Phase separation and superselective binding of multivalent associative polymers

ANDREW TREVOR RANDAL CHRISTY

Phase separation and superselective binding of multivalent associative polymers

Andrew Christy

A thesis presented for the degree of
Doctor of Philosophy

Department of Chemistry
Durham University
United Kingdom

November 2021

Abstract

Multivalent polymers play pivotal roles in many biological and synthetic systems. Multivalency can often be tuned to provide control over both the enhancement and suppression of binding. In this thesis, Monte Carlo simulations of simplified model chains on a cubic lattice are used to study multivalent polymers in liquid-liquid phase separation, and in superselective binding to a host.

We are particularly interested in liquid-liquid phase separation due to its potential role in the formation of membraneless organelles, which support cellular fitness and exhibit interesting properties and functions. Multivalent polymers are broadly representative of many membraneless organelle components, such as RNA and proteins. We investigate the different ways in which linkers can be represented on a cubic lattice and study how this impacts the individual chain properties and the aggregates they form. We find that the phase separation of multivalent polymers is promoted when the polymers can form dense, highly interconnected structures; this is achieved through increasing the polymer length and valency.

We then study how the binding of individual multivalent polymers can be tuned to achieve highly selective binding. We show that superselective binding of multivalent polymers to a 3D host can be controlled by the presence of crowder species in the 3D host. Superselectivity describes binding that is very sensitive to the density of receptors, resulting in a sharp binding transition at a receptor density threshold. Superselectivity is shown to be a potential mechanism by which membraneless organelles control their composition.

Superselective binding of multivalent polymers at surfaces is also investigated. We find that superselectivity is enhanced on moving from a flat surface to a pitted surface. Furthermore, rough surfaces show potential for polymer sorting or selective recruitment, whereby the binding of polymers to a given surface is dependent on the interaction strengths, polymer properties and pore geometry. The underlying ther-

modynamics is analysed, allowing us to identify the enthalpic and entropic drivers, and also indicate how this phenomenon can be harnessed in applications.

Declaration

The work presented in this thesis has not been submitted for another degree at Durham University or any other institution. The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

I would like to offer my immeasurable thanks to Mark Miller and Halim Kusumaatmaja; their insight and guidance has been invaluable during my studies. This work is a reflection of the excellent support and wisdom that they provided during this work.

My heartfelt thanks are also due to my family, for their unending support, love and encouragement during this process. Many thanks also to Dr Tiemessen and Dr Mackay for their steadfast friendship and their crucial, if impatient, role in improving my coding aptitude. I would also like to thank all the members, both past and present, of 200x; you have made this PhD a happy and memorable experience, ensuring that there was never a dull moment. It has been a pleasure to train with Durham University Triathlon Club, and the friendships made there will last a lifetime. Special thanks go to Leon and Peter for their great company and stimulating discussions! Finally, none of this work would have been possible without the love and kindness of Eliza, who has seen me through the darkest depths and provided the most joyful times during my period in Durham.

Contents

Chapter 1

Introduction

Multivalent polymers are crucial components in a plethora of natural and synthetic systems. Their multivalency allows them to drive phase separation through cooperative effects and by connecting many components of lower valency [1]. In this chapter we shall first discuss liquid-liquid phase separation, and the role of multivalent species in such processes. One area where the phase separation of multivalent polymers is of interest is in the formation of membraneless organelles. Membraneless organelles are an ever-growing area of research and a summary of these biological bodies provides context to our investigation of multivalent polymers. Following this, we focus more on individual polymer behaviour, and how multivalency can provide a pathway for selective binding. We provide an overview of superselectivity, a binding phenomenon which depends on multivalency, and indicate the means by which, to date, it has been controlled and tuned.

1.1 Condensation of multivalent polymers

1.1.1 Membraneless organelles

Over the past two decades there has been an ever growing surge in research into Liquid Liquid Phase Separation (LLPS) and its role in the formation of membraneless organelles. The cellular bodies differ from the more widely known membrane-bound organelles in their absence of a lipid bilayer to delimit them from the cell cytoplasm. An example of membraneless organelles is shown in the fluorescence microscopy images in Figure 1.1. The bright green structures are Fused in Sarcoma (FUS) granules, which form in response to a variety of environmental stimuli and

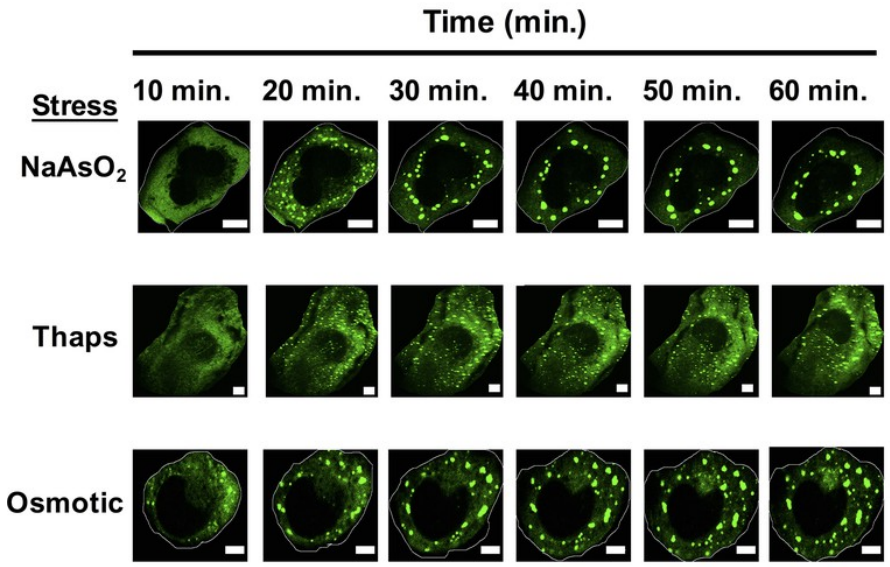


Figure 1.1: Fluorescence microscopy images of Fused in Sarcoma (FUS) granules formed *in vivo*. The granules form from the cytoplasm and mature over time to become distinct compartments of concentrated biological material. Image cropped and reprinted with permission from [2], DOI: https://doi.org/10.7554/eLife.18413.001. Copyright ©2016 by eLife. Creative Commons Attribution 4.0 International Public License.

exist as polydisperse structures within the cytoplasm.

Despite their discovery being almost four decades ago [3], it is only recently that these membraneless organelles have drawn significant attention. They play crucial roles in cell functions and are connected to a number of neuro-degenerative diseases [4]. Furthermore, the ubiquity and importance of LLPS in cellular processes has driven research in this direction.

Membraneless organelles, also known as biomolecular condensates, form within the cytoplasm or nucleus of cells. Formation occurs through the association of different polymeric species, as illustrated in Figure 1.2A, until there is sufficient interconnectivity for the organelle to be distinct from the cytoplasm. This involves a competition between minimising the entropy loss on proteins aggregating, but maximising the interaction energy of the components. Multivalent polymers, in the form of proteins and RNA, enable these highly interconnected structures to form [1, 5–8]. Membraneless organelles encompass a huge range of cellular structures, a full review of which would likely require a textbook. Here we focus on the family of organelles known as RNP granules, which consist largely of ribonucleic acid (RNA) and RNA-binding proteins [4]. Within this family of organelles there are numerous different subspecies of condensates, several examples of which are given in Figure

1.2B, alongside their location within the cell. To give a sense of the variety of RNP granules and their roles, P-bodies are thought to have a regulatory role in germ cells [9], FUS granules are strongly associated with Amyotrophic Lateral Sclerosis (ALS) [10], and stress granules form in response to cellular stress to aid cell survival [4]. The remainder of this section gives a brief overview of these fascinating and diverse cellular bodies.

Membraneless organelles are necessary for maintaining cellular fitness. A key means by which this is achieved is their ability to recruit and expel species from or to the cytoplasm [12]. As referenced previously, stress granules form in response to external stresses such as pH, temperature, glucose deprivation, osmotic shock or poisoning. These organelles, once formed, are then able to sequester components not required for cell survival, and to inhibit messenger RNA (mRNA) translation [9], allowing the cell to focus its resources on survival [8, 13]. Membraneless organelles can also purportedly act as reaction vessels, again facilitated by their sensitive control of composition [14]. Additionally, their formation and recruitment has been closely linked to cellular signalling [8]. Moreover, they have been found to aggregate anti-viral proteins in the event of a cell being infected by a virus, which stimulates their activation and replication [4]. However, they are also susceptible to disruption by invading viruses, which may chemically alter components crucial for organelle formation to prevent the cell from generating anti-viral production bodies [15].

Membraneless organelles are closely linked with disease due to their ability to mature and develop into amyloid fibrils, whose formation is the hallmark of neurodegenerative diseases. The exact reasons for this are unknown. However, mutations which are believed to cause certain neuro-degenerative conditions increase the rate of formation and longevity of stress granule-like structures when studied *in vitro* [4]. In addition, mutations have been found to impede cells' ability to clear the membraneless organelles [16]. Furthermore, LLPS is believed to increase the risk of amyloid formation by increasing the local concentration of the aggregating species [16]. Therefore, gaining a deeper understanding of this process should aid the mission to disrupt amyloid formation and the subsequent neurological effects.

Although the exact mechanisms behind membraneless organelle formation and function are currently disputed, there is general consensus that they are held together through a combination of weak isotropic and stronger specific interactions. These biomolecular condensates are believed to consist of proteins with Intrinsi-

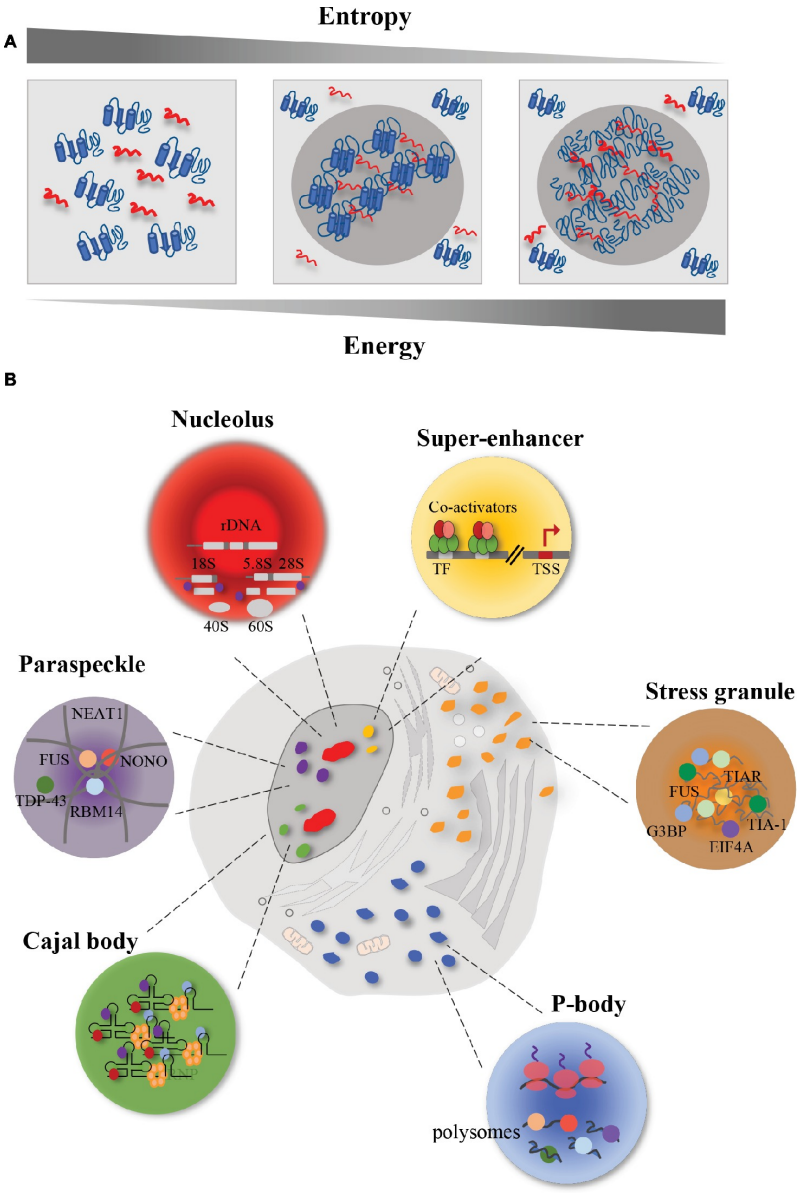


Figure 1.2: (A) Illustration of the formation of a membraneless organelle from a bulk fluid, e.g. cytoplasm. (B) Examples of different membraneless organelles, including whether they are found within the cytoplasm or nucleus. Within each of these sub-categories of organelle there are numerous different species of structure. Despite the compositional differences between the different organelles, there is commonality between the different structures. A fundamental point of agreement is that they all exploit multivalency in order to form and function. Image reprinted with permission from [11], DOI: https://doi.org/10.3389/fgene.2019.00173. Copyright ©2019 by Verdile, De Paola and Paronetto. Creative Commons Attribution 4.0 International Public License.

