffC values and viscous values. It is unclear what may cause this discrepancy but they do not lie within the range of our interest because they are low viscosity molecules. 4) **Trend Line (TL) group**: Molecules having ACSINS diffusion values close with the regression line lie in this group.

89
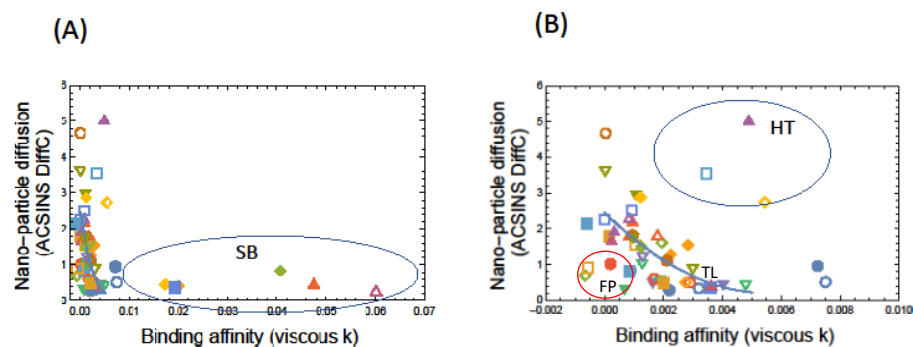
Figure 4.11: *Comparision of the ACSINS diffusion and solution viscosity reveals four groups. (A) Molecules having very high binding affinity are the Strong Binder(SB) groups. (B) The same plot as '(A)' but binding affinity range is limited. Molecules having both diffusion and binding affinity high are classified into HT group. Similarly, molecules having both diffusion and binding affinity low are classified into the False Positive (FP) group. Other molecules which follow the expected trend between diffusion and viscosity are the Trend Line (TL) group.*

4.3.2 Filtering HT molecules

If we plot nanoparticle diffusion (ACSINS DiffC) vs free particle diffusion (DLS kD), they show an linear relationship as shown in Fig. 4.12. There are big outliers in the upper left corner of the graph which we predict are HT binding molecules because those interactions can not be seen by the ACSINS method.
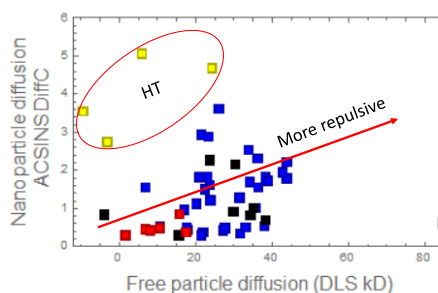
Figure 4.12: *We expect the linear relationship even though nano-particle diffusion and free particle diffusion are two distinct diffusion parameters. The relation between two different diffusion measurements identifies the HT outliers enclosed in a red circle.*

We have proposed that these molecules can be assessed by ACSINS if we reverse the capture orientation on half of the nanoparticles so that HT binding between mAbs is possible 4.13).

4.3.3 Finding regression line

Our entanglement model implies that the viscosity is the function of concentration and length of complexes ($\eta = AC^{3.75}L^3$, where length L is the function of concentration and binding affinity). We calculate binding affinity for each molecule using viscosity at two different point concentrations. Then, we plot the ACSINS diffusion coefficient ($DiffC$) vs binding affinity and the DLS interaction parameter (k_D) vs binding affinity as shown in Fig. 4.6 and Fig 4.7.
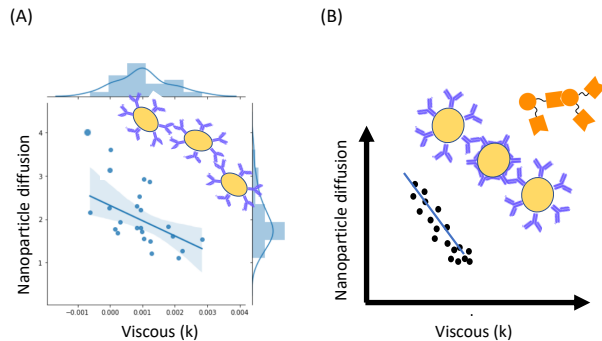
Figure 4.13: *(A) ACSINS measurement does not detect HT interactions (see cartoon) so that we see the data points scattered from the best fit line. (B) We propose inverted ACSINS (see cartoon) to measure HT interactions and generate a best fit line for HT molecules (data and trend line are cartoons).*

The best regression equations for these two cases are as follows:

$$DiffC = 4 \exp^{-K*750} \tag{4.13}$$

$$k_D = 55 - K * 14000 \tag{4.14}$$

From these relations we can predict the antibody binding affinity from ACSINS and DLS measurements. Then, we can predict the concentration dependent viscosity using our entanglement theory.

4.3.4   Testing our hypothesis in new data set

After developing the algorithm described above, AMGEN provided us with a test set of 17 molecules with ACSINS diffusion coefficient ($DiffC$) and DLS interaction parameter (k_D). The predicted viscosity as the function of concentration for these test data set are shown in Fig. 4.14. The comparison of actual viscosity and viscosity predicted from ACSINS and

DLS measurement are shown in bar plot (Fig 4.15).
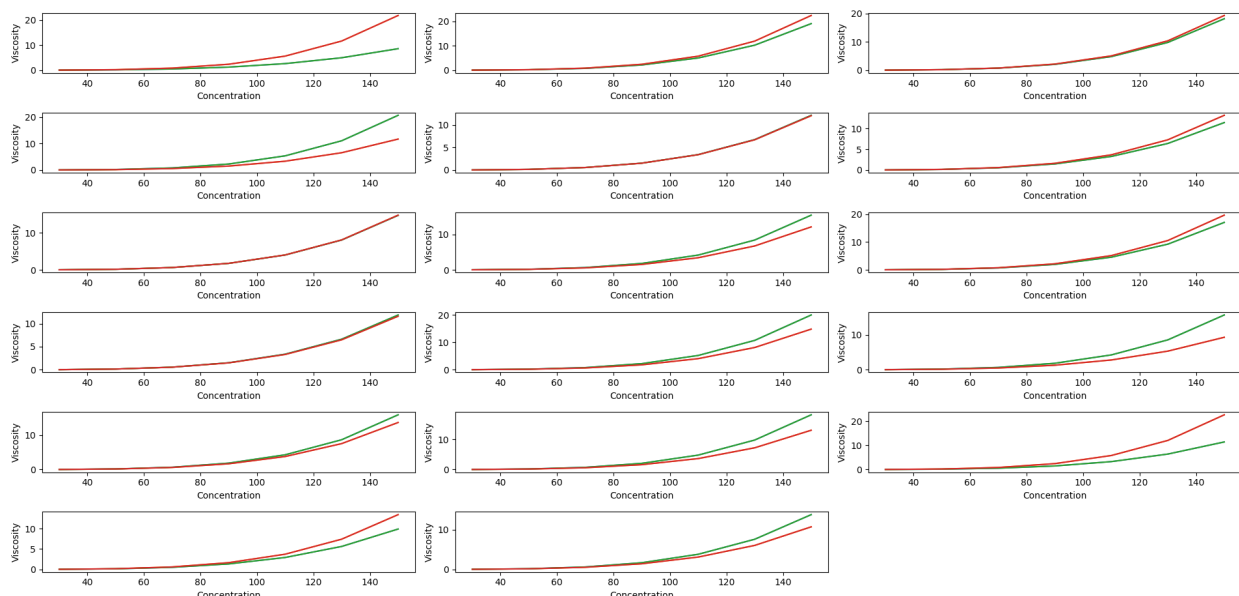
Figure 4.14: *The predicted viscosity from ACSINS and DLS measurements for test set of molecules.*
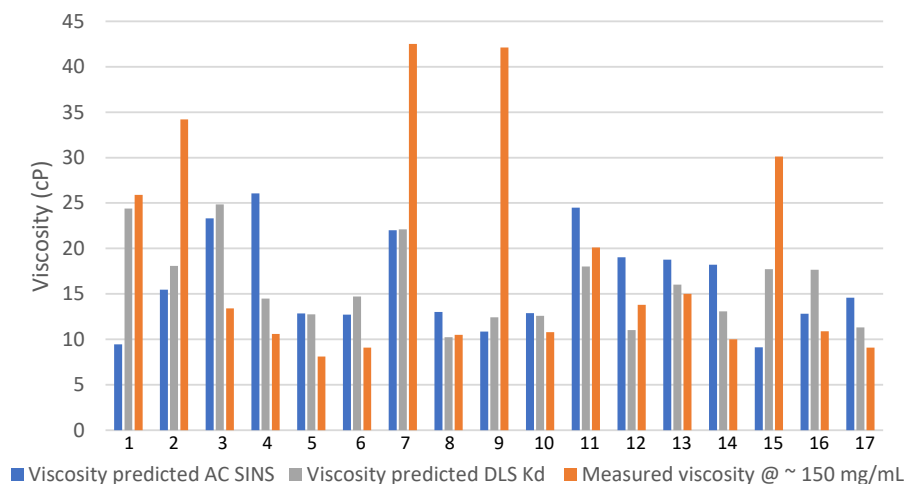
Figure 4.15: *Dilute solution measurements looks promising to predict the antibody viscosity.*

It is clearly seen from bar diagram (Fig 4.15) that our predicted viscosity matches well with actual viscosity although there are also outliers. In cases where the ACSINS viscosity prediction is much lower than the DLS viscosity prediction, the DLS viscosity and the actual viscosity are in good agreement. This is the expected signature of HT binders. If the

actual viscosity is much larger than the ACSINS viscosity prediction and the DLS viscosity prediction, such molecules belong to the strong binder (SB) group. If we apply the sample set observation into test set molecules, molecule 1 is an HT candidate and molecules 2, 7, 9 and 15 are possible SB candidates (Fig 4.15). The modified dilute solution experiments suggested above can be used to test these predictions and validate our method.