cally Disordered Regions (IDRs), alongside proteins with modular binding sites, as depicted in model form in Figure 1.3. Interestingly, despite the weakness of the isotropic interactions, long-lived aggregates purely consisting of Intrinsically Disordered Proteins (IDPs) can form due to the cumulative strength of having many weak interactions [17]. This is in contrast to the specific binding domains which interact selectively with their corresponding binding sites with much stronger anisotropic interactions [18]. Therefore these membraneless organelles are held together through an assortment of interactions, which on closer inspection include π-π stacking, hydrophobic interactions, hydrogen bonding and electrostatic interactions [12, 18]. Indeed, charge distribution and protein sequence can also control LLPS — as complex coacervation plays a role in some systems [19]. Lastly, external species, such as molecular crowders in the bulk, can promote or even induce phase separation [20], adding further complexity to this process.

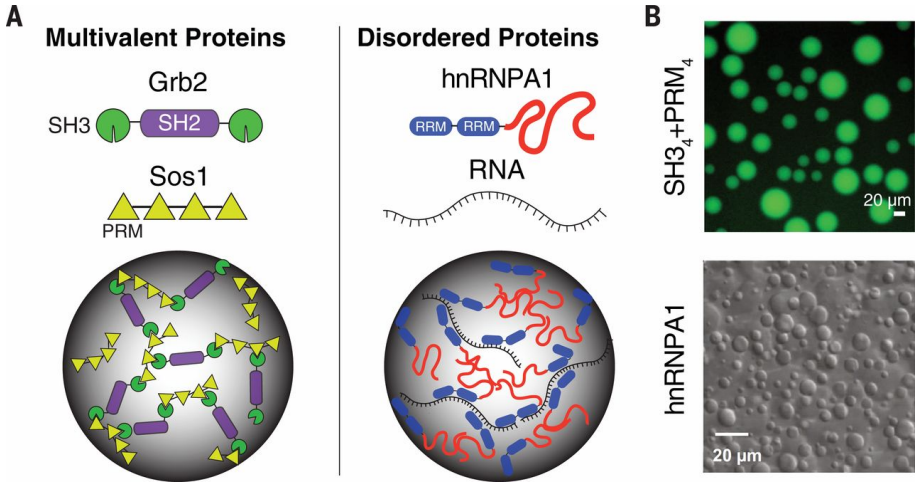


Figure 1.3: (A) Cartoon of membraneless organelle composition, where an example of an organelle with well defined, structurally stable binding sites is shown adjacent to an organelle where the interactions are predominantly between disordered proteins which lack well defined binding sites. (B) Microscopy images of biomolecular condensates formed from model proteins. The top image is for structures consisting of well defined binding sites, and the bottom image is for a condensate consisting of disordered proteins and RNA. Image from [8], DOI: https://doi.org/10.1126/science.aaf4382. Reprinted with permission from AAAS.

1.1.2 Role of multivalent interactions

Multivalent binding is ubiquitous within nature, playing a crucial role in controlling the adhesion of viruses to cells, intercellular interactions, antibody response, polymer phase separation and DNA transcription [21, 22]. Counterintuitively, mul-

tivalency can also be utilised to limit and control undesired binding and aggregation, through multivalent inhibitor species [21]. Multivalency facilitates behaviour that is unachievable for monovalent species, and therefore has attracted significant research over the past three decades. Indeed, in systems where the individual interactions are weak, multivalent binding can even result in strong overall interactions due to cooperative effects [23], which will be discussed in Section 1.2.2. Multivalent species are capable of driving LLPS, as shown by Bracha *et al.* where the concentration of IDRs required for phase separation was drastically reduced by the addition of multivalent species capable of recruiting IDRs [24]. This reveals one means by which multivalency can promote LLPS, by increasing the local concentration of the IDRs.

Although multivalent species can drive phase separation, even in their absence, IDPs and IDRs are inextricably linked to membraneless organelles. They can control the formation, behaviour and composition of the membraneless organelles to which they associate. It is believed that these regions interact promiscuously with each other resulting in a transient bonding network, and it is this network that gives the granules their properties [18]. Research into granule-related proteins has also revealed that they are usually rich in polar and aromatic residues, but charged residues are rare. All this gives some insight into the likely bonding present in granules [25].

Another crucial multivalent component in membraneless organelles, especially in RNP granules, is RNA. In fact, not only has RNA been implicated in the formation of cellular granules, it has also been found to significantly affect the behaviour of membraneless organelles in the cytoplasm. An example of where RNA appears to impact the condensate dynamics was in a study by Elbaum–Garfinkle and co-workers. They found that by adding short RNA strands to the protein LAF-1, droplets of lower viscosity formed [26], but did not have any effect on the minimum concentration of LAF-1 required for the phase transition. In further work done by Wei *et al.* they found that adding longer RNA strands increased the viscosity of the droplets, implying that RNA entanglement can play a role in the droplet viscosity. As the RNA found in RNP granules can have a range of lengths, this may be a mechanism by which the cell controls the physical properties of its granules [27]. The stiffness of RNA is also believed to have profound impact on the formation and properties of biomolecular condensates [28]. The formation and properties of organelles depend heavily on the distribution of charges on RNA. For example,

the highly charged polyuridylic RNA promotes LLPS through its ability to form electrostatic and π interactions with other RNAs in the system [20]. Therefore RNA is pivotal in determining the organelle properties.

The roles of proteins in RNP granules vary considerably. Whereas many protein species are redundant in the formation of membraneless organelles, certain proteins have been found to be crucial. For instance, three such proteins are GTPase-activating protein-binding protein (G3BP), Eukaryotic Initiation Factor 2 (eIF2-α) and T-cell intracellular antigen (TIA) [29].

Beyond composition, processes such as protein mutations and Post Translational Modificationss (PTMs) are also critical in membraneless organelle formation and physical properties. An example of where mutations impact organelle function is found in mutated proteins destroying the functionality of FUS granules; instead of being dynamic structures, aggregation is promoted [30], and the structures lack liquid-like properties. When proteins containing these mutations were investigated it was found that they form fibres from the phase-separated droplets more readily [31]. Therefore, the protein history (any chemical reactions or sequence alterations they have undergone) is also important in the formation of membraneless granules [29]. PTMs are vital in the behaviour and formation of biomolecular condensates, as they alter the protein-protein and protein-RNA interactions (either through altering configurations of protein regions or by chemical changes to binding sites). As such, they can trigger the formation or dissolution of organelles. There are numerous PTMs observed within cells, and these modifications are often achieved through enzymatic activity [32]. Significant alterations to organelle environment and components therein are possible due to the sheer range available, a sample of which are: phosphorylation, glycosylation, acetylation, ubiquitination, SUMOylation and methylation [33–37].

1.1.3 Client–scaffold model

There is growing evidence to support a theory that membraneless organelles consist of a *scaffold*, which then recruits and expels *clients*, known as the *scaffold-client* theory [32]. The scaffold molecules tend to be RNAs and RNA binding proteins. These remain in the condensate far longer than the clients, which are often shorter proteins and enzymes and are transiently located in the condensate [32]. The two

different dynamics regimes allow the distinction between clients and scaffolds to be made [38], as evidenced by Fluorescence Recovery After Photobleaching (FRAP) experiments [39]. Indeed, RNA base pairing is thought to facilitate this decoupling [38] through the strong interconnectivity of RNA and weaker binding to the client proteins. It has also been observed that many components often present in RNP granules are redundant in their formation; the granules still form even when the component is removed from the system [40]. The client–scaffold model explains this by labelling the redundant species as clients, with a vastly diminished role in the initial phase separation. These clients do not form the scaffold, and instead are just recruited into the granule after its formation. Finally, it has been shown that client recruitment is very sensitive to the stoichiometry of the scaffold species within test systems – giving credence to the idea that the scaffold structure controls client recruitment and could even be altered by PTMs which would cause the expulsion of certain clients whilst recruiting others [32]. This would allow for a scaffold to recruit enzymes and substrates as required to promote reactions or sequester components to prevent them reacting and using up cell resources during periods of stress. Therefore this mechanism further supports the theory of a client-scaffold structure.

A cartoon model of a multivalent scaffold recruiting clients is shown in Figure 1.4, where the clients are recruited into the scaffold depending on their valency. The valency of clients will be an important parameter later in this thesis, but at this point it is worth noting that the recruitment of clients to a scaffold of excess receptor sites varies non-linearly with client valency [32]. Further discussions of the merit of this model can be found in Chapter 4.

One limitation of the client-scaffold model is that, despite being likely to represent many membraneless organelles, there are some structures where the concentrations and molecular weights of the species blur the distinction between clients and scaffolds. In this case, the decoupling of dynamics between clients and scaffolds is much less pronounced. In such systems, the phase separation process, the structures formed, and the functions are impacted [42]. Whilst this is interesting in its own right, in this thesis we will focus on the case where the clients are much shorter than the scaffold species and the distinction between their dynamics is clear.

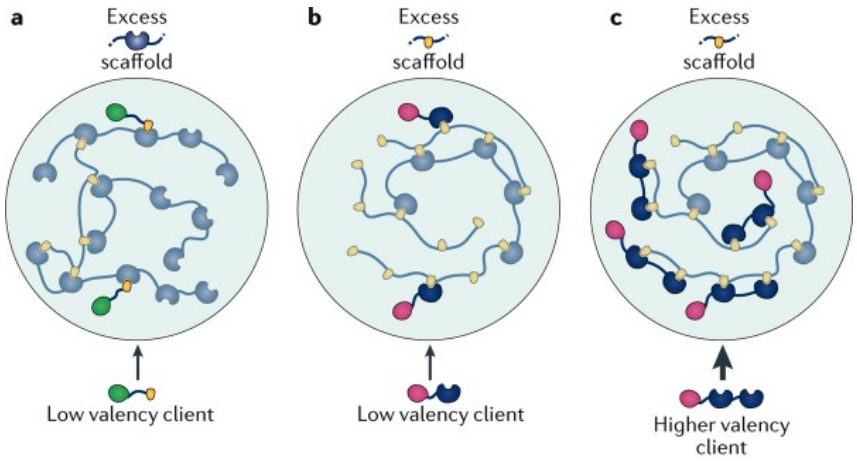


Figure 1.4: Illustration of the mechanisms of the client scaffold model; (A) and (B) are of low valency clients being weakly recruited to the scaffold, whereas (C) is for a high valency clients being strongly recruited to the scaffold. Figure reprinted with permission from [41], DOI: https://doi.org/10.1038/nrm.2017.7. Copyright ©2017 by Springer Nature.

1.1.4 Ostwald ripening and arrested phase separation

A passive two phase liquid system, consisting of droplets of one phase dispersed within the other, would be expected to coarsen over time. This is known as Ostwald ripening, and involves the largest droplet growing at the expense of the smaller droplets, until only one droplet remains. For this description we will define the two phases as the droplet phase and the continuous phase. Ostwald ripening drives the growth of large droplets and the shrinkage of smaller droplet in order to minimise the interfacial free energy. To illustrate why the smaller droplets are less favourable than a single large droplet, we must consider the Laplace pressure. The Laplace pressure is the difference in pressure p on moving across the droplet interface, from inside the droplet into the continuous phase. The Laplace pressure is given by

$$\Delta p = 2\gamma/R$$

where R is the droplet radius and γ is the surface tension of the droplet submerged in the continuous phase. Therefore, a passive two phase system, consisting of droplets of one phase dispersed in the other, will coarsen over time, to maximise the R of the droplet and minimise the Laplace pressure. This coarsening involves material moving from smaller droplets to larger droplets [43]. Systems which do not coarsen in this manner are resisting this process.

Cells contain polydisperse biomolecular condensates, indicating that Ostwald ripening is somehow being suppressed. At least three mechanisms could explain this observation. Firstly, Dufresne *et al.* proposed that an elastic matrix can be used to allow a bi-component system to phase separate into droplets of one species within a bulk of the other, without the droplets eventually forming a single large drop [44]. An illustration of this model is shown in Figure 1.5, alongside a phase diagram for the mixing and demixing of bi-component mixtures. The elastic network was capable of suppressing Ostwald ripening, and even more strikingly, they found Ostwald ripening to be reversed in some cases. Here, material moved from large droplets to smaller droplets to reduce the elastic penalty of the large droplets distorting the elastic cross-linked polymer gel network that they are in. This elastic penalty can override the surface tension driven material transfer from the small to the large droplets. Moreover, they subsequently generated synthetic polymer networks capable of suppressing nucleation, or controlling the locality of the droplets by having a stiffness gradient (droplets moved to the lower elastic strength regions). The authors suggest this could be a mechanism within cytoplasms, which also have complex properties; however, due to the complex nature of cells *in vivo* studies are currently limited [45].

A second factor that limits droplet growth was proposed by Shin *et al.* [46]. Despite cellular complexity, these authors were able to study the mechanical impact of the nucleolar matrix on LLPS within cells. They found that IDPs self-assembling within a nucleolus can exclude chromatin — soft genomic material is readily recruited and allows coalescence of droplets, whereas genomic material with high chromatin density is not recruited due to the large elastic penalty associated with deforming this material during droplet growth [46]. Elasticity of the droplet components is therefore yet another consideration in the study of these droplets.

A third way to limit droplet growth is through chemical reactions. Wurtz and Lee developed an idealised model of arrested phase separation where the aggregating component could be reversibly converted into a soluble form by an active process such as phosphorylation, thereby counteracting the growth of large aggregates [47]. For low conversion rates, Ostwald ripening still leads to a single, large aggregate, just as for a passive system where there is no conversion. At the other extreme, high conversion rates lead to a homogeneous solution. However, for intermediate rates, a steady-state distribution of finite droplet sizes is sustained as a result of the

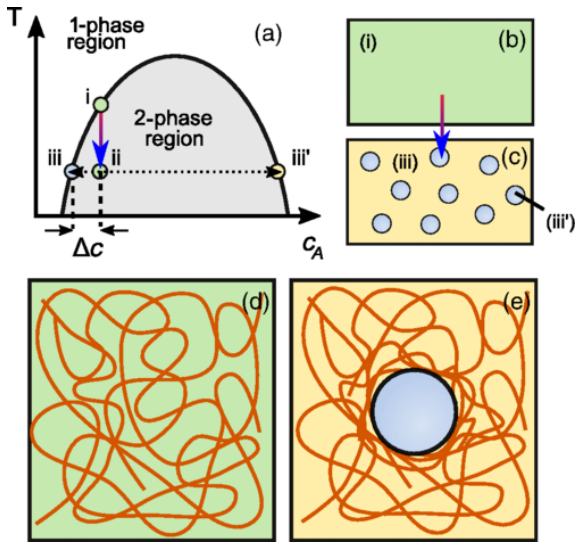


Figure 1.5: (a) Bi-component phase diagram, (b) and (c) illustration of phase separation of blue and yellow components. (d) Sketch of phase separation suppressed by an elestic network (red lines) and (e) of arrested phase separation of the blue and yellow components within an elastic network. Figure reproduced with permission from [44], DOI: https://doi.org/10.1103/PhysRevX.8.011028. Copyright 2018 by the American Physical Society. Creative Commons Attribution 4.0 International Public License.

dynamic balance between Ostwald ripening and molecular solubilisation.

1.1.5 Membraneless organelle properties and structures

Membraneless organelles are often described as exhibiting liquid-like properties, such as a largely spherical shape, the ability to fuse with each other to form larger spheres and their wetting of substrates [48]. FRAP is commonly used in this field to study how quickly particles diffuse both within the biomolecular condensates and between the cytoplasm and the condensates. Many biomolecular condensates display rapid recovery of fluorescence after photobleaching, indicative of dynamic structures where diffusion within the structures is swift. Separate experiments have also shown rapid diffusion between the condensate and the cytoplasm [49]. Crucially, this diffusion does not result in loss of the organelle structure, as the condensates can maintain their shape for minutes to hours whilst exchanging material with the cytoplasm [50]. Therefore, membraneless organelles can have stable structures with dynamic properties. The observed diffusion within these organelles further implies that covalently cross-linked gels are unlikely to be the structure of these cellular bodies [4]. Instead, physical crosslinking, where bonds can form and be broken more easily, is thought to dominate. The presence of ATP in cells, providing energy

to break bonds, facilitates such processes [17]. The nature and timescale of the bonds determines the physical properties of organelles, with stronger and longer lasting bonds increasing the viscosity of the organelles. Additionally, rapid diffusion is likely to be aided by the apparent porosity of many membraneless organelles [51].

Intriguingly, many membraneless organelles also show more complex behaviours. As illustrated in Figures 1.6B and C, it has been shown that increasing interaction strength generally leads to fibre formation and thus a non-spherical shape, whereas slightly weaker interactions result in glass-like structures [8]. Moreover, although many membraneless organelles form initially as liquid structures, some mature into more solid-like structures [2, 30]. They can have different properties depending on the type of cell they are in [18]. For instance, stress granules are liquid-like within mammalian cells, but solid-like in yeast cells. The exact reason for the different states is largely unknown, especially as in the case of the yeast stress granules, these solid structures are able to redissolve into the cell faster than more liquid-like misfolded protein aggregates. Whether a granule has liquid-like or solid-like properties is often dependent on the stress that induced its formation [18]. Additionally, time plays an important role in their state, as maturation of the stress granules often results in the transition from being liquid-like to solid-like [31]. This process can occur through mutation of the proteins [30], PTMs of the components [52] or through significant changes in the cellular environment and composition [16]. Solidification is believed to result in irreversible damage to cells, as the granules can no longer perform their cellular roles once in a solid state. Hence, this maturation is of great interest.

Looking more closely at the structures of membraneless organelles, it has been argued that some consist of an inner core, which is relatively stable and an outer shell which exhibits dynamic properties. Stress granules (membraneless organelles which form in response to unfavourable cellular conditions) are widely documented to exhibit this structure. The reported liquid-like dynamics of the granules — such as their observed coalescence, flow and break-up — are suggested to be due to this outer shell [2]. The core–shell structure was successfully reproduced using dynamic Monte Carlo simulations over an intermediate range of interaction strengths in [38].

There are two current proposed routes of stress granule formation which satisfy the experimental observations of having an inner core and outer shell, and these are illustrated in Figure 1.7. The first model, proposed by Jain *et al.* asserts that

mRNAs initially form a stable core with a relatively strong network of bonds holding them together, Figure 1.7 (A). Following this, since the concentration of IDRs is increased, LLPS is induced and thus the dynamic shell forms around the core. The second mechanism proposed by Lin *et al.*, states that liquid–liquid phase separation (LLPS) occurs first, due to the weak interactions between the components, and then the more stable core forms subsequently due to the resulting increased concentration of stress granule components in the phase separated region [53]. Therefore, although there is a growing understanding of the types of bonding that is present in stress granules, the order of, and the initiating factors in, their formation are still under investigation.

We also note that there is increasing experimental evidence that multiple immiscible membraneless organelles can coexist. As illustrated in Figure 1.6A, different arrangements can occur when two immiscible droplets form and interact, including droplet encapsulation to minimise the interfacial energy of the system [8]. Bracha *et al.* illustrated the complex interactions between immiscible membraneless organelles by designing peptides that had the ability to control condensate structure [54].

Finally, the coexistance of a two component system is determined by the interactions between the two components and their respective concentrations. The system can either be a single mixed phase or can demix into two separate phases. The demixing process can either occur through nucleation, or spontaneously by spinodal decomposition [55]. The pathway by which the phase separation occurs, through nucleation or spinodal decomposition is crucial in determining the mechanism by which membraneless organelles form [55]. Therefore generating full phase diagrams of these systems, indicating the concentrations and interaction strengths at which either nucleation or spinodal decomposition occur, will be essential in ascertaining how these membraneless organelles form within real cells. Recent developments in experimental techniques — such as manufacturing systems which exhibit spatially controlled, photoinduced phase separation [24] — have increased the likelihood of generating detailed phase diagrams for these biological system within this decade.

1.1.6 Experimental model systems for LLPS study

Due to the complexity of membraneless organelle formation and the plethora of proteins that are implicated, one of the most promising streams of research involves

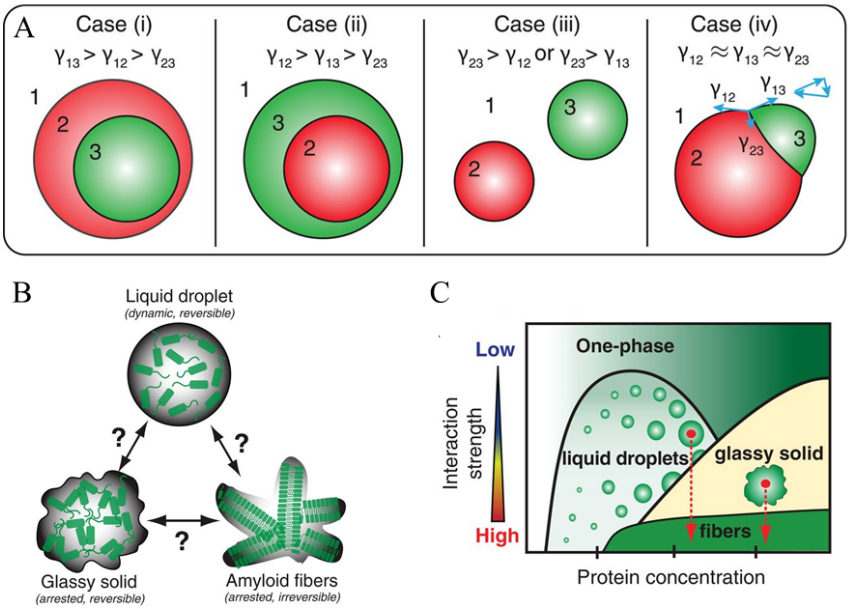


Figure 1.6: (A) Schematic of a bi-droplet system for various interfacial tensions γ between the phases involved. By altering the relative interfacial tensions of the different liquids, the arrangement of the droplets can be controlled. (B) Illustrations of the various droplet shapes that IDRs in solution can form, including amyloid formation. (C) Phase diagram of an IDR in solution for varying IDR interaction strengths and concentrations. Images from [8], DOI: https://doi.org/10.1126/science.aaf4382. Reprinted with permission from AAAS.