4.4 Discussion

The large mass of antibody required for therapeutic doses requires high concentration that can result in problematic viscosity in production and delivery. Predicting antibody viscosity in the early stage of development pipeline could assist in cost reduction. The reason for sharp rise in antibody viscosity at the required dosing is identified as similar to the entanglement of long, flexible polymers. Recent years, there have been many efforts to predict antibody viscosity from dilute solution measurements. Our entanglement model[26] links two-body binding affinity to the many-body viscosity. In this study we have connected dilute solution measurements to the binding affinity of molecule in order to allow binding affinity to act as a bridge between dilute solution measurements and viscosity. Our model has been successful in predicting viscosity from dilute solution measurements except some outliers. We have not only identified the outliers but also explained the physical mechanics behind these outliers and proposed the solutions to make the dilute solution measurements more efficient. The utility of our method will depend on the experimental verification of the mechanism of these outliers.

Chapter 5

Conclusion and future works

5.1   Conclusion

In the concentrated biomolecules system, there is the potential for many types of inter-molecular interactions. We have found that the structure of these systems can be described in a hierarchical manner such that relatively strong interactions, form the molecular complexes, while weak interaction drives the many body effect. In the SUMO/SIM system, strong SUMO/SIM interactions form zipper-like network, whereas weak cross-links in defect sites drive phase separation[44;45]. In human mAbs, strong interactions between domains form the elongated complexes, whereas entanglement causes the sharp rise in viscosity. In the nucleolus, strong interaction among the N-terminal domains of NPM1 drives NPM1 pentamers whereas weak interaction (NPM1-SURF6, NPM1-rRNA) drives the phase separation and facilitates the formation of pre-ribosomal sub units.

In chapter two, we discussed the structure-function properties in bio-molecular condensates. There are large number of membrane-less organelles inside the cell which are formed by the condensation of large number of molecules. A small number of molecules, called "scaffolds" produce a phase separated network with liquid-like properties, and these networks recruit other molecules, called "clients". We studied the synthetic system of phase separating molecules designed by Rosen and co-workers[7]. There are two types of scaffolds

(polySUMO/polySIM) and two types of clients(SUMO/SIM). Since the condensed phase a lower density than we would expect for a random network, we propose that molecules are aligned to form a zipper structure and these zippers have defects in the bonding structure that allow for the subsequent formation of a network and client recruitment. Our theory predicts the non-monotonic client partitioning behavior similar to that are seen in experiments. At equal scaffold concentrations, even N-mer complexes are more favorable than odd N-mers in the condensates. At unequal scaffold concentrations, odd N-mers are more favorable than even N-mers whereas at high scaffold stoichiometric mismatches, shorter odd N-mers are more favorable and excess monomer starts to accumulates in the solution. The filamentous structure constrains the scaffold stoichiometries and the distribution of client recruitment sites that the network can accommodate. The filamentous micro-structure of the droplet is sensitive to the valence and binding affinites of clients, which provides independent mechanisms to tune the magnitude (via the affinity) and the location (via the valence) of client recruitment. With tuned client affinities, it is possible for the network to specifically select between closely related clients in the different regimes of parameter space. This confirms that even though condensate are liquid droplets, they have enough structure to bind different valence clients and the change in body composition due to different scaffolds/proportions which allows the system to change the functionality of the droplets.

In chapter three, we discussed the concurrent pre-ribosomal assembly and rRNA folding process in the nucleolus. The nucleolus is the largest condensate in the nucleoplasm which is the site of DNA transcription, rRNA pre-processing, and pre-ribosomal assembly[16–18;46]. We applied our modeling techniques to the nucleolus to determine its internal structure and understand how this structure relates to the ribosome assembly function. The outermost layer of nucleolus is the site of pre-ribosomal assembly and rRNA folding. The primary protein for these processes is NPM1. We modeled the interaction between the NPM1 and positively charged protein like SURF6. Phase separation is a competition between collapses of oppositely charged blocks and the long range repulsion. We also modeled the folding and unfolding of rRNA in early assembly. We found that NPM1 acts as a chaperone to fully unzip the misfolded state so that rRNA re-attempts to fold. In our model, we have developed

(a) matrix method which accounts the probability of unzipping misfolded rRNA due to chaperone. (b) First passage calculation (which accounts the time taken by misfolded rRNA to fully unzip with chaperone assistance) (c) Random walk method: In this computation, we account total number of discrete time step required to unzip the misfolded rRNA in case of chaperone assistance). We have found that the chaperone NPM1 helps to unfold misfolded rRNA by lowering the zipper barrier and the middle of the rRNA zipper is the best position of chaperone to unzip the rRNA zipper.

In chapter four, we discussed the importance of dilute solution measurements to predict antibody viscosity. The concentrated solutions required for efficient dosing possesses high viscosity that creates problems in drug formulation and subcutaneous delivery. We present a physics-based model to predict viscosity at early stage of development which promises to reduce the cost of drug development. Our model assumes that interaction between antibody variable domains results in the formation of elongated complexes and these complexes entangle with each other[25;26]. Our entanglement model accounts for the reptation dynamics of these complexes and shows that the viscosity not only depends on concentration, but also on the binding affinity of molecules. Therefore, knowing the binding affinity of molecules assists in predicting antibody viscosity as a function of concentration. We have used the data from dilute solution measurements to predict the binding affinity of molecules. The dilute solution experiments (ACSINS and DLS) measure the strength of self association of molecules[104]. Our entanglement theory determines the binding affinity of each molecules from measured viscosity values in sample data set. We plot the regression line between these binding affinities and the corresponding dilute solution measurements. These regression lines are used to predict the binding affinity of molecules from ACSINS and DLS measurements. Our entanglement theory then predicts the viscosity from these binding affinities. Our model enables the prediction of antibody viscosity from dilute solution measurements. We have proposed physical mechanism for the outliers as well as experimental modification to account for these mechanisms in the future.

5.2 Future works

We can extend our work on SUMO/SIM droplets to study the formation of other cellular condensates inside the cell and their structure-function relationships. In SUMO/SIM system, the size of linker length is fixed (which is 12 amino acids) and both scaffolds have same length (deccavalent scaffold). We can extend our work to study how variable linker length and unequal valence scaffolds affect condensate formation and the functionality of droplets. In the nucleolus project we observe only the phase separation due to hetero-type NPM1-SURF6 interactions. We can extend our work to understand how homotype (NPM1-NPM1) interactions with the aid of crowding agents, drives phase separation. Also, our three (matrix, analytic, random walk) methods tells us that NPM1 chaperone lowers the zipper barrier in misfolded rRNA. We can test the prediction of the model using a computer simulation that more closely mimics the real system. In the antibody project, we have proposed physical mechanisms for outlier molecules in the dilute solution measurements and experiments to test for similar molecules. Experiment testing these proposed mechanisms will allow us to analyze more about the outliers and it will significantly contribute to better viscosity prediction. We can also work on improving viscosity and stability of antibody solutions in different concentration regimes.

Bibliography

[1] Maad Shatnawi. Chapter 6 - review of recent protein-protein interaction techniques. In Quoc Nam Tran and Hamid Arabnia, editors, *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, Emerging Trends in Computer Science and Applied Computing, pages 99–121. Morgan Kaufmann, Boston, 2015. ISBN 978-0-12-802508-6. doi: https://doi.org/10.1016/B978-0-12-802508-6.00006-5. URL `https://www.sciencedirect.com/science/article/pii/B9780128025086000065`.

[2] Simon Alberti. Phase separation in biology. *Current Biology*, 27(20):R1097 – R1102, 2017. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2017.08.069. URL `http://www.sciencedirect.com/science/article/pii/S0960982217311090`.

[3] Clifford P. Brangwynne, Peter Tompa, and Rohit V. Pappu. Polymer physics of intracellular phase transitions. *Nature Physics*, 11:899–904, 2015. ISSN 1745-2481. doi: 10.1038/nphys3532. URL `https://doi.org/10.1038/nphys3532`.

[4] Tam T.M. Phan and Jeremy D. Schmit. Thermodynamics of huntingtin aggregation. *Biophysical Journal*, 118(12):2989–2996, 2020. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2020.05.013. URL `https://www.sciencedirect.com/science/article/pii/S0006349520304148`.

[5] Yongdae Shin and Clifford P. Brangwynne. Liquid phase condensation in cell physiology and disease. *Science*, 357(6357), 2017. ISSN 0036-8075. doi: 10.1126/science.aaf4382. URL `https://science.sciencemag.org/content/357/6357/eaaf4382`.

[6] Tien M. Phan, Stephen Whitelam, and Jeremy D. Schmit. Catalystlike role of impurities in speeding layer-by-layer growth. *Phys. Rev. E*, 100:042114, Oct 2019.

doi: 10.1103/PhysRevE.100.042114. URL `https://link.aps.org/doi/10.1103/PhysRevE.100.042114`.

[7] Salman F. Banani, Allyson M. Rice, William B. Peeples, Yuan Lin, Saumya Jain, Roy Parker, and Michael K. Rosen. Compositional control of phase-separated cellular bodies. *Cell*, 166(3):651–663, 2016. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2016.06.010. URL `https://www.sciencedirect.com/science/article/pii/S0092867416307395`.

[8] Jeremy D. Schmit, Jill J. Bouchard, Erik W. Martin, and Tanja Mittag. Protein network structure enables switching between liquid and gel states. *Journal of the American Chemical Society*, 142(2):874–883, 2020. ISSN 0002-7863. doi: 10.1021/jacs.9b10066. URL `https://doi.org/10.1021/jacs.9b10066`. PMID: 31845799.

[9] Yuntao S. Mao, Bin Zhang, and David L. Spector. Biogenesis and function of nuclear bodies. *Trends in Genetics*, 27(8):295 – 306, 2011. ISSN 0168-9525. doi: https://doi.org/10.1016/j.tig.2011.05.006. URL `http://www.sciencedirect.com/science/article/pii/S016895251100076X`.

[10] Avinash Patel, Hyun O. Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y. Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M. Franzmann, and et al. A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell*, 162(5):1066 – 1077, 2015. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2015.07.047. URL `http://www.sciencedirect.com/science/article/pii/S0092867415009630`.

[11] Edward Gomes and James Shorter. The molecular language of membraneless organelles. *Journal of Biological Chemistry*, 294(18):7115–7127, 2019. ISSN 0021-9258. doi: https://doi.org/10.1074/jbc.TM118.001192. URL `https://www.sciencedirect.com/science/article/pii/S0021925820367910`.

[12] Steven Boeynaems, Simon Alberti, Nicolas L. Fawzi, Tanja Mittag, Magdalini Poly-

menidou, Frederic Rousseau, Joost Schymkowitz, James Shorter, Benjamin Wolozin, Ludo Van Den Bosch, and et al. Protein phase separation: A new phase in cell biology. *Trends in Cell Biology*, 28(6):420 – 435, 2018. ISSN 0962-8924. doi: https://doi.org/10.1016/j.tcb.2018.02.004. URL `http://www.sciencedirect.com/science/article/pii/S096289241830028X`.

[13] Peter St George-Hyslop, Julie Qiaojin Lin, Akinori Miyashita, Emma C. Phillips, Seema Qamar, Suzanne J. Randle, and GuoZhen Wang. The physiological and pathological biophysics of phase separation and gelation of rna binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration. *Brain Research*, 1693: 11 – 23, 2018. ISSN 0006-8993. doi: https://doi.org/10.1016/j.brainres.2018.04.036. URL `http://www.sciencedirect.com/science/article/pii/S0006899318302312`. RNA Metabolism in Neurological Disease 2018.

[14] Marina Feric, Nilesh Vaidya, Tyler S. Harmon, Diana M. Mitrea, Lian Zhu, Tiffany M. Richardson, Richard W. Kriwacki, Rohit V. Pappu, and Clifford P. Brangwynne. Coexisting liquid phases underlie nucleolar subcompartments. *Cell*, 165(7):1686 – 1697, 2016. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2016.04.047. URL `http://www.sciencedirect.com/science/article/pii/S0092867416304925`.

[15] Xiaolei Su, Jonathon A. Ditlev, Enfu Hui, Wenmin Xing, Sudeep Banjade, Julia Okrut, David S. King, Jack Taunton, Michael K. Rosen, and Ronald D. Vale. Phase separation of signaling molecules promotes t cell receptor signal transduction. *Science*, 352(6285):595–599, 2016. ISSN 0036-8075. doi: 10.1126/science.aad9964. URL `https://science.sciencemag.org/content/352/6285/595`.

[16] Lian Zhu, Tiffany M. Richardson, Ludivine Wacheul, Ming-Tzo Wei, Marina Feric, Gena Whitney, Denis L. J. Lafontaine, and Clifford P. Brangwynne. Controlling the material properties and rrna processing function of the nucleolus using light. *Proceedings of the National Academy of Sciences*, 116(35):17330–17335, 2019. ISSN 0027-

8424. doi: 10.1073/pnas.1903870116. URL `https://www.pnas.org/content/116/35/17330`.

[17] François-Michel Boisvert, Silvana van Koningsbruggen, Joaquín Navascués, and Angus I. Lamond. The multifunctional nucleolus. *Nature Reviews Molecular Cell Biology*, 8:574–585, 2007.

[18] Marie-Line Dubois and Francois-Michel Boisvert. *The Nucleolus: Structure and Function.* Springer International Publishing, Cham, 2016. ISBN 978-3-319-38882-3. doi: 10.1007/978-3-319-38882-3_2. URL `https://doi.org/10.1007/978-3-319-38882-3_2`.

[19] Carl C Correll, Jiri Bartek, and Miroslav Dundr. The nucleolus: a multiphase condensate balancing ribosome synthesis and translational capacity in health, aging and ribosomopathies. *Cells*, 8(8):869, 2019.

[20] Albert E Dahlberg. The functional role of ribosomal rna in protein synthesis. *Cell*, 57(4):525–529, 1989.

[21] Krishan Maggon. Monoclonal antibody "gold rush". *Current Medicinal Chemistry*, 14(18):1978–1987, 2007. ISSN 0929-8673/1875-533X. doi: 10.2174/092986707781368504. URL `http://www.eurekaselect.com/node/59601/article`.

[22] Janice M. Reichert. Monoclonal antibodies as innovative therapeutics. *Current Pharmaceutical Biotechnology*, 9(6):423–430, 2008. ISSN 1389-2010/1873-4316. doi: 10.2174/138920108786786358. URL `http://www.eurekaselect.com/node/68138/article`.

[23] LH Bookbinder, A Hofer, MF Haller, ML Zepeda, G-A Keller, JE Lim, TS Edgington, HM Shepard, JS Patton, and GI Frost. A recombinant human enzyme for enhanced interstitial transport of therapeutics. *Journal of Controlled Release*, 114(2):230–241, 2006.

[24] Dheeraj S. Tomar, Sandeep Kumar, Satish K. Singh, Sumit Goswami, and Li Li. Molecular basis of high viscosity in concentrated antibody solutions: Strategies for high

concentration drug product development. *mAbs*, 8(2):216–228, 2016. doi: 10.1080/ 19420862.2015.1128606. URL `https://doi.org/10.1080/19420862.2015.1128606`. PMID: 26736022.

[25] Nelson Ramallo, Subhash Paudel, and Jeremy Schmit. Cluster formation and entanglement in the rheology of antibody solutions. *The Journal of Physical Chemistry B*, 123(18):3916–3923, 2019. doi: 10.1021/acs.jpcb.9b01511. URL `https://doi.org/10.1021/acs.jpcb.9b01511`. PMID: 30986054.

[26] Jeremy D. Schmit, Feng He, Shradha Mishra, Randal R. Ketchem, Christopher E. Woods, and Bruce A. Kerwin. Entanglement model of antibody viscosity. *The Journal of Physical Chemistry B*, 118(19):5044–5049, 2014. doi: 10.1021/jp500434b. URL `https://doi.org/10.1021/jp500434b`. PMID: 24758234.

[27] Sandeep Yadav, Steven J. Shire, and Devendra S. Kalonia. Factors affecting the viscosity in high concentration solutions of different monoclonal antibodies. *Journal of Pharmaceutical Sciences*, 99(12):4812–4829, 2010. ISSN 0022-3549. doi: https://doi. org/10.1002/jps.22190. URL `https://www.sciencedirect.com/science/article/pii/S0022354915323662`.

[28] Zhenhuan Zhang and Yun Liu. Recent progresses of understanding the viscosity of concentrated protein solutions. *Current Opinion in Chemical Engineering*, 16:48–55, 2017. ISSN 2211-3398. doi: https://doi.org/10.1016/j.coche.2017.04.001. URL `https://www.sciencedirect.com/science/article/pii/S2211339816301198`. Nanotechnology / Separation Engineering.

[29] Ming-Tzo Wei, Shana Elbaum-Garfinkle, Alex S. Holehouse, Carlos Chih-Hsiung Chen, Marina Feric, Craig B. Arnold, Rodney D. Priestley, Rohit V. Pappu, and Clifford P. Brangwynne. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry*, 9:1118–1125, 2017. ISSN 1755-4349. doi: 10.1038/nchem.2803. URL `https://doi.org/10.1038/nchem.2803`.

[30] Peiguo Yang, Cécile Mathieu, Regina-Maria Kolaitis, Peipei Zhang, James Messing, Ugur Yurtsever, Zemin Yang, Jinjun Wu, Yuxin Li, Qingfei Pan, and et al. G3bp1 is a tunable switch that triggers phase separation to assemble stress granules. *Cell*, 181 (2):325 – 345.e28, 2020. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2020.03. 046. URL `http://www.sciencedirect.com/science/a\left(\right)rticle/pii/S0092867420303391`.

[31] Wenmin Xing, Denise Muhlrad, Roy Parker, and Michael K. Rosen. A quantitative inventory of yeast p body proteins reveals principles of compositional specificity. *bioRxiv*, 2018. doi: 10.1101/489658. URL `https://www.biorxiv.org/content/early/2018/12/07/489658`.

[32] Lindsay B. Case, Jonathon A. Ditlev, and Michael K. Rosen. Regulation of transmembrane signaling by phase separation. *Annual Review of Biophysics*, 48(1):465–494, 2019. doi: 10.1146/annurev-biophys-052118-115534. URL `https://doi.org/10.1146/annurev-biophys-052118-115534`. PMID: 30951647.

[33] Yi-Hsuan Lin, Jianhui Song, Julie D. Forman-Kay, and Hue Sun Chan. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids*, 228: 176 – 193, 2017. ISSN 0167-7322. doi: https://doi.org/10.1016/j.molliq.2016.09.090. URL `http://www.sciencedirect.com/science/article/pii/S0167732216322449`. From simple liquids to macromolecular solutions: recent experimental and theoretical developments. In Honor of the 70th birthday of Vojko Vlachy.

[34] Tyler S Harmon, Alex S Holehouse, Michael K Rosen, and Rohit V Pappu. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife*, 6:e30294, nov 2017. ISSN 2050-084X. doi: 10.7554/eLife. 30294. URL `https://doi.org/10.7554/eLife.30294`.