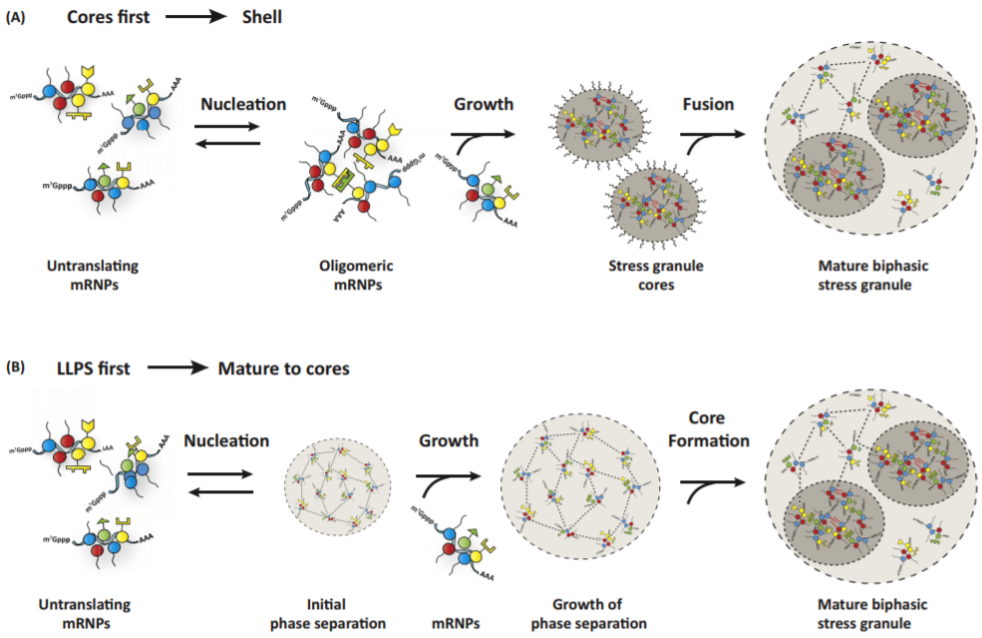


Figure 1.7: Schematic showing two proposed routes for the formation of membraneless organelles with a core-shell structure. In both systems the different coloured circles represent binding sites, which can interact. The light shaded area shows low density whereas the dark shaded regions are the dense cores of the structure. Figure cropped and reprinted with permission from [4], DOI: https://doi.org/10.1016/j.tcb.2016.05.004. Copyright ©2016 by Elsevier.

a bottom-up approach. In order to do this, simple model systems containing some of the key proteins (or even just the most interesting domains of key proteins) found in granules are used to probe the fundamental causes of liquid-liquid phase separation. Previous studies have looked at how molecular weight impacts the phase separation of globular proteins [56]. However, more insight can be obtained by the decoupling of valency and molecular weight achieved by studying model systems of modular protein domains with set valencies. A multitude of such systems of modular domains, including their impact on signalling, can be found in [57].

One such study looked at the impact of valency on the mechanism by which toxins aggregate [58]. Sisu *et al.* synthesised inhibitor proteins of set valency and probed how the valency impacts the aggregation, and thus inhibition, of *E. coli* heat-labile toxin and cholera toxin. In a development from previous studies, where increasing inhibitor valency had been shown to increase inhibition [58–61], they found that the interplay between the toxin and the inhibitor valency impacts both aggregation kinetics and stability.

In a later landmark study by Li *et al.* [7], *in vivo* studies on the components of a naturally occuring, three-component system of Waskott-Aldrich syndrome protein

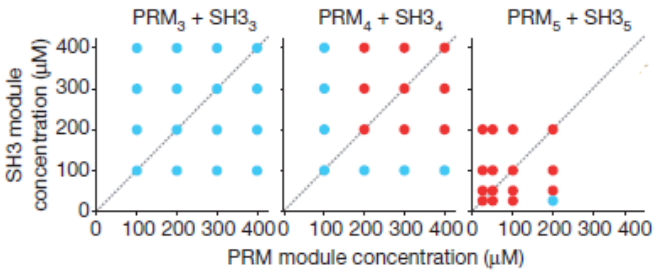


Figure 1.8: Phase diagram for system of PRM and SH3 proteins at varying concentrations, where red dots represent phase separation occuring and blue dots show concentrations where no phase separation was observed. The valency of the SH3 and PRM species in each study are shown above the respective plots. Image reprinted with permission from [7], DOI: https://doi.org/10.1038/nrm.2017.7. Copyright ©2012 by Springer Nature.

(N-WASP), the protein NCK and phosphorylated nephrin were carried out, showing a sharp phase transition which is dependent on the protein valencies. The system they looked at was a simplified model system consisting of segments of several proteins found in signalling pathways of cells. However, it was further simplified to an *in vitro* study of two domains found in these signalling proteins. The domains were SRC homology 3 (SH3) and its proline-rich motif (PRM). These engineered biopolymers consisted of repeats of either of the two domains, where the number of domains on each biopolymer was controlled. Various concentrations of these two engineered proteins were mixed in solution, and their phase behaviour was probed using microscopy techniques. The results of this study, showing how the phase separation of these two components varied with valency, can be found in Figure 1.8. The nomenclature used for these experiments was $\mathrm{DOMAIN}_{\mathrm{valency}}$, therefore $\mathrm{SH3}_3$ refers to a protein of three SH3 domains, and PRM_5 is five PRM binding domains. Varying the protein valency showed that by increasing the protein valency, the minimum concentration of protein domains required for phase separation to be observed is reduced. The extended discussion of this particular model is presented here as this is the model on which we have based our simulation study.

Similar model systems have also been studied using repeating binding sites of numerous proteins [7, 32, 62–64]. In recent studies, the role of dimeric species on phase separation and stress granule formation has been probed, with a focus on G3BP [29, 65, 66] and RNA. By reducing the number of cytoplasmic components, the interplay between G3BP and RNA could be more clearly observed.

Lastly, Low Complexity Domains (LCDs) which often lack a well defined sec-

ondary structure, can also influence the phase separation and properties of RNA binding proteins. The relationship between well defined binding regions and IDRs have also been studied in model systems. One such instance is in the study of the RNA binding Protein hnRNPA1 [67], which contains both LCD and modular regions. Altering this protein to remove the modular binding regions or the LCD, showed that the combination of these two regions is necessary for the condensate to exhibit the desired dynamic properties — without the modular regions less dynamic hydrogels formed. Moreover, despite the propensity of LCDs to phase separate in solution, the modular regions facilitate this at much lower concentrations by crosslinking with the RNA.

1.1.7 Simulation of LLPS

Computational studies of LLPS and membraneless organelles are also interspersed in the experimental literature. For example, the model system of SH3 and PRM domains, realised by Li *et al.* [7], was also studied computationally by the same authors. A stochastic algorithm was used to simulate SH3 and PRM species at various concentrations, and with a prescribed number of binding domains. These protein domains were randomly positioned in a 20 nm sided cube. Each protein binding site could bond to one binding site of the other protein type. The cube was divided into smaller cubes, with proteins only able to bond to other proteins within the same cube, the probability of which was found from their experimental work. Similarly, two protein domains that were already bound could also dissociate, with a specified probability. After each binding or unbinding event, the affected proteins had a chance of moving into one of the six neighbouring cubes — which depended of the number of each type of protein present in the system. The phase separation of the system was defined as a cluster having formed containing at least 50% of one of the domains. This phase separation was found to occur at lower protein concentrations as the valency of the proteins increased from 3 to 6 [7]. This allowed, in contrast to the experimental work, for straightforward, rapid and inexpensive generation of phase diagrams for species of varying valencies. However, this simple model did not account for any steric effects, either due to the volume of the proteins or the orientation of the domains, as well as omitting calculations of conformational entropy. Further simulations have also been carried out, by various athors, using

these model system components [32, 68, 69].

A particular study of note by Pappu and co-workers [68], involved cubic lattice simulations of the two proteins domains studied by Li *et al.* They investigated the role of biopolymer valency, linker type, linker length and interaction strength on phase separation. They modelled their biopolymers as binding sites connected by linkers — with the options of hard linkers, implicit linkers, or a hybrid of the two. They found that, where clustering was exhibited, the explicit linkers resulted in system-spanning networks, whereas the implicit linkers resulted in a compact structure forming. Therefore, they concluded that, for the explicit linkers, gelation occurred without phase separation. Conversely, the implicit linkers resulted in phase separation. They also carried out further work where simulations of a mixture of polymers with hard and implicit linkers respectively were conducted; there was a spatial sorting of the species, which indicated a mechanism by which the core-shell structure could be achieved, i.e. a fluid core and more rigid shell (see Figure 1.7).

In a separate study, using quite different tools, Falkenberg, Blinov and Leow studied another family of RNP ganules known as A2 granules [70]. The simulations involved three components: RNA (both with A2 specific binding sites and without), the A2 protein and a protein named TOG. The methodology of their study is set out Figure 1.9. In step 1, the authors utilised the Flory-Stockmayer theory to determine the percentage of each binding site type that were bound. Step 2 involved filling the simulation system with a predetermined concentration of each protein binding site. These binding sites were then permitted to bind such that the calculated binding percentages were satisfied. In step 3, the binding sites of each species, whether bound or unbound, were connected to form protein chains of set length. Therefore this process resulted in a list of proteins that were bound together, and by extension, the size and composition of the clusters formed. Due to the stochastic nature of steps 2 and 3, these steps were repeated multiple times to obtain more representative results. Additionally, as the probability of binding has a time dependency, the whole process was repeated to observe how clustering progresses with time [70]. This method is highly efficient, but does not account for conformational entropy, steric effects or the spatial orientation of the molecules.

Beyond these examples, there have been numerous computational studies, in addition to extensive theoretical studies, of the phase separation of polymers and model proteins. Some highlights of this research include simulations predicting the

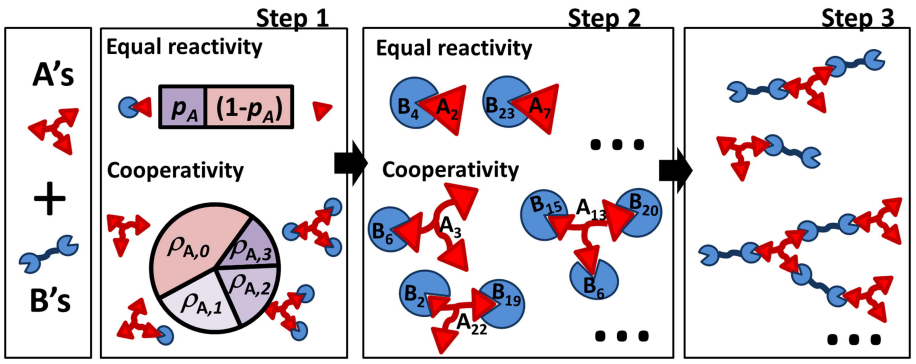


Figure 1.9: Schematic of deterministic and stochastic phase separation studies carried out by Falkenberg, Blinov and Leow [71]. The three steps are determining probability of binding, generating system of binding sites satisfying the binding requirements and then connecting the binding sites to form chains. Image taken from [70], DOI: https://doi.org/10.1016/j.bpj.2013.10.016. Copyright ©2013 by the Biophysical Society.

phase behaviour of IDPs from experimental radii of gyration data, utilising patchy particles to probe the role of RNA in phase separation [5, 6, 72] and investigating the impact of competitor species on phase separation [73].