[35] Jie Wang, Jeong-Mo Choi, Alex S. Holehouse, Hyun O. Lee, Xiaojie Zhang, Marcus Jahnel, Shovamayee Maharana, Régis Lemaitre, Andrei Pozniakovsky, David Drechsel,

and et al. A molecular grammar governing the driving forces for phase separation of prion-like rna binding proteins. *Cell*, 174(3):688 – 699.e16, 2018. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2018.06.006. URL `http://www.sciencedirect.com/science/article/pii/S0092867418307311`.

[36] Alexander N. Semenov and Michael Rubinstein. Thermoreversible gelation in solutions of associative polymers.1.statistics. *Macromolecules*, 31:1373–1385, 1998. ISSN 0024-9297. doi: 10.1021/ma970616h. URL `https://doi.org/10.1021/ma970616`.

[37] Jeong-Mo Choi, Alex S. Holehouse, and Rohit V. Pappu. Physical principles underlying the complex biology of intracellular phase transitions. *Annual Review of Biophysics*, 49(1):107–133, 2020. doi: 10.1146/annurev-biophys-121219-081629. URL `https://doi.org/10.1146/annurev-biophys-121219-081629`. PMID: 32004090.

[38] B. H. Zimm and J. K. Bragg. Theory of the phase transition between helix and random coil in polypeptide chains. *The Journal of Chemical Physics*, 31(2):526–535, 1959. doi: 10.1063/1.1730390. URL `https://doi.org/10.1063/1.1730390`.

[39] J B Pendry. Transfer matrices and conductivity in two- and three-dimensional systems. i. formalism. *Journal of Physics: Condensed Matter*, 2(14):3273–3286, apr 1990. doi: 10.1088/0953-8984/2/14/012. URL `https://doi.org/10.1088`.

[40] Elizabeth S. Freeman Rosenzweig, Bin Xu, Luis Kuhn Cuellar, Antonio Martinez-Sanchez, Miroslava Schaffer, Mike Strauss, Heather N. Cartwright, Pierre Ronceray, Jürgen M. Plitzko, Friedrich Förster, and et al. The eukaryotic co2-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell*, 171(1):148 – 162.e19, 2017. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2017.08.008. URL `http://www.sciencedirect.com/science/article/pii/S0092867417309339`.

[41] Bin Xu, Guanhua He, Benjamin G Weiner, Pierre Ronceray, Yigal Meir, Martin C Jonikas, and Ned S Wingreen. Rigidity enhances a magic-number effect in polymer

phase separation. *Nat Commun*, 11:1561, 2020 Mar 25 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15395-6.

[42] Melissa R Marzahn, Suresh Marada, Jihun Lee, Amanda Nourse, Sophia Kenrick, Huaying Zhao, Gili Ben-Nissan, Regina-Maria Kolaitis, Jennifer L Peters, Stanley Pounds, and et al. Higher-order oligomerization promotes localization of spop to liquid nuclear speckles. *The EMBO Journal*, 35(12):1254–1275, 6 2016. ISSN 1460-2075. doi: 10.15252/embj.201593169. URL `https://doi.org/10.15252/embj.201593169`.

[43] Jill J. Bouchard, Joel H. Otero, Daniel C. Scott, Elzbieta Szulc, Erik W. Martin, Nafiseh Sabri, Daniele Granata, Melissa R. Marzahn, Kresten Lindorff-Larsen, Xavier Salvatella, and et al. Cancer mutations of the tumor suppressor spop disrupt the formation of active, phase-separated compartments. *Molecular Cell*, 72(1):19 – 36.e8, 2018. ISSN 1097-2765. doi: https://doi.org/10.1016/j.molcel.2018.08.027. URL `http://www.sciencedirect.com/science/article/pii/S1097276518306877`.

[44] Kamal Bhandari, Michael A. Cotten, Jonggul Kim, Michael K. Rosen, and Jeremy D. Schmit. Structure–function properties in disordered condensates. *The Journal of Physical Chemistry B*, 125(1):467–476, 2021. doi: 10.1021/acs.jpcb.0c11057. URL `https://doi.org/10.1021/acs.jpcb.0c11057`. PMID: 33395293.

[45] Jeremy D. Schmit, Marina Feric, and Miroslav Dundr. How hierarchical interactions make membraneless organelles tick like clockwork. *Trends in Biochemical Sciences*, 46(7):525–534, 2021. ISSN 0968-0004. doi: https://doi.org/10.1016/j.tibs.2020.12.011. URL `https://www.sciencedirect.com/science/article/pii/S0968000420303236`.

[46] Diana M. Mitrea, Jaclyn A. Cika, Christopher B. Stanley, Amanda Nourse, Paulo L. Onuchic, Priya R. Banerjee, Aaron H. Phillips, Cheon-Gil Park, Ashok A. Deniz, and Richard W. Kriwacki. Self-interaction of npm1 modulates multiple mechanisms of liquid-liquid phase separation. *Nature Communications*, 2018.

[47] Danièle Hernandez-Verdun. Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus*, 2(3):189–194, 2011. doi: 10.4161/nucl.2.3.16246. URL `https://doi.org/10.4161/nucl.2.3.16246`. PMID: 21818412.

[48] Asen A Hadjiolov. *The nucleolus and ribosome biogenesis*, volume 12. Springer Science & Business Media, 2012.

[49] Angus I Lamond and David L Spector. Nuclear speckles: a model for nuclear organelles. *Nature reviews Molecular cell biology*, 4(8):605–612, 2003.

[50] DL Spector and AI Lamond. Nuclear speckles. cold spring harb perspect biol 3: a000646. *Either first page or author must be supplied. AMBIGUOUS (137 citations)*, 2011.

[51] Archa H Fox and Angus I Lamond. Paraspeckles. *Cold Spring Harbor perspectives in biology*, 2(7):a000687, 2010.

[52] Archa H Fox, Yun Wah Lam, Anthony KL Leung, Carol E Lyon, Jens Andersen, Matthias Mann, and Angus I Lamond. Paraspeckles: a novel nuclear domain. *Current Biology*, 12(1):13–25, 2002.

[53] Valérie Lallemand-Breitenbach et al. Pml nuclear bodies. *Cold Spring Harbor perspectives in biology*, 2(5):a000661, 2010.

[54] Elizabeth C Batty, Kirsten Jensen, and Paul S Freemont. Pml nuclear bodies and other trim-defined subcellular compartments. *TRIM/RBCC Proteins*, pages 39–58, 2012.

[55] Mario Cioce and Angus I Lamond. Cajal bodies, a long history of discovery. *Annu. Rev. Cell Dev. Biol.*, 21:105–131, 2005.

[56] Anthony A Hyman and Clifford P Brangwynne. Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. *Developmental cell*, 21(1):14–16, 2011.

[57] Sudeep Banjade and Michael K Rosen. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *Elife*, 3:e04123, 2014.

[58] Pilong Li, Sudeep Banjade, Hui-Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, Javoris V Hollingsworth, David S King, Salman F Banani, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483(7389):336–340, 2012.

[59] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18:285–298, 2017.

[60] Meiyan Jin, Gregory G. Fuller, Ting Han, Yao Yao, Amelia F. Alessi, Mallory A. Freeberg, Nathan P. Roach, James J. Moresco, Alla Karnovsky, Misuzu Baba, John R. Yates, Aaron D. Gitler, Ken Inoki, Daniel J. Klionsky, and John K. Kim. Glycolytic enzymes coalesce in g bodies under hypoxic stress. *Cell Reports*, 20(4):895–908, 2017. ISSN 2211-1247. doi: https://doi.org/10.1016/j.celrep.2017.06.082. URL `https://www.sciencedirect.com/science/article/pii/S2211124717309300`.

[61] J.B. Woodruff, B. Ferreira Gomes, P.O. Widlund, J. Mahamid, A. Honigmann, and A.A. Hyman. The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. *Cell*, 169(6):1066–1077.e10, 2017. doi: 10.1016/j.cell.2017.05.028. URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85020053848&doi=10.1016`. cited By 240.

[62] Amayra Hernández-Vega, Marcus Braun, Lara Scharrel, Marcus Jahnel, Susanne Wegmann, Bradley T. Hyman, Simon Alberti, Stefan Diez, and Anthony A. Hyman. Local nucleation of microtubule bundles through tubulin concentration into a condensed tau phase. *Cell Reports*, 20(10):2304–2312, 2017. ISSN 2211-1247. doi: https://doi.org/10.1016/j.celrep.2017.08.042. URL `https://www.sciencedirect.com/science/article/pii/S221112471731149X`.

[63] S. Gueroussov, R.J. Weatheritt, D. O'Hanlon, Z.-Y. Lin, A. Narula, A.-C. Gingras, and B.J. Blencowe. Regulatory expansion in mammals of multivalent hnrnp assemblies that globally control alternative splicing. *Cell*, 170(2):324–339.e23, 2017. doi: 10.1016/ j.cell.2017.06.037. URL `https://www.scopus.com/inward/record.uri?eid=2-s2. 0-85023645460&doi=10.1016.` cited By 55.

[64] M. Arrasate, S. Mitra, E.S. Schweitzer, M.R. Segal, and S. Finkbeiner. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, 431(7010):805–810, 2004. doi: 10.1038/nature02998. URL `https://www.scopus.com/ inward/record.uri?eid=2-s2.0-7244236320&doi=10.1038.` cited By 1524.

[65] Claudia M. Cremers, Daniela Knoefler, Stephanie Gates, Nicholas Martin, Jan-Ulrik Dahl, Justine Lempart, Lihan Xie, Matthew R. Chapman, Veronica Galvan, Daniel R. Southworth, and Ursula Jakob. Polyphosphate: A conserved modifier of amyloidogenic processes. *Molecular Cell*, 63(5):768–780, 2016. ISSN 1097-2765. doi: https://doi. org/10.1016/j.molcel.2016.07.016. URL `https://www.sciencedirect.com/science/ article/pii/S1097276516303690.`

[66] Edward W.J. Wallace, Jamie L. Kear-Scott, Evgeny V. Pilipenko, Michael H. Schwartz, Pawel R. Laskowski, Alexandra E. Rojek, Christopher D. Katanski, Joshua A. Riback, Michael F. Dion, Alexander M. Franks, Edoardo M. Airoldi, Tao Pan, Bogdan A. Budnik, and D. Allan Drummond. Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. *Cell*, 162(6):1286–1298, 2015. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2015.08.041. URL `https://www.sciencedirect.com/ science/article/pii/S0092867415010934.`

[67] Mylene C Ferrolino, Diana M Mitrea, J Robert Michael, and Richard W Kriwacki. Compositional adaptability in npm1-surf6 scaffolding networks enabled by dynamic switching of phase separation mechanisms. *Nature communications*, 9(1):1–11, 2018.

[68] Hanieh Falahati, Bobbie Pelham-Webb, Shelby Blythe, and Eric Wieschaus. Nucle-

ation by rrna dictates the precision of nucleolus assembly. *Current Biology*, 26(3): 277–285, 2016.

[69] Alice Grob and Brian McStay. Construction of synthetic nucleoli and what it tells us about propagation of sub-nuclear domains through cell division. *Cell cycle*, 13(16): 2501–2508, 2014.

[70] Diana M Mitrea, Jaclyn A Cika, Clifford S Guy, David Ban, Priya R Banerjee, Christopher B Stanley, Amanda Nourse, Ashok A Deniz, and Richard W Kriwacki. Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying r-rich linear motifs and rrna. *eLife*, 5:e13571, feb 2016. ISSN 2050-084X. doi: 10.7554/eLife.13571. URL https://doi.org/10.7554/eLife.13571.

[71] Mikael S Lindström. Npm1/b23: a multifunctional chaperone in ribosome biogenesis and chromatin remodeling. *Biochemistry research international*, 2011, 2011.

[72] Ludmila G Romanova, Martin Anger, Olga V Zatsepina, and Richard M Schultz. Implication of nucleolar protein surf6 in ribosome biogenesis and preimplantation mouse development. *Biology of reproduction*, 75(5):690–696, 2006.

[73] Kensaku Murano, Mitsuru Okuwaki, Miharu Hisaoka, and Kyosuke Nagata. Transcription regulation of the rrna gene by a multifunctional nucleolar protein, b23/nucleophosmin, through its histone chaperone activity. *Molecular and Cellular Biology*, 28(10):3114–3126, 2008. doi: 10.1128/MCB.02078-07.

[74] Leonard B. Maggi, Michael Kuchenruether, David Y. A. Dadey, Rachel M. Schwope, Silvia Grisendi, R. Reid Townsend, Pier Paolo Pandolfi, and Jason D. Weber. Nucleophosmin serves as a rate-limiting nuclear export chaperone for the mammalian ribosome. *Molecular and Cellular Biology*, 28(23):7050–7065, 2008. doi: 10.1128/ MCB.01548-07.

[75] Casey Lee, Brian A Smith, Keya Bandyopadhyay, and Ruth A Gjerset. Dna damage

disrupts the p14arf-b23 (nucleophosmin) interaction and triggers a transient subnuclear redistribution of p14arf. *Cancer research*, 65(21):9834–9842, 2005.

[76] Daniel D. Scott and Marlene Oeffinger. Nucleolin and nucleophosmin: nucleolar proteins with multiple functions in dna repair. *Biochemistry and Cell Biology*, 94 (5):419–432, 2016. doi: 10.1139/bcb-2016-0068. URL `https://doi.org/10.1139/bcb-2016-0068`. PMID: 27673355.

[77] Diana M. Mitrea, Christy R. Grace, Marija Buljan, Mi-Kyung Yun, Nicholas J. Pytel, John Satumba, Amanda Nourse, Cheon-Gil Park, M. Madan Babu, Stephen W. White, and Richard W. Kriwacki. Structural polymorphism in the n-terminal oligomerization domain of npm1. *Proceedings of the National Academy of Sciences*, 111(12):4466–4471, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1321007111. URL `https://www.pnas.org/content/111/12/4466`.

[78] Jose M Eirin-Lopez, Lindsay J Frehlick, and Juan Ausio. Long-Term Evolution and Functional Diversification in the Members of the Nucleophosmin/Nucleoplasmin Family of Nuclear Chaperones. *Genetics*, 173(4):1835–1850, 08 2006. ISSN 1943-2631. doi: 10.1534/genetics.106.058990. URL `https://doi.org/10.1534/genetics.106.058990`.

[79] Anastasiia Moraleva, Charalambos Magoulas, Mikhail Polzikov, Sabine Hacot, Hichem C Mertani, Jean-Jacques Diaz, and Olga Zatsepina. Involvement of the specific nucleolar protein surf6 in regulation of proliferation and ribosome biogenesis in mouse nih/3t3 fibroblasts. *Cell Cycle*, 16(20):1979–1991, 2017.

[80] Miharu Hisaoka, Kyosuke Nagata, and Mitsuru Okuwaki. Intrinsically disordered regions of nucleophosmin/B23 regulate its RNA binding activity through their inter- and intra-molecular association. *Nucleic Acids Research*, 42(2):1180–1195, 10 2013. ISSN 0305-1048. doi: 10.1093/nar/gkt897. URL `https://doi.org/10.1093/nar/gkt897`.

[81] J.L. Camberg, S.M. Doyle, D.M. Johnston, and S. Wickner. Molecular chaper-

ones. In Stanley Maloy and Kelly Hughes, editors, *Brenner's Encyclopedia of Genetics (Second Edition)*, pages 456–460. Academic Press, San Diego, second edition edition, 2013. ISBN 978-0-08-096156-9. doi: https://doi.org/10.1016/B978-0-12-374984-0.00221-7. URL `https://www.sciencedirect.com/science/article/pii/B9780123749840002217`.

[82] Martina Beissinger and J Buchner. How chaperones fold proteins. *Biological chemistry*, 379(3):245–259, 1998.

[83] Catalin M. Filipeanu. Chapter eleven - temperature-sensitive intracellular traffic of alpha2c-adrenergic receptor. In Guangyu Wu, editor, *Trafficking of GPCRs*, volume 132 of *Progress in Molecular Biology and Translational Science*, pages 245–265. Academic Press, 2015. doi: https://doi.org/10.1016/bs.pmbts.2015.02.008. URL `https://www.sciencedirect.com/science/article/pii/S187711731500037X`.

[84] Takaho Terada and Shigeyuki Yokoyama. Chapter thirteen - escherichia coli cell-free protein synthesis and isotope labeling of mammalian proteins. In Zvi Kelman, editor, *Isotope Labeling of Biomolecules - Labeling Methods*, volume 565 of *Methods in Enzymology*, pages 311–345. Academic Press, 2015. doi: https://doi.org/10.1016/bs.mie.2015.08.035. URL `https://www.sciencedirect.com/science/article/pii/S0076687915005029`.

[85] R.B. Wickner, H. Edskes, T. Nakayashiki, F. Shewmaker, L. McCann, A. Engel, and D. Kryndushkin. Prions of yeast and fungi. In Brian W.J. Mahy and Marc H.V. Van Regenmortel, editors, *Encyclopedia of Virology (Third Edition)*, pages 336–341. Academic Press, Oxford, third edition edition, 2008. ISBN 978-0-12-374410-4. doi: https://doi.org/10.1016/B978-012374410-4.00407-6. URL `https://www.sciencedirect.com/science/article/pii/B9780123744104004076`.

[86] A.J.L. Macario and E. Conway de Macario. Chaperone proteins and chaperonopathies. In George Fink, editor, *Encyclopedia of Stress (Second Edition)*, pages 438–444.

Academic Press, New York, second edition edition, 2007. ISBN 978-0-12-373947-6. doi: https://doi.org/10.1016/B978-012373947-6.00075-1. URL `https://www.sciencedirect.com/science/article/pii/B9780123739476000751`.

[87] Mark R. Woodford, Diana Dunn, Jonelle B. Miller, Sami Jamal, Len Neckers, and Mehdi Mollapour. Chapter two - impact of posttranslational modifications on the anticancer activity of hsp90 inhibitors. In Jennifer Isaacs and Luke Whitesell, editors, *Hsp90 in Cancer: Beyond the Usual Suspects*, volume 129 of *Advances in Cancer Research*, pages 31–50. Academic Press, 2016. doi: https://doi.org/10.1016/bs.acr.2015.09.002. URL `https://www.sciencedirect.com/science/article/pii/S0065230X15000949`.

[88] Daniel Jeffery, Katrina Podsypanina, Tejas Yadav, and Geneviève Almouzni. Chromatin dynamics in cancer: Epigenetic parameters and cellular fate. In Paolo Boffetta and Pierre Hainaut, editors, *Encyclopedia of Cancer (Third Edition)*, pages 372–388. Academic Press, Oxford, third edition edition, 2019. ISBN 978-0-12-812485-7. doi: https://doi.org/10.1016/B978-0-12-801238-3.65276-5. URL `https://www.sciencedirect.com/science/article/pii/B9780128012383652765`.

[89] BW Pontius and Paul Berg. Rapid assembly and disassembly of complementary dna strands through an equilibrium intermediate state mediated by a1 hnrnp protein. *Journal of Biological Chemistry*, 267(20):13815–13818, 1992.