1.2 Selectivity through multivalent binding

We now turn to focus more closely on multivalency and the mechanisms behind the overall binding behaviour that it promotes. Specifically, we are interested in the selective binding that can be achieved using a multivalent species.

1.2.1 Selectivity

Of late, significant insight has been gained into the selective binding achievable by multivalent species [74]. This selectivity depends on the receptor species present, but crucially, it also depends heavily on the concentrations of receptors at the host. Therefore, the effect differs significantly from the common monovalent 'lock and key' binding which tends to be very strong and specific and is primarily selective with respect to the type of receptor species regardless of its density. [21]

Multivalent selectivity is illustrated in Figure 1.10, showing that a multivalent species will only bind to a surface if the receptor density is above a threshold value. This is because, for this system, the individual bonds are too weak to bind the species to the surface; instead multiple bonds are required to attach the body strongly [74].

Multivalency is widely present within nature, but it is also drawing significant

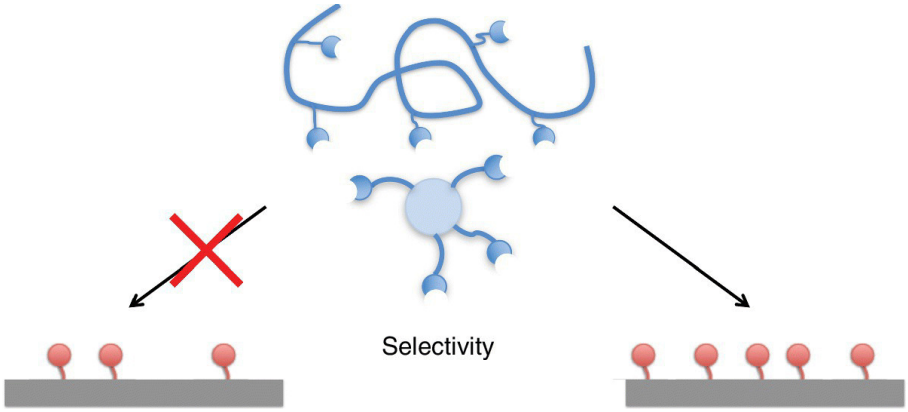


Figure 1.10: Schematic from [74], illustrating the ability of multivalent nanoparticles and polymers to bind selectively, whereby they only attach to surfaces with sufficiently high receptor density. Figure reprinted with permission from [74]. Copyright ©2018 by John Wiley and Sons Ltd.

attention from synthetic chemists and medical researchers, due to the fine control that it can achieve. A summary of some of the most promising avenues of research in this domain are discussed in this section.

Firstly, multivalency shows promise in the synthesis of biomolecules and nanostructures, through reversible adsorption and desorption of functionalised multivalent nanoparticles at surfaces [23, 75]. The multivalent binding controls the adsorption of the molecules and the desorption can be stimulated through electrochemical oxidation. Furthermore, functionalised nanostructures can also self-assemble in a controlled manor within bulk solutions, so this technology is not limited to surface chemistry [76]. This controlled binding has shown great potential for generating finely tuned nanostrucures structures for biomedical applications, such as drug delivery, theraputics and bioimaging [76, 77] and is highly desirable for the miniaturisation of electronic devices [75].

Secondly, we have already seen how multivalent binding of proteins and RNA can result in LLPS within cells and can dictate the structures and properties of the condensates [8, 43]. It is further implicated in facilitating the spatiotemporal control of components within the cytoplasm [78], a key interest of this thesis, and a requirement for the client-scaffold model.

Thirdly, and highly pertinent due to the current COVID-19 pandemic, multivalency appears to offer more effective general immunology [79, 80] when compared to the specific toll-like receptors which act to signal an immune response when specific proteins or antigens bind to the cells [81]. Toll-like receptors are those which only bind to a specific binding site. Moreover, multivalent protein antigens appear

(in many cases) to provide longer-lived immunity from vaccines than monovalent ones [82]. In an *in vivo* study of synthetic multivalent (polymeric) antigens, it was found that high valency antigens promote antibody production, where lower-valency antigens fail [83]. Additionally, multivalent antigens have been shown to give an enhanced T-cell activation — a crucial step in immune response [80]. Furthermore, cell signalling, as an immune response, is also an area where multivalent binding appears to play a fundamental role [83].

Fourthly, multivalency in medicine is proving to have great potential; multivalent nanoparticles capable of targetting tumour cells have been produced [84], as have targeted drug delivery vehicles [85] and materials to aid imaging [86]. Crucially, the reversibility of multivalent binding means a self-assembled drug delivery vehicle can be stimulated to break apart once inside the cells, and in this way they can release their payload [85]. Multivalent selective tumour treatment involves developing a drug or drug vehicle which only targets the diseased cells and not healthy ones due to the over-expression of specific proteins in diseased cells [87]. This in turn then reduces the damage of cancer treatment to the rest of the body.

1.2.2 Superselectivity

Multivalent selectivity is often accompanied by extremely responsive binding, known as superselectivity. Superselective binding involves a 'switch-like' transition, whereby a small change in receptor density on the surface results in a very sharp change in the probability of binding. Therefore, a client exploring a system containing a surface of receptors can go from never being bound to always bound within a very small receptor density range. The term superselectivity was first coined by Martinez-Veracoechea and Frenkel to describe the binding of a nanoparticle coated in many ligands binding to a surface of receptors [88].

Superselectivity is illustrated through the comparison of monovalent and multivalent binding curves in Figure 1.11, where the binding transition for the multivalent species is notably steeper than that of the monomer. The sharp binding transition, dependent on receptor density, is only possible for a multivalent species. Monovalent species lack the ability to bind in multiple equivalent ways required for superselectivity [74]. The on-off binding behaviour of superselective systems is much more rapid than the conventional Langmuir adsorption isotherm [89]. Since the discovery

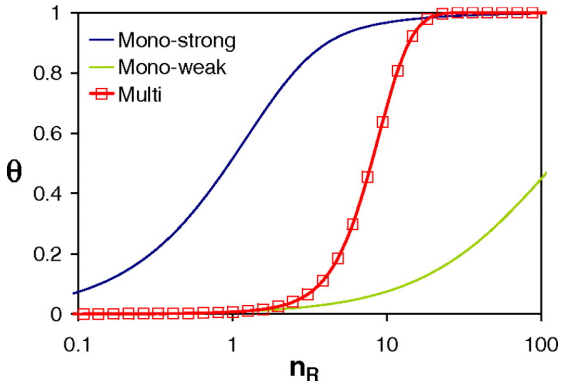


Figure 1.11: Typical superselective behaviour of model multivalent nanoparticles coated with ligands, binding to a surface of receptors. The variation in the percentage of occupied receptors (θ) against the surface density of receptors (n_R) is shown. The blue and green lines are for monovalent nanoparticles (with strong and weak interaction strengths respectively). These curves are significantly less steep than the red line for multivalent nanoparticles binding to an equivalent surface (where the number of ligands in the system is constant, and it is the number of nanoparticles that varies between the mono and multi results). Image taken from [88], doi: https://doi.org/10.1073/pnas.1105351108. Copyright ©2011 by Martinez-Veracoechea and Frenkel.

of the superselective binding transition, numerous different systems have been studied, including polymers binding to a surface [90] and the use of mobile receptors on a surface [91].

To explain why this superselective binding occurs, the example of a multivalent client (either a nanoparticle or polymer) is useful. On initial binding of the client to the surface (formation of its first bond) it loses a significant amount of entropy (translational and conformational in different proportions depending on the surface). Subsequent binding to the surface by the client is then more favourable due to cooperativity: the polymer does not lose as much entropy on forming additional bonds with the surface as it did with the first. There is thus a free energy penalty to binding which can only be overcome if the density of receptor sites on the surface (and multivalency of the client) is sufficiently high. Once this barrier is overcome, binding to the surface is highly favourable, explaining why the species then becomes predominantly bound at these higher receptor densities. This behaviour can be described in terms of concentrations of bound and unbound multivalent particles [88, 90, 92, 93], but can equally well be interpreted in terms of the binding probability of a single multivalent particle.

A multivalent species is bound to a surface if any one of its binding sites is attached to a receptor. The fact that there are so many combinations of binding

sites and receptors means that there is a high chance of an attachment existing even though any one connection is highly transient if the individual interactions are weak. Hence, for a multivalent species to be unbound, all of its binding sites must be unbound simultaneously. Consider a n-valent species binding to a surface with binding probability p_{bound} at each of its binding sites. The probability of the multivalent species being entirely unbound is given by

$$P_{\text{unbound}} = (1 - p_{\text{bound}})^n. \tag{1.1}$$

If p_{bound} changes smoothly from 0 to 1 in response to some external parameter (such as receptor density), Equation 1.1 changes increasingly sharply with respect to that parameter as n increases. Superselectivity exploits this combinatoric effect of the binding sites and receptors, which is a type of entropy.

The scope of superselectivity is huge, whether in enhancing the performance of the targeted drug delivery species discussed previously [94] or in other systems where step-like binding transitions are required, such as in diagnostics [93], where altering the system from being selective to superselective can vastly improve precision. In fact, work has already begun on using superselectivity to reversibly produce multi-component supramolecular polymers [95] and for the targeted binding of polymers at a surface [90]. Superselectivity could be utilised to overcome obstacles previously encountered in this field as a result of the selective binding transition being insufficiently sharp. This could give rise to significant improvements in precision and targeting in areas such as self-assembly of supramolecular materials, diagnostic tests and medical imaging processes.

1.2.3 Tuning multivalent binding

Multivalent binding can be controlled by the geometry of the binding target and the internal properties of the binding species. These considerations are highly relevant to this thesis so we shall expand on them here. In the case of nanoparticles binding to a surface, constraining the ligands of the nanoparticles and/or the receptors, can significantly impede binding, thereby reducing the overall attraction of the multivalent species to the surface [96]. Further studies have involved investigating the impact of the mobility of receptors on surfaces [97] and ligands on nanoparticle(s) [98], tethering the receptors to the surface [99] and changing the lig-

and length [100, 101], density, valency [102] and rigidity [103]. The parameter space of these problems is therefore huge. Additional complexity arises when considering that nanoparticles can also be coated with polymers in order to inhibit interactions, such as by coating them with polyethylene glycol (PEG) to prevent the nanoparticle binding non-specifically to receptors [104]. Moreover, nanoparticles expressing two different types of linkers could achieve binding specificity, whereby attachment to cells would only occur if the cell was expressing both receptor types at high enough densities [105]. Taking all this into account, it is clear that the binding behaviour of multivalent species is highly adaptable and open to manipulation. For a more complete overview of nanoparticles in medicine, including parameter optimisation, the reader is referred to [106].

An interesting finding from simulation work by Angioletti-Uberti, and an imperative consideration for those developing systems to utilise multivalency, was that non-selective binding can disrupt selectivity. This is due to the nanoparticle ligands binding to non-target receptors on the surface. One mechanism for overcoming this non-specific binding is to put receptors on the nanoparticle to compete with the cell receptors, as shown in Figure 1.12, thereby reducing the overall amount of non-specific binding and allowing the specific binding to control the binding behaviour again. These receptors grafted to the nanoparticle are known as 'protective' receptors. This process can be successful because the targeted interactions are often stronger than the non-targeted interactions so will not be as negatively affected by the 'protector' receptors competing with the surface receptors [107]. A futher non-trivial biological consideration, not to be overlooked in this field, is the environments that drug delivery vehicles operate in. As they travel through the bloodstream they get coated with proteins (a protein corona)[108], thus impacting the binding behaviour of the nanoparticle [109].