[90] Daniel Herschlag. Rna chaperones and the rna folding problem. *Journal of Biological Chemistry*, 270(36):20871–20874, 1995. ISSN 0021-9258. doi: https://doi.org/10.1074/jbc.270.36.20871. URL `https://www.sciencedirect.com/science/article/pii/S0021925818902863`.

[91] S H Munroe and X F Dong. Heterogeneous nuclear ribonucleoprotein a1 catalyzes rna.rna annealing. *Proceedings of the National Academy of Sciences*, 89(3):895–899, 1992. ISSN 0027-8424. doi: 10.1073/pnas.89.3.895. URL `https://www.pnas.org/content/89/3/895`.

[92] Joshua A. Riback, Christopher D. Katanski, Jamie L. Kear-Scott, Evgeny V. Pilipenko, Alexandra E. Rojek, Tobin R. Sosnick, and D. Allan Drummond. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell*, 168(6):1028 – 1040.e19, 2017. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2017.02.027. URL `http://www.sciencedirect.com/science/article/pii/S0092867417302428`.

[93] Edward A Codling, Michael J Plank, and Simon Benhamou. Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834, 2008. doi: 10.1098/rsif. 2008.0014. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2008.0014`.

[94] Lisha Sheng, Zhenqian Chen, Bo Xu, and Juan Shi. Molecular dynamics study of the dispersion stability and fluidity of porous liquids with different canopy structures. *The Journal of Physical Chemistry B*, 125(20):5387–5396, 2021. doi: 10.1021/acs.jpcb. 1c02450. URL `https://doi.org/10.1021/acs.jpcb.1c02450`. PMID: 33983737.

[95] William J. Galush, Lan N. Le, and Jamie M.R. Moore. Viscosity behavior of high-concentration protein mixtures. *Journal of Pharmaceutical Sciences*, 101(3):1012–1020, 2012. ISSN 0022-3549. doi: https://doi.org/10.1002/jps.23002. URL `https://www.sciencedirect.com/science/article/pii/S0022354915316877`.

[96] Brian D. Connolly, Chris Petry, Sandeep Yadav, Barthélemy Demeule, Natalie Ciaccio, Jamie M.R. Moore, Steven J. Shire, and Yatin R. Gokarn. Weak interactions govern the viscosity of concentrated antibody solutions: High-throughput analysis using the diffusion interaction parameter. *Biophysical Journal*, 103(1):69–78, 2012. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2012.04.047. URL `https://www.sciencedirect.com/science/article/pii/S000634951200522X`.

[97] Ann L. Daugherty and Randall J. Mrsny. Formulation and delivery issues for monoclonal antibody therapeutics. *Advanced Drug Delivery Reviews*, 58(5):686–706, 2006. ISSN 0169-409X. doi: https://doi.org/10.1016/j.addr.2006.03.011. URL

https://www.sciencedirect.com/science/article/pii/S0169409X06000895. Engineered antibody therapeutics.

[98] Branden A. Salinas, Hasige A. Sathish, Steven M. Bishop, Nick Harn, John F. Carpenter, and Theodore W. Randolph. Understanding and modulating opalescence and viscosity in a monoclonal antibody formulation. *Journal of Pharmaceutical Sciences*, 99(1):82–93, 2010. ISSN 0022-3549. doi: https://doi.org/10.1002/jps.21797. URL https://www.sciencedirect.com/science/article/pii/S0022354916303495.

[99] Jun Liu, Mary D.H. Nguyen, James D. Andya, and Steven J. Shire. Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. *Journal of Pharmaceutical Sciences*, 94(9):1928–1940, 2005. ISSN 0022-3549. doi: https://doi.org/10.1002/jps.20347. URL https://www.sciencedirect.com/science/article/pii/S0022354916318524.

[100] Thomas M. Scherer, Jun Liu, Steven J. Shire, and Allen P. Minton. Intermolecular interactions of igg1 monoclonal antibodies at high concentrations characterized by light scattering. *The Journal of Physical Chemistry B*, 114(40):12948–12957, 2010. doi: 10.1021/jp1028646. URL https://doi.org/10.1021/jp1028646. PMID: 20849134.

[101] Shantanu V. Sule, Jason K. Cheung, Valentyn Antochshuk, Amardeep S. Bhalla, Chakravarthy Narasimhan, Steven Blaisdell, Mohammed Shameem, and Peter M. Tessier. Solution ph that minimizes self-association of three monoclonal antibodies is strongly dependent on ionic strength. *Molecular Pharmaceutics*, 9(4):744–751, 2012. doi: 10.1021/mp200448j. URL https://doi.org/10.1021/mp200448j. PMID: 22221144.

[102] Akhilesh Bhambhani, Julian M. Kissmann, Sangeeta B. Joshi, David B. Volkin, Ramesh S. Kashi, and C. Russell Middaugh. Formulation design and high-throughput excipient selection based on structural integrity and conformational stability of dilute and highly concentrated igg1 monoclonal antibody solutions. *Journal of Pharmaceutical Sciences*, 101(3):1120–1135, 2012. ISSN 0022-3549. doi: https://doi.org/

10.1002/jps.23008. URL `https://www.sciencedirect.com/science/article/pii/` `S0022354915316932`.

[103] Steven J. Shire, Zahra Shahrokh, and Jun Liu. Challenges in the development of high protein concentration formulations. *Journal of Pharmaceutical Sciences*, 93(6):1390–1402, 2004. ISSN 0022-3549. doi: https://doi.org/10.1002/jps.20079. URL `https://www.sciencedirect.com/science/article/pii/S0022354916315234`.

[104] Shantanu V. Sule, Craig D. Dickinson, Jirong Lu, Chi-Kin Chow, and Peter M. Tessier. Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. *Molecular Pharmaceutics*, 10(4):1322–1331, 2013. doi: 10.1021/mp300524x. URL `https://doi.org/10.1021/mp300524x`. PMID: 23383873.

[105] Yuqi Liu, Isabelle Caffry, Jiemin Wu, Steven B Geng, Tushar Jain, Tingwan Sun, Felicia Reid, Yuan Cao, Patricia Estep, Yao Yu, Maximiliano Vásquez, Peter M Tessier, and Yingda Xu. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *mAbs*, 6(2):483–492, 2014. doi: 10.4161/mabs.27431. URL `https://doi.org/10.4161/mabs.27431`. PMID: 24492294.

[106] Jiemin Wu, Jason S. Schultz, Caroline L. Weldon, Shantanu V. Sule, Qing Chai, Steven B. Geng, Craig D. Dickinson, and Peter M. Tessier. Discovery of highly soluble antibodies prior to purification using affinity-capture self-interaction nanoparticle spectroscopy. *Protein Engineering, Design and Selection*, 28(10):403–414, 09 2015. ISSN 1741-0126. doi: 10.1093/protein/gzv045. URL `https://doi.org/10.1093/protein/gzv045`.

[107] Isidro E. Zarraga, Rosalynn Taing, Jonathan Zarzar, Jacob Luoma, Jenny Hsiung, Ankit Patel, and Fredric J. Lim. High shear rheology and anisotropy in concentrated solutions of monoclonal antibodies. *Journal of Pharmaceutical Sciences*, 102(8):2538–2549, 2013. ISSN 0022-3549. doi: https://doi.org/10.1002/jps.23647. URL `https://www.sciencedirect.com/science/article/pii/S0022354915310005`.

[108] Eric J. Yearley, Paul D. Godfrin, Tatiana Perevozchikova, Hailiang Zhang, Peter Falus, Lionel Porcar, Michihiro Nagao, Joseph E. Curtis, Pradad Gawande, Rosalynn Taing, Isidro E. Zarraga, Norman J. Wagner, and Yun Liu. Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. *Biophysical Journal*, 106(8):1763–1770, 2014. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2014.02.036. URL `https://www.sciencedirect.com/science/article/pii/S0006349514002835`.

[109] Jayant Arora, John M Hickey, Ranajoy Majumdar, Reza Esfandiary, Steven M Bishop, Hardeep S Samra, C Russell Middaugh, David D Weis, and David B Volkin. Hydrogen exchange mass spectrometry reveals protein interfaces and distant dynamic coupling effects during the reversible self-association of an igg1 monoclonal antibody. In *MAbs*, volume 7, pages 525–539. Taylor and Francis, 2015.

[110] Jayant Arora, Yue Hu, Reza Esfandiary, Hasige A Sathish, Steven M Bishop, Sangeeta B Joshi, C Russell Middaugh, David B Volkin, and David D Weis. Charge-mediated fab-fc interactions in an igg1 antibody induce reversible self-association, cluster formation, and elevated viscosity. In *MAbs*, volume 8, pages 1561–1574. Taylor and Francis, 2016.

[111] Masao Doi and Samuel Frederick Edwards. *The theory of polymer dynamics*, volume 73. oxford university press, 1988.

[112] PG De Gennes. Dynamics of entangled polymer solutions. i. the rouse model. *Macromolecules*, 9(4):587–593, 1976.

[113] PG De Gennes. Dynamics of entangled polymer solutions. ii. inclusion of hydrodynamic interactions. *Macromolecules*, 9(4):594–598, 1976.

[114] Steven B. Geng, Michael Wittekind, Adam Vigil, and Peter M. Tessier. Measurements of monoclonal antibody self-association are correlated with complex biophysical properties. *Molecular Pharmaceutics*, 13(5):1636–1645, 2016. doi: 10.1021/

117

acs.molpharmaceut.6b00071. URL `https://doi.org/10.1021/acs.molpharmaceut.6b00071`. PMID: 27045771.

[115] Pilong Li, Sudeep Banjade, Hui-Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, Javoris V. Hollingsworth, David S. King, Salman F. Banani, and et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483:336–340, 2012. ISSN 1476-4687. doi: 10.1038/nature10879. URL `https://doi.org/10.1038/nature10879`.