Although much of the experimental work to date on multivalent binding has largely focussed on nanoparticles, due to both their designability and their potential use as drug delivery vessels, many of the principles of superselectivity are also applicable to multivalent biopolymers, such as proteins. Crucially, as we focus largely on such polymers, the most relevant of the controlling factors are ligand length, valency, binding competition, confinement and stiffness.

One key consideration to bear in mind when comparing the binding of proteins and polymers to that of nanoparticles, is the ability for the former to change confor-

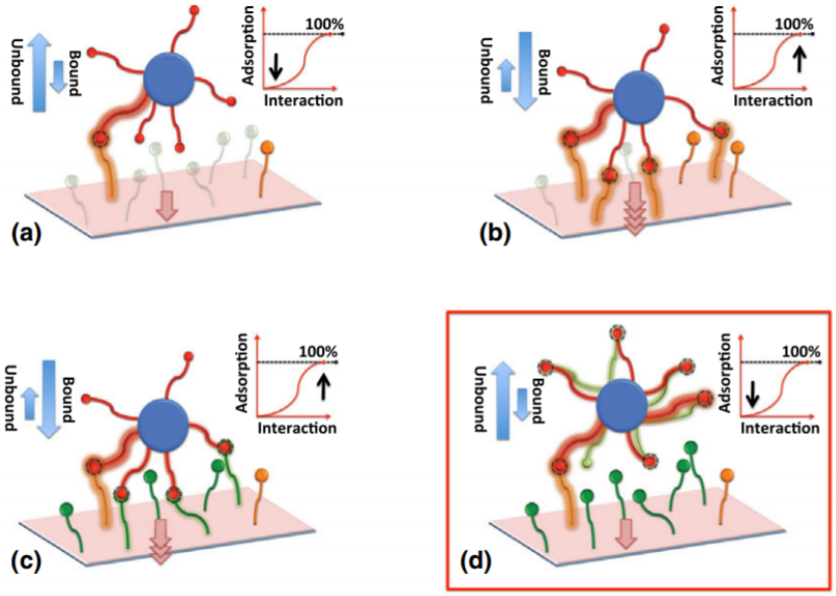


Figure 1.12: Schematic showing how deliberate competition between ligands can allow a system to regain selectivity after being lost due to non-specific interactions. Figures (a) and (b) are for specific binding only and clearly show density-dependent binding, weak in (a) and strong in (b). Non-specific binding — to the green receptors — is included in (c) and (d). Whereas in panel (c) the particle is bound strongly at low orange receptor density, in (d) the presence of competing ligands means that the nanoparticle is no longer strongly bound at the low receptor density. Reprinted figure with permission from [107], DOI: https://doi.org/10.1103/PhysRevLett.118.068001. Copyright ©2017 by the American Physical Society.

mation. One such example of this is due to pH changes, which can even be localised in certain regions — such as at a surface coated in a polymer brush [110]. From an entropic perspective, coating a surface in polymers and proteins can also significantly alter both the binding behaviour and geometries of the binding proteins. For instance, the binding, or indeed repulsion, of proteins to a surface coated in polymers can be tuned by varying the polymer length, rigidity and surface interactions [111, 112]. Additionally, surfaces coated in proteins have been shown in simulations to alter the conformations and geometries of proteins binding at that surface [113]. Beyond this, polymer behaviour is similar to that of nanoparticles on binding, with clear parallels between the properties of the polymers and the nanoparticle linkers on the binding behaviour observed.

A fundamental consideration in this thesis is of how weak binding, often associated with superselectivity, can be insignificant compared to the cumulative binding strength of the multitudinous other non-specific interactions, such as van der Waals and hydrogen bonding [107]. The impact of weak binding on selective binding

was investigated by Angioletti-Uberti, who proposed that coating the surface with inert proteins (such as PEG) can reduce these non-specific interactions. This is of biological relevance as most cells are coated in proteins and glycans, known as the glycocalyx, which can form polymer brushes to reduce non-specific interactions [114, 115]. Having multiple different types of ligand on the nanoparticle has been shown to allow the particle to target cells expressing each receptor at the correct density. Angioletti-Uberti argues this is less sensitive to biological fluctuations in receptor density — but that the main advantage is it allows the specific binding of nanoparticles to cells expressing the correct 'barcode' of receptors [107]. We will discuss this further in Chapters 4 and 5, where we study superselective systems.

1.2.4 Simulations of superselectivity

We now present a brief overview of the simulations which have been carried out to model and investigate superselectivity as background to the models used in this thesis. Simulations play an essential role in explaining the thermodynamics, kinetics and underlying behaviour of multivalent species because they allow systematic control of parameters and characterisations that are much harder to achieve experimentally.

The first such study is the seminal work by Martinez-Veracoechea and Frenkel [88] in which the sharp switch-like binding behaviour was observed, and first coined superselectivity. This was presented through a combination of an analytical model and Monte Carlo lattice simulations in the grand-canonical ensemble. Here we will focus on the simulation details to briefly describe their system. Nanoparticles coated in ligands were allowed to explore a flat surface coated in mobile receptors. The receptors were tethered to the surface by a harmonic potential, in the direction perpendicular to the surface, but were free to move across the surface within their lattice square. The nanoparticles were given an excluded volume, so could not penetrate the surface. The nanoparticle ligands were not explicitly represented; instead, whenever they bound to the surface it resulted in a tether (harmonic potential) forming between the surface and nanoparticle. The ligands and receptors were 'monovalent', so the multivalency is a result of the nanoparticle having multiple ligands. In this study, they found that for multivalent nanoparticles there was a sharp transition with increasing receptor density, at which the nanoparticles went

from being predominantly unbound to being predominantly bound to the surface. This was borne out in a sharp transition in the percentage of receptor sites occupied, as receptor density increased. The behaviour was sensitive to the nanoparticle valency, increasing with increasing number of ligands per nanoparticle [88]. This simple model system provided inspiration for the system used in this research.

The next example is that of the Monte Carlo soft Gaussian blob model in the grand canonical ensemble, by Curk and co-workers [91, 116], where polymers are modelled as Gaussian blobs connected by harmonic springs. Each blob represents one polymer, and binds to the surface by harmonic springs. The receptors are modelled as points on a surface and can bind to the blob if it is in range. The ligands are not explicitly represented, instead the polymer blob has a valency determining the maximum number of bonds it can form. Blobs can connect by harmonic springs to form longer polymers, and explore the system by translations and reinsertions. On calculating the free energy of the system to account for entropy, they found that the entropy loss of the polymer and receptor ligand (U_{poly}) decreased as the number of bonds formed increased, by which the authors were able to explain discrepancies between their simulation and analytical model results. This is the fundamental reason for superselectivity — lower entropic cost on subsequent binding due to high degeneracy. One limitation of this study is that it must be in the dilute/semidilute regime so there is not significant blob overlap with other blobs.

The last such study we shall summarise is a seperate soft-blob study by Curk *et al.* that involves the soft blob model with Langevin dynamics, to study the feasibility of utilising the superselective effect to develop a probe for pathogen genome (DNA) recognition [93]. A schematic of the system is shown in Figure 1.13. Using short, single-stranded, oligonucleotide DNA probes attached to a surface — with binding sites at spacings that will favour the desired bacterial DNA (single-stranded genomic DNA) — the authors can successfully generate a highly specific sensor. By tuning the genome length (fragmentation) and the length of DNA strands on the surface the superselectivity can be optimised. Short DNA strands grafted to the surface work better to distinguish between different bacteria — in addition to low fragmentation of the bacterial DNA. Short DNA strands have a lower entropic barrier to binding, therefore the difference between the correct genome and the undesired one is lessened at short lengths because neither can form many bonds, whereas long strands are unlikely to bind unless they are from the correct genome and can form many bonds

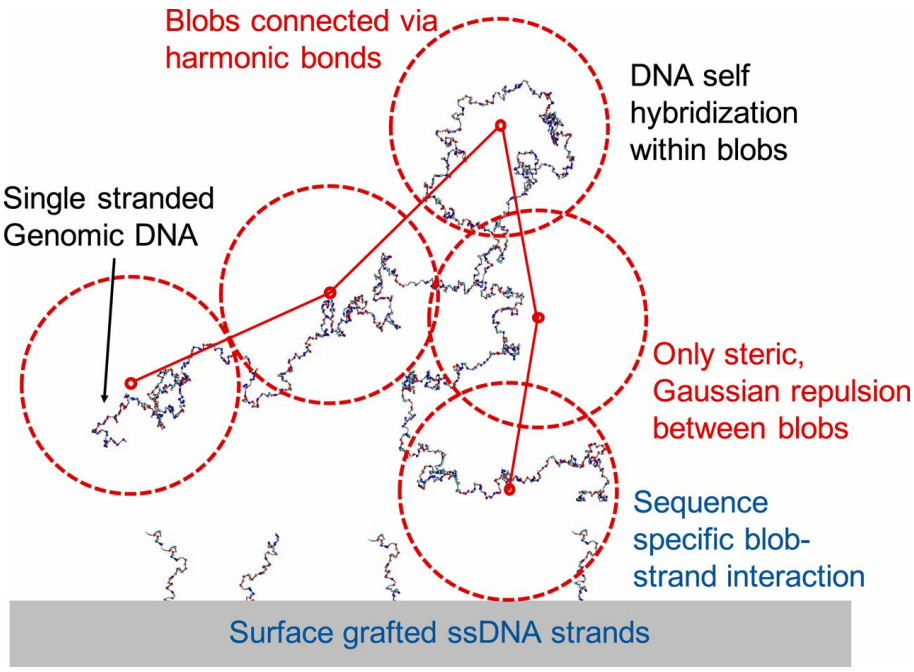


Figure 1.13: Schematic of the soft blob study on the binding of multivalent DNA to a surface. Figure reprinted from [93], DOI: https://doi.org/10.1073/pnas.1918274117. Copyright ©2020 the authors. Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0.

to offset the high entropic penalty. Interestingly, these simulations indicate that such a surface should be able differentiate between different strains of the same bacteria, but this requires long DNA on the surface and long sections of target DNA. [93] Another potential use of superselectivity is in diagnostics where DNA-based probes have been shown to effectively target specific pathogens or genes.

1.3 Project aims and workplan

In the following chapter, the methodology of the investigations in this thesis are detailed, alongside justification of the simulation set-up. This involves an in-depth description of cubic lattice systems, the moves carried out and the merits of Monte Carlo (MC) lattice simulations.

Despite the research that has been done so far into LLPS and membraneless organelles, questions still remain about the fundamental physics behind the components of the cytoplasm phase separating. Therefore, in Chapter 3, the phase behaviour of multivalent species will be investigated using MC simulations on a cubic lattice. This will involve simulating two coarse-grained associative polymer species, thereby allowing exploration of the effect of polymer valency, interaction strengths,

linker length and linker type on the phase transition of these components. Furthermore, the structures formed during this phase separation will be scrutinised, allowing conclusions to be drawn on the relationship between the polymers and the crude representations of membraneless organelles that they form. This is with an aim to discover how to promote or suppress phase separation, and to uncover whether these properties can be tuned to promote client recruitment.

There is still much to be understood and many more potential applications to explore with regards to multivalent binding and superselectivity. In Chapter 4 we investigate whether superselective binding to a 3D host of receptors is possible, in a significant departure from the conventional focus on surfaces. This has important implications for membraneless organelles and their formation by LLPS; moreover, we propose superselectivity as a mechanism by which membraneless organelles recruit and expel clients due to environmental changes in the cell, as described in [32]. Due to the step-like nature of superselectivity this would facilitate complete expulsion or recruitment (to saturation) of proteins and enzymes in the cytoplasm which has been observed experimentally [34].