[116] Feng He, Gerald W. Becker, Jennifer R. Litowski, Linda O. Narhi, David N. Brems, and Vladimir I. Razinkov. High-throughput dynamic light scattering method for measuring viscosity of concentrated protein solutions. *Analytical Biochemistry*, 399(1):141–143, 2010. ISSN 0003-2697. doi: https://doi.org/10.1016/j.ab.2009.12.003. URL `https://www.sciencedirect.com/science/article/pii/S0003269709008173`.

[117] Hardeep S Samra and Feng He. Advancements in high throughput biophysical technologies: applications for characterization and screening during early formulation development of monoclonal antibodies. *Molecular pharmaceutics*, 9(4):696–707, 2012.

[118] Carl Mieczkowski, Alan Cheng, Thierry Fischmann, Mark Hsieh, Jeanne Baker, Makiko Uchida, Gopalan Raghunathan, Corey Strickland, and Laurence Fayadat-Dilman. Characterization and modeling of reversible antibody self-association provide insights into behavior, prediction, and correction. *Antibodies*, 10(1), 2021. ISSN 2073-4468. doi: 10.3390/antib10010008. URL `https://www.mdpi.com/2073-4468/10/1/8`.

[119] Deepak Langhe and Michael Ponting. 1 - introduction to multilayered films. In Deepak Langhe and Michael Ponting, editors, *Manufacturing and Novel Applications of Multilayer Polymer Films*, pages 1–15. William Andrew Publishing, Boston, 2016. ISBN 978-0-323-37125-4. doi: https://doi.org/10.1016/B978-0-323-37125-4.00001-0. URL `https://www.sciencedirect.com/science/article/pii/B9780323371254000010`.

[120] Atul Saluja, R. Matthew Fesinmeyer, Sabine Hogan, David N. Brems, and Yatin R.

Gokarn. Diffusion and sedimentation interaction parameters for measuring the second virial coefficient and their utility as predictors of protein aggregation. *Biophysical Journal*, 99(8):2657–2665, 2010. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2010.08.020. URL `https://www.sciencedirect.com/science/article/pii/S0006349510009896`.

Appendix A

A.1 Binding affinity parameter

We have some unknown parameters in grand partition function of zipper molecules. To facilitate the comparison of our theory with experiment, we have to reduce these number of unknown parameters.

The total partition function of a molecule which is the product of terms,

$$
\begin{aligned}
Z_1 &= Z_{trans} Z_{momentum} Z_{rot} Z_{elec}....... \\
&= V \lambda^{-3} Z_{int} \\
&= \frac{V}{\lambda^3} Z_{int}
\end{aligned}
\tag{A.1}
$$

where $Z_{trans} = V$ is total volume of system, $Z_{momentum} = \lambda^{-3}$ where λ is the de Broglie wavelength and $Z_{int} = Z_{rot} Z_{elec}.....$

For identical molecules, $Z_N = \frac{Z_1^N}{N!}$

Total free energy for a system of N molecules,

$$
\begin{aligned}
F_N &= -N K_{\mathrm{B}} T \log Z_N \\
&= -N K_{\mathrm{B}} T \log \frac{Z_1^N}{N!} \\
&= -N K_{\mathrm{B}} T (\log \frac{Z_1}{N} + 1) \\
&= -N K_{\mathrm{B}} T (\log \frac{V}{N \lambda^3} + \log Z_{int} + 1) \\
&= -N F_{int} + N K_{\mathrm{B}} T \log \frac{N \lambda^3}{V} - N K_{\mathrm{B}} T
\end{aligned}
\tag{A.2}
$$

And chemical potential of the system of N molecules is

$$
\begin{aligned}
\mu_N &= \frac{\partial F_N}{\partial N} \\
&= F_{int} + K_{\mathrm{B}} T \log c \lambda^3 \\
&= F_{int} + K_{\mathrm{B}} T \log \frac{c}{c_0} + K_{\mathrm{B}} T \log c_0 \lambda^3 \\
&= F_{int} + K_{\mathrm{B}} T \log \frac{c}{c_0} + \mu_0
\end{aligned}
\tag{A.3}
$$

If two molecules bind together and at chemical equilibrium, we can write

$$
\begin{aligned}
\mu_2 &= 2\mu_1 \\
F_2 + K_{\mathrm{B}} T \log \frac{c_2}{c_0} + \mu_0 &= 2F_1 + 2K_{\mathrm{B}} T \log \frac{c_1}{c_0} + 2\mu_0 \\
F_2 - F_1 &= K_{\mathrm{B}} T \log \frac{c_1^2}{c_2 c_0} c_0 \lambda^3 \\
\frac{\Delta F}{K_{\mathrm{B}} T} &= \log \frac{c_1^2}{c_2 c_0} c_0 \lambda^3 \\
\frac{c_1^2}{c_2 c_0} &= K_d = \frac{1}{c_0 \lambda^3} \exp^{\frac{\Delta F}{K_{\mathrm{B}} T}}
\end{aligned}
\tag{A.4}
$$

Finally,

$$
\therefore \log K_d = \Delta F + F_0
\tag{A.5}
$$

where F_0 is a constant belongs to entropic arrangement of system. For two molecules

N	$K_d(\mu M)$
1	10
2	0.7
3	0.07
5	0.001

Table A.1: *Scaffold Dissociation constant*

with same binding sites ℓ, $\Delta F = \exp^{-l\epsilon} + 2\sum_{i=1}^{\ell-1}\exp^{-i\epsilon}$.

TableA.1 shows the experimental data for polySUMO-polySIM dissociation constant for different scaffold valency. The fitted value of unknown parameter through Eq.A.5 using experimental data [TableA.1] are $\epsilon = -2.23K_BT$ and $F_0 = -9.4K_BT$. With these two values $\epsilon = -2.23K_BT$ and $F_0 = -9.4K_BT$, scaffold association constant can be written as

$$K_a = \frac{1}{K_d} = K_s \exp^{-\Delta F} \tag{A.6}$$

where $\epsilon^{-\Delta F_0} = K_s$ and $\Delta F = \exp^{-l\epsilon} + 2\sum_{i=1}^{\ell-1}\exp^{-i\epsilon}$.

After getting fitted value of ϵ and F_0, we have an unknown parameter left. It is client-scaffold binding affinity. We have taken appropriate value of client-scaffold binding affinity through which our theory gives pretty close result with the experiment which are shown in following table.

Client	$K_a(M^{-1})$
SIM_1	4000
SIM_2	77000
SIM_3	625000

Table A.2: *Client binding affinity*

A.2 End Vectors and Transfer Matrices

Total grand partition function (Eq. 2.10) of the system is:

$$
\begin{aligned}
\mathcal{Q} &= c_{\mathrm{U}} + c_{\mathrm{I}} + c_{\mathrm{U}}c_{\mathrm{I}}e^{-n\epsilon} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{UL}}(\mathbf{M}_{\mathrm{I}}\mathbf{M}_{\mathrm{U}})^{N}\mathbf{V}_{\mathrm{UR}} + \sum_{N=0}^{\infty} \mathbf{V}_{\mathrm{IL}}(\mathbf{M}_{\mathrm{U}}\mathbf{M}_{\mathrm{I}})^{N}\mathbf{V}_{\mathrm{IR}} \quad \text{(A.7)} \\
&= c_{\mathrm{U}} + c_{\mathrm{I}} + c_{\mathrm{U}}c_{\mathrm{I}}e^{-n\epsilon} + \mathbf{V}_{\mathrm{UL}}\frac{1}{1 - \mathbf{M}_{\mathrm{I}}\mathbf{M}_{\mathrm{U}}}\mathbf{V}_{\mathrm{UR}} + \mathbf{V}_{\mathrm{IL}}\frac{1}{1 - \mathbf{M}_{\mathrm{U}}\mathbf{M}_{\mathrm{I}}}\mathbf{V}_{\mathrm{IR}} \quad \text{(A.8)}
\end{aligned}
$$

The two summation terms represent filaments starting with SUMO or SIM modules. In our formalism, filaments begin at the right end and grow right to left by application of the transfer matrices. The starting vectors $\mathbf{V}_{\mathrm{UR}}$ and $\mathbf{V}_{\mathrm{IR}}$ initiate the assembly with an imperfectly aligned dimer. These dimers have $v - 1$ different states where v is the valence of the scaffolds. These states have free energy $e^{-(v-n)\epsilon}g_i(n)$ where $n = 1, 2, 3....v - 1$, and $g_i(n)$ is the gap partition function for the sticky end defect at right end of the filament. These alignment states comprise the elements of the right vectors

$$
\mathbf{V}_{\mathrm{UR}} = c_0 z_{\mathrm{U}} z_{\mathrm{I}}
\begin{bmatrix}
e^{-9\epsilon}g_{\mathrm{I}}(1) \\
e^{-8\epsilon}g_{\mathrm{I}}(2) \\
e^{-7\epsilon}g_{\mathrm{I}}(3) \\
e^{-6\epsilon}g_{\mathrm{I}}(4) \\
e^{-5\epsilon}g_{\mathrm{I}}(5) \\
e^{-4\epsilon}g_{\mathrm{I}}(6) \\
e^{-3\epsilon}g_{\mathrm{I}}(7) \\
e^{-2\epsilon}g_{\mathrm{I}}(8) \\
e^{-\epsilon}g_{\mathrm{I}}(9)
\end{bmatrix}
\quad \text{(A.9)}
$$

and

$$\mathbf{V}_{\mathrm{IR}} = c_0 z_{\mathrm{U}} z_{\mathrm{I}} \begin{bmatrix} e^{-9\epsilon} g_{\mathrm{U}}(1) \\ e^{-8\epsilon} g_{\mathrm{U}}(2) \\ e^{-7\epsilon} g_{\mathrm{U}}(3) \\ e^{-6\epsilon} g_{\mathrm{U}}(4) \\ e^{-5\epsilon} g_{\mathrm{U}}(5) \\ e^{-4\epsilon} g_{\mathrm{U}}(6) \\ e^{-3\epsilon} g_{\mathrm{U}}(7) \\ e^{-2\epsilon} g_{\mathrm{U}}(8) \\ e^{-\epsilon} g_{\mathrm{U}}(9) \end{bmatrix} \tag{A.10}$$

$\mathbf{V}_{\mathrm{UR}}$ and $\mathbf{V}_{\mathrm{IR}}$ are identical apart from the client species that can bind in the gap partition functions g_{U} and g_{I}, where the gap partition functions are labeled with subscripts that describe the client that they bind.