Following this, in-depth exploration of superselective binding of polymeric species to a surface will be presented in Chapter 5. Most notably, we study binding at structured surfaces, such as penetrable and pitted surfaces. We also investigate whether superselectivity could be utilised to sort polymers, based on their intrinsic properties. This sorting behaviour has potential applications within analytical chemistry and in separation processes.

Finally, the conclusions of this work will be presented alongside directions for future work in Chapter 6. We will address both the studies that could be carried out in the MC cubic lattice system we have been working in, and the useful studies that could be undertaken using different models and tools to further the scientific understanding in these fields.

Chapter 2

Models and Methods

Monte Carlo (MC) lattice simulations in the canonical ensemble are used throughout this thesis. Whilst the nature of the system and exact set-up vary between the calculations, the foundations of the simulations remain unchanged. They are all coarse-grained studies of proteins and other polymers on a 3D cubic lattice, where the polymers are modelled as binding beads connected by linkers, illustrated in Figure 2.1 as the archetypal system of study.

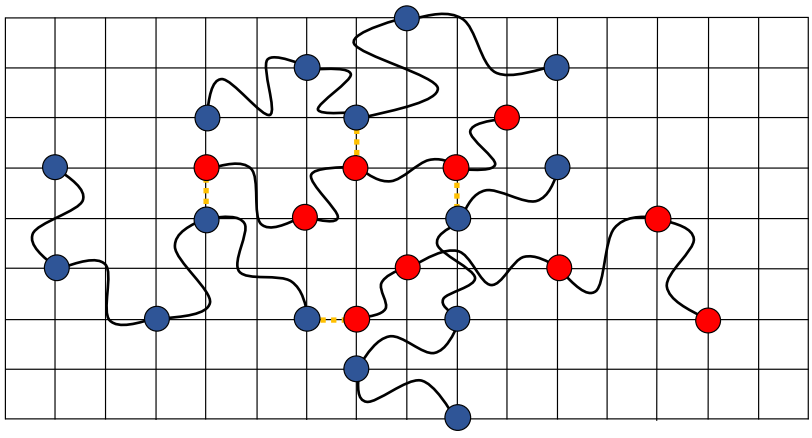


Figure 2.1: 2D representation of typical 3D lattice system studied, where the yellow dashed lines represent interactions, the curved black lines are linkers and the circles are polymer binding beads.

2.1 Coarse-grained simulations

There are two main motivating factors for utilising coarse-grained simulations for the polymeric systems of interest in this thesis, rather than simulating them at an atomistic resolution. Firstly, the creation of an idealised model captures the key factors that control the system behaviour. This allows general and widely applicable findings to be reached, rather than generating results that are specific to a single system. Secondly, the reduction in the number of degrees of freedom greatly reduces the computational expense compared to atomistic models, thereby allowing us to study larger systems. In brief, they offer conceptual clarity and are efficient [117].

Coarse-graining can take many forms, and can be implemented at various levels of resolution, depending on the system of study [117–119]. One of the finest levels of coarse-graining is to approximate molecular groups as single beads or entities, such as in the united atom model. For example a methyl group can be represented by a single bead. The MARTINI model is a coarser model which approximates larger atomic groups into single entities [120]. Due to the large number of atoms and functional groups within the polymers we wish to simulate, this coarse-graining is insufficient for modelling systems at a level where we can draw widely applicable conclusions to a range of similar systems — the results obtained from such coarse-graining would still be very system specific.

A higher level of coarse-graining is to model sections (such as monomers) of a polymer as single entities, which can be connected by linkers of various forms. This is an example of a 'beads on a string model' [121]. This approximation removes the configurations available to the linkers as well as the internal configurations of the coarse-grained sections, thereby reducing the total entropy of the polymer. Therefore, this coarse-graining is most effective when the overall polymer shape is key to the system behaviour, rather than the individual atoms and side chain configurations. Due to the coarse-graining of entire binding domains/regions, the potential energy between beads must be an effective potential incorporating multiple components of interactions, such as van der Waals, electrostatic, hydrogen-bonds and internal entropy etc. Coarse-graining reduces the number of spatial configurations that the polymer can occupy and results in a coarse-grained energy landscape, allowing for the overall system behaviour to be probed, without minor configurational changes hampering the rate of exploration [118]. This level of coarse-graining has

been extensively used to study polymer folding [122], and in polymeric phase separation [1, 68]. It enables more general conclusions to be made, which we desire for the project, and allows for system behavioural changes to be attributed to well defined intrinsic polymer properties, such as a change in polymer length or interaction potential.

The coarse-grained model can be further simplified by restricting the polymer binding beads to a cubic lattice [123]. This allows for the discretisation of space, which increases the rate at which configurational space can be sampled, thus facilitating simulations of far greater size and timescales compared to off-lattice models, without losing the overall physical properties of the polymer. Given the numerous successes of modelling proteins on cubic lattices [124–127, 127–131], they are an attractive model for the system we wish to study.

2.1.1 Successes of lattice models

The following examples give some justification for using such a model in our studies. A significant achievement of simulating polymers on a cubic lattice is the reproduction of the universal polymer scalings predicted by de Gennes [132] for the mean squared polymer radius of gyration R_{G}^2 and end-to-end separation $\langle R^2 \rangle$, with respect to the polymer length. This indicates the reliability and relevance of cubic lattice polymer models in predicting polymer behaviour. These polymer scaling relations were used within this work to benchmark the code. This is pertinent as we model polymers with different linker types, each with different scaling exponents.

Wide-ranging studies into protein folding have been carried out on cubic lattice systems. Such work includes investigation of protein designability, determining the number of different amino acid sequences that can fold into a given spatial configuration. Designability allows nature to favour protein structures and shapes which are thermodynamically and evolutionarily stable [124]. Lattice simulations have allowed many studies of this kind, where huge numbers of sequences can be tested against numerous conformations. Lattice studies have identified a relatively small number of highly designable lattice polymer shapes amongst a plethora of low or non-designable conformations [133], giving some insight into protein designability within biology. This has also been done by monitoring the protein contact trace (a protein fold's tertiary topology) which is highly correlated to the designability [134].

Successes in protein folding simulations on a cubic lattice include further insights into the kinetics [125–127] and thermodynamics of folding [135]. The generality of lattice models has allowed for the kinetics of folding to be predicted based on the equilibrium collapse temperature and the folding temperature [136]. Both of these temperatures can be estimated from experimental results. Lattice models have also been used to investigate the Levinthal paradox; there are a huge number of protein conformations available but proteins fold on very short timescales (on the order of milliseconds to seconds), implying that the search for the folded conformation cannot be random [137]. Lattice simulations enable comprehensive coverage of configurational space, which would be prohibitively expensive without coarse-graining [127–130].

Lattice simulations have been used to probe the role of folding cooperativity in protein behaviour, where folding cooperativity describes an "all-or-nothing" transition from an unfolded to a native state as opposed to proteins folding partially or in a step-wise manner towards their native state [138]. Stemming from this, the cooperativity of side chain interactions has been found to be crucial in achieving the sharp folding transitions into well-defined configurations, often observed in biology, as opposed to random coil to globule transitions [136].

In addition to single-species studies, lattice models have been successfully used to study the global behaviour of multi-protein systems. This includes both homopolymer phase behaviour [131], as well as multicomponent polymer mixtures [139]. For example, 3D cubic lattices were used to obtain phase diagrams of two species of associating polymers in a solvent (as predicted by Flory [140]) and the corresponding equations of state, connecting the system energy to temperature [141]. Additionally, these models have allowed for the efficient exploration of protein packing. The folding of proteins, to facilitate packing, is hindered by side chains, as folding involves restriction of the side chains' freedom of movement resulting in a large entropic barrier [142].

Thorough exploration of polymer entanglement and knotting has been undertaken using cubic lattice polymers [143]. For instance, cubic polymer studies have allowed research into the kinetics of folding and, from this, knotting was shown to occur late in the folding process once many contacts had been made [144, 145]. Furthermore, investigations into mutations which promote or prevent folding, through knotting, have also been informed by cubic lattice simulations [146]. These studies

have contributed to the rapidly expanding field of knotting and entanglement of polymers and proteins [143].

2.2 Model

2.2.1 Interaction beads

Returning to the model used in this thesis, the binding beads of the biopolymers are represented according to three different interaction cases. Two of these cases are specific and directional interactions. The final case is isotropic binding. One motivation for modelling directional and isotropic interactions is that it allows us to determine to what extent the behaviour we observe is controlled by the binding mechanism. The other motivation is so that we can represent a wider range of multivalent polymer species. Specific, directional interactions are employed by proteins with well defined binding regions, as in the model experimental proteins constructed of repeated binding domains used in [7]. In contrast, intrinsically disordered regions of proteins, which have no stable conformation to fold into, are believed to interact much more promiscuously with the other species in the system, as well as with each other, and can thus be modelled as binding isotropically [53]. For the specific case of membraneless organelles, species with both binding types have been observed [147].

The interaction H_{ij} between two beads i and j of types a and b respectively is:

$$H_{ij} = -\delta_{ij}\varepsilon_{ab}. \tag{2.1}$$

In Equation 2.1, ε_{ab} is the interaction strength between beads of type a and b. The delta δ_{ij} determines whether or not the two beads are bound, taking the value of 1 for bound beads and zero for unbound beads. In the majority of work undertaken for this project the interaction strength has a single non-zero value for all beads that may bind. For simplicity, the interaction energy is expressed in the dimensionless form $f_{ab} = -\varepsilon_{ab}/kT$, where $\varepsilon_{ab} > 0$. Therefore the pairwise energy can be expressed as

$$H_{ij}/kT = \delta_{ij} f_{ab}. \tag{2.2}$$

The first of the two directional binding cases is where the bonding is determined

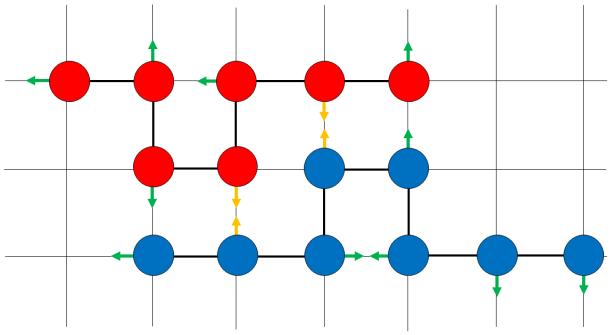


Figure 2.2: Schematic of two neighbouring polymers showing the direction of their bonds at an arbitrary time as green arrows. A pair of red and blue beads are bonded directionally if the arrows point towards each other — and these arrows are shown in yellow. This figure relates to the vector binding model.

by a single binding vector protruding from each binding bead. This vector must point along the x, y or z axis, in either direction. If two beads are on adjacent sites, and the vectors point towards each other, as shown in Figure 2.2, then they are deemed to be bound, with the magnitude of the interaction given by H_{ij}. We will refer to this as the vector binding model.

The second specific, directional binding case is implemented in order to align our simulations more closely with the approach described by Martinez *et al.* [88]. This facilitates more straightforward comparison between our results and existing studies, as the entropies of this binding model and the vector model differ. This non-vector binding involves the generation of a list of adjacent beads (which have the capacity to bind), alongside a list of bound beads. Pairs of beads are then randomly chosen and bond breaks or formations are attempted.

Isotropic binding is arguably the simplest to model. In this case, δ_{ij} is 0 for any beads that are not adjacent and 1 for those that are. Therefore, all adjacent beads are bound with interaction strength f_{ab}, and, unlike the directional cases, a single bead may be involved in two or more bonds simultaneously.