The transfer matrix $\mathbf{M_i}$ for deccavalent ($v = 10$) scaffolds is given by

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z_i e^{-9\epsilon} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & z_i e^{-8\epsilon} & z_i e^{-8\epsilon} g_i(1)) \\ 0 & 0 & 0 & 0 & 0 & 0 & z_i e^{-7\epsilon} & z_i e^{-7\epsilon} g_i(1) & z_i e^{-7\epsilon} g_i(2) \\ 0 & 0 & 0 & 0 & 0 & z_i e^{-6\epsilon} & z_i e^{-6\epsilon} g_i(1) & z_i e^{-6\epsilon} g_i(2) & z_i e^{-6\epsilon} g_i(3) \\ 0 & 0 & 0 & 0 & z_i e^{-5\epsilon} & z_i e^{-5\epsilon} g_i(1) & z_i e^{-5\epsilon} g_i(2) & z_i e^{-5\epsilon} g_i(3) & z_i e^{-5\epsilon} g_i(4) \\ 0 & 0 & 0 & z_i e^{-4\epsilon} & z_i e^{-4\epsilon} g_i(1) & z_i e^{-4\epsilon} g_i(2) & z_i e^{-4\epsilon} g_i(3) & z_i e^{-4\epsilon} g_i(4) & z_i e^{-4\epsilon} g_i(5) \\ 0 & 0 & z_i e^{-3\epsilon} & z_i e^{-3\epsilon} g_i(1) & z_i e^{-3\epsilon} g_i(2) & z_i e^{-3\epsilon} g_i(3) & z_i e^{-3\epsilon} g_i(4) & z_i e^{-3\epsilon} g_i(5) & z_i e^{-3\epsilon} g_i(6) \\ 0 & z_i e^{-2\epsilon} & z_i e^{-2\epsilon} g_i(1) & z_i e^{-2\epsilon} g_i(2) & z_i e^{-2\epsilon} g_i(3) & z_i e^{-2\epsilon} g_i(4) & z_i e^{-2\epsilon} g_i(5) & z_i e^{-2\epsilon} g_i(6) & z_i e^{-2\epsilon} g_i(7) \\ z_i e^{-\epsilon} & z_i e^{-\epsilon} g_i(1) & z_i e^{-\epsilon} g_i(2) & z_i e^{-\epsilon} g_i(3) & z_i e^{-\epsilon} g_i(4) & z_i e^{-\epsilon} g_i(5) & z_i e^{-\epsilon} g_i(6) & z_i e^{-\epsilon} g_i(7) & z_i e^{-\epsilon} g_i(8) \end{pmatrix}$$

Where z_i is the fugacity of the attached molecule and $g_i(n)$ is the gap partition function for a gap of n unbound modules. Again, the matrices $\mathbf{M}_{\mathrm{U}}$ and $\mathbf{M}_{\mathrm{I}}$ are identical apart from the

fugacity of the added molecule and the client species that binds in the gap. Since adding a SUMO molecule will leave a gap of unbound SIM modules, the gap partition function g_U describes the binding of clients composed of SUMO modules.

Applying N transfer matrices to the right vector generates a vector $\mathbf{V}(N+2)$ describing a filament containing $N+2$ scaffold molecules. Each element of the vector is a polynomial giving the partition function of a filament terminating with a different length sticky end. The left vectors, $\mathbf{V}_{iL}$, serve three purposes. First, they collapse the filament vector into a scalar polynomial that gives the complete partition function of the filament. Secondly, the left vector provides the statistical weights for client binding at the left sticky end. These two functions can be served by a vector of the form

$$\mathbf{V}_i = \{g_i(1), g_i(2), g_i(3), g_i(4), g_i(5), g_i(6), g_i(7), g_i(8), g_i(9)\}$$

The third function of the left vector is that it must account for the fact that a filament can terminate with either species of scaffold. That means that the final left vectors are given by

$$\mathbf{V}_{UL} = \mathbf{V}_U + \mathbf{V}_I \mathbf{M}_U \tag{A.11}$$

$$\mathbf{V}_{IL} = \mathbf{V}_I + \mathbf{V}_U \mathbf{M}_I \tag{A.12}$$

A.3 Monomer Concentration vs total concentration

The quantities c_U and c_I appearing in our calculations represent the concentration of scaffold molecules that have not formed intermolecular bonds. In contrast, the more experimentally accessible quantity is the total scaffold concentration C_{tot}, which includes molecules that have formed assemblies. These quantities are related as follows

$$C_{Utot} = c_U \frac{\partial Q}{\partial c_U} \tag{A.13}$$

$$C_{\text{Itot}} = c_{\text{I}} \frac{\partial Q}{\partial c_{\text{I}}} \tag{A.14}$$

Fig 2.16 plots the free scaffold concentrations as a function of the total SUMO module concentration. There are several observations to make from this plot. First, when the scaffold concentrations are equal the monomer concentrations are very low, on the order of 10^{-10} M. This is much lower than the dilute phase concentration reported by[7]. The discrepancy is due to the fact that most molecules in the dilute phase are in the perfectly aligned dimer state. This state satisfies all the available bonding sites, rendering the scaffolds inert to further assembly processes.

Second, when one scaffold is in excess, SUMO in this case, the monomer concentrations diverge widely (note the logarithmic vertical axis). This is because the low concentration scaffolds are mostly consumed in complexes with the higher concentration module. The depletion of low concentration modules results in an excess of high concentration modules that increases as the stoichiometry asymmetry increases.

Third, it is useful to compare the concentrations in Fig 2.16 to the 10^5 M^{-1} affinity for monovalent SUMO-SIM binding[7]. At equal scaffold stoichiometry, the scaffold monomer concentrations are much too low for monovalent binding to occur. This justifies our approximation of strictly 1D filament formation because most of the gaps only have one or two unbound sites. However, when the SUMO module concentration reaches 90 μM (a 9:5 excess over SIM), the concentration of free SUMO module scaffolds is $\simeq 14$ μM. This is comparable to the $\simeq 0.1$ μM concentration of clients, so monovalent binding cannot be neglected.

A.4   Excess monomer scaffold binding at defect sites

When the scaffolds are present at unequal stoichiometries, the excess scaffold accumulates at concentrations where monovalent binding in the gaps becomes significant. At large asymmetry, the excess scaffold monomer concentration ($\simeq 1$ μM) is about ten times greater than the total client concentration ($\simeq 0.1$ μM) so that monomer scaffold binding is more favorable than client binding at defect sites. Therefore, we need to correct our 1D filament model to
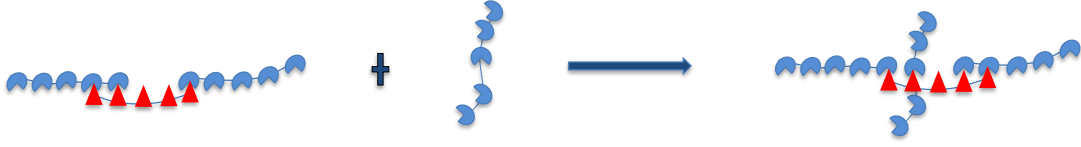
Figure 1: *Excess monomer scaffold concentration binding at defect sites in crosslinked fashion.*

allow for perpendicular binding of scaffolds. As a first correction, we consider the "t" configuration illustrated in Fig. 1. This correction only allows for monovalent binding, which we expect to dominate given the small size of gaps in the filament.

The concentration of scaffolds that bind in the "t" configuration is calculated as

$$c_{\text{extra}}^{\text{monomer}} = 10 K_{\text{a}}^{1} * \rho_{\text{g}}^{\text{I}} c_{\text{I}}^{\text{droplet}} * c_{\text{U}}$$

Where $K_{\text{a}}^{1} = 10^{5}$ M^{-1} is the monovalent SUMO-SIM binding affinity[7]. This affinity is multiplied by a factor of 10 to account for the degeneracy of binding of a decavalent scaffold to a single-site defect. c_{U} is the scaffold monomer concentration (Fig. 2.16) and $\rho_{\text{g}}^{\text{I}} * c_{\text{I}}^{\text{droplet}}$ is the concentration of SIM defects in the network droplet. The latter quantity is calculated from the concentration of SIM scaffolds in the droplets, $c_{\text{I}}^{\text{droplet}\,7}$, and the density of defect sites per SIM scaffold, which is calculated from $\rho_{\text{g}}^{\text{I}} = N_{\text{tot}}^{-1} g_{\text{U}} \partial \ln Q_{\text{droplet}} / \partial g_{\text{U}}$, where we have used the substitution $g_{\text{U}}(m) = g_{\text{U}}^{m}$ to define g_{U} as the statistical weight of an unbound SIM module. Again, the subscript follows from our definition of gap partition functions based on the client that they bind. Note that this correction has not been applied to our PC calculations, which explains the systematic underestimate of the PC at high SUMO concentration.