2.2.2 Linker types

Despite the geometric restrictions of 3D lattices, there are a number of ways to model the linkers between the binding beads on the polymers we study. The three linkers we use are: (A) hard linkers where the linkers, of length l, are explicitly modelled with excluded volume; (B) ideal linkers consisting of linker beads being explicitly modelled but having no excluded volume; and (C) implicit linkers where

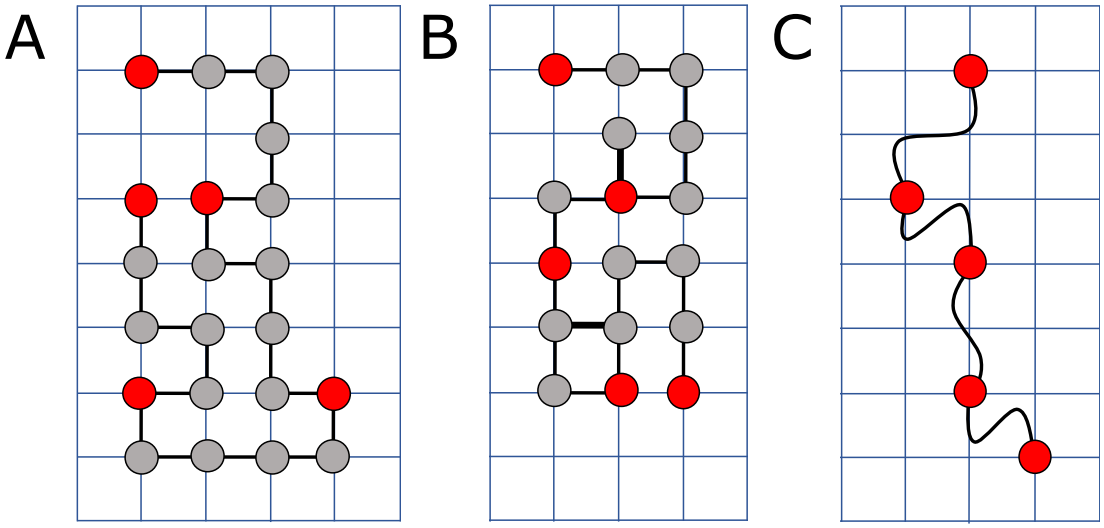


Figure 2.3: Schematic of the three different linker types, where red beads are binding beads, the thick black lines are linkers and the grey beads are linker beads. The linker types are (A) hard, (B) ideal and (C) implicit. Lines of increased thickness are used in (B) to show where the linker route is retraced, due to overlapping beads. The thick black lines in (C) are not explicitly modelled but are shown here to indicate that the beads are connected, albeit by invisible tethers.

the linkers are invisible tethers which occupy no volume and solely act to prevent adjacent binding beads from being too far apart. These models are illustrated in Figure 2.3. Using a variety of linkers allows a wider range of proteins and solvent conditions to be roughly replicated. The different linker types can also be used to show whether the behaviour observed in the simulations is linker dependent or universal.

Hard

Hard linkers, as shown in Figure 2.3A, are modelled as beads with excluded volume equal to that of binding beads. These form a contiguous chain connecting the binding beads, and are therefore self-avoiding random walks. All beads forming the polymer must be exactly one lattice spacing from their nearest neighbour, and this cannot be across a lattice diagonal. Hard linkers have several advantages; they capture the linker entropy of the polymer and they have been extensively studied alongside being verified as reliable for replicating polymer behaviour. They also account for polymer entanglement [148] and occupy an expanded configuration [68]. However, self-avoiding walks (SAW) are computationally expensive to model and their scaling behaviour is not suitable for some of the proteins we desire to study,

especially in a poor solvent; therefore they are not universally applicable.

Ideal

Ideal linkers are similar to hard linkers, but crucially the linker beads can overlap with any other bead, see Figure 2.3B, and these are therefore random walks (RW). The polymers with ideal linkers tend to be more compact than their corresponding hard polymers, and are therefore more representative of polymers in an unfavourable solvent or where linkers are much smaller than the binding sites. They do not account for entanglement of polymers, and as such they perform best for short polymers and for simulations in the dilute limit, where entanglement plays a smaller role. Both hard and ideal linkers account for the entropy associated with the route the linker follows to connect two binding beads on a polymer; this is in contrast to the implicit linkers which we shall discuss next.

Implicit

Polymers with implicit linkers have binding beads, with an exluded volume, connected by invisible tethers, Figure 2.3C. These tethers occupy no volume but prevent adjacent beads from moving too far apart i.e., they capture the overall length of the linker via a maximum extension. Implicit linkers are advantageous in that they are computationally much less expensive to simulate than explicit linkers, and Holehouse *et al.* found that they represented the solvation volumes of common stress granule proteins better than the hard linkers [68]. They allow for simulations in dense areas where excluded volume can make such studies prohibitively expensive. Despite their advantages, two key considerations to be made when using these linkers are that they fail to account for polymer entanglement and other excluded volume effects, and their behaviour does not account for the linker route entropy.

Linker scaling properties

The critical scaling parameters for polymer radius of gyration and end-to-end separation are essential in the testing of our systems to ensure the polymers are being modelled correctly. The radius of gyration $R_{\mathrm{G}}(N)$ of a polymer of length N can be found using:

$$R_{\mathrm{G}}^2(N) = \frac{1}{N}\left\langle \sum_{k=1}^{N} (\mathbf{r}_k - \mathbf{r}_{\mathrm{COM}})^2 \right\rangle, \tag{2.3}$$

where $\mathbf{r}_k$ is the position of bead k and $\mathbf{r}_{\mathrm{COM}}$ is the centre of mass of the polymer. The angular brackets indicate that the value of R_{G} is averaged over all configurations. The end-to-end separation is a simple measurement of the separation between the two polymer ends (beads 1 and N).

The mean squared end-to-end separation $\langle R^2(N)\rangle$ of a self-avoiding random walk (SAW) polymer consisting of N segments of length a is proportional to $a^2N^{2\nu}$. Similarly, the radius of gyration $R^2_{\mathrm{G}}(N)$ is also proportioanl to $a^2N^{2\nu}$, but with a different constant of proportiality to the mean squared end-to-end separation [140]. The exponent, ν, for 3D lattices in the mean field approximation is 3/5 [132]. Using lattice polymer simulations, a more accurate scaling law for a 3D SAW of $\nu = 0.588$ has subsequently been obtained [149].

Random walks (RW), which correspond to a freely jointed chain, are similar to SAWs, but with the distinction that they have no excluded volume, so beads can overlap with each other. Hence, self-avoiding walks are swelled compared to random walks. This means that the scaling of a RW on a cubic lattice is governed by slightly different equations to a SAW. For a random walk, the mean squared end-to-end displacement is $\langle R^2(N)\rangle = a^2N$. This can also be presented as $\langle R^2(N)\rangle = a^2N^{2\nu}$, where ν is equal to 1/2. A similar equation exists for the square of the radius of gyration, $R^2_{\mathrm{G}}(N) = (a^2/6)N^{2\nu}$ [150] with $\nu = 1/2$, for a random walk in 3D.

2.3 Methods

2.3.1 Monte Carlo overview

We use Monte Carlo simulations to generate statistics on the configurations of the polymeric systems of interest in this thesis. In practice, this involves making trial moves, and accepting or rejecting them due to the energy change resulting from the move. The acceptance criterion used is the Metropolis algorithm, Equation 2.4 [151]. Trial moves within the present work include polymer and bead displacements, in addition to bond formation and breaking, the complete descriptions of which can be found in section 2.3.2. The probability of accepting a trial move P^{acc}_{ab} with energy change ΔE_{ab} from state a to b at temperature T is [152]:

$$P^{\mathrm{acc}}_{ab} = \min[1,\ \exp\left(-\Delta E_{ab}\,/\,kT\right)]. \tag{2.4}$$

The Metropolis algorithm accepts energetically favourable moves whilst only accepting some energetically unfavourable moves, depending on the energy change. A crucial criterion of Monte Carlo simulations is that they must be ergodic, meaning that all states within the system are accessible to each other through the moves available [151].

The probability of the system being in state i, P_i, is given by the Boltzmann distribution:

$$P_i = \frac{\exp\left[-E_i/kT\right]}{\Sigma_j \exp\left[-E_j/kT\right]} \tag{2.5}$$

where E_i is the energy of state i and Σ_j is a sum over all discrete states in the system. A well equilibrated Monte Carlo simulation should reproduce the Boltzmann distribution.

The moves, and the beads/polymers on which these are attempted, are chosen randomly to obey detailed balance, which dictates that the probability π_{ij} of going from state i to j is related to the reverse move probability π_{ji} by

$$P_i \pi_{ij} = P_j \pi_{ji}. \tag{2.6}$$

Finally, in order to generate the equilibrium statistics of a given system, MC simulations include an equilibration stage, during which the polymer species explore the system through random MC moves, followed by a production stage, during which the statistics are gathered.

2.3.2 Initialisation and local moves

Initialisation of the system involves the random insertion of polymers, in a bead by bead manner, onto the lattice, as illustrated in Figure 2.4. Each polymer is grown from a random starting point, and growth is restarted if there are no unoccupied sites adjacent to the last bead placed.

A number of moves are available for simulating polymers on a lattice. The choice of moves is dictated by the simulation model and the system of study, and should allow for efficient and ergodic simulations. These considerations include linker type and the system density, where the requirements for polymer melts will differ for those of single polymer studies [153]. The first moves we shall discuss are for the hard and ideal linkers, both of which are represented explicitly.

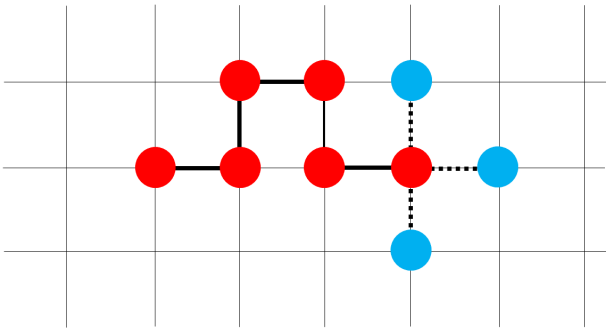


Figure 2.4: 2D representation of a step in the formation of a polymer on a 3D cubic lattice. The red beads have already been placed and the blue sites are the potential locations that are allowed for the next bead.

Once initialised, the polymers explore conformational space. In similar studies by Coluzza *et al.* [154], the following moves are carried out: right angle flip, crankshaft rotation, pivot. We choose to also use these moves, and additionally employ the reptation move [155]. These four moves will now be described in turn.

The right angle move consists of selecting a bead, and if the beads adjacent to it are at right angles to each other, the bead is flipped 180° around the axis which connects the two neighbouring beads, as shown in Figure 2.5A.

The crankshaft move involves rotation of a section of the polymer delimited by two non-bonded beads that lie along a Cartesian axis. This rotation can be by 90°, 180° or 270° around the axis, as illustrated in Figure 2.5B.

The pivot move is utilised as it was shown to be an efficient method of exploring configurational space, despite having a low acceptance rate, because the moves that are accepted are usually a significant alteration to the polymer shape [156]. To carry out this move, a bead is chosen at random and the shorter end of the polymer is rotated around this bead, as illustrated in Figure 2.5C. The shorter end of the polymer is rotated instead of the longer end for computational efficiency. The polymer section can be rotated by 90° , 180° or 270° around the x, y or z axis.

The reptation move involves random selection of either the first or last bead of the polymer. A new bead is then placed on one of the available sites neighbouring the chosen end of the polymer, such that it becomes the new end bead. The end bead at the opposite end of the polymer is simultaneously deleted as depicted in Figure 2.5D. This is equivalent to the polymer moving in a new direction and the rest of the polymer following along its path.

During the simulations, the moves types are selected at random, to obey detailed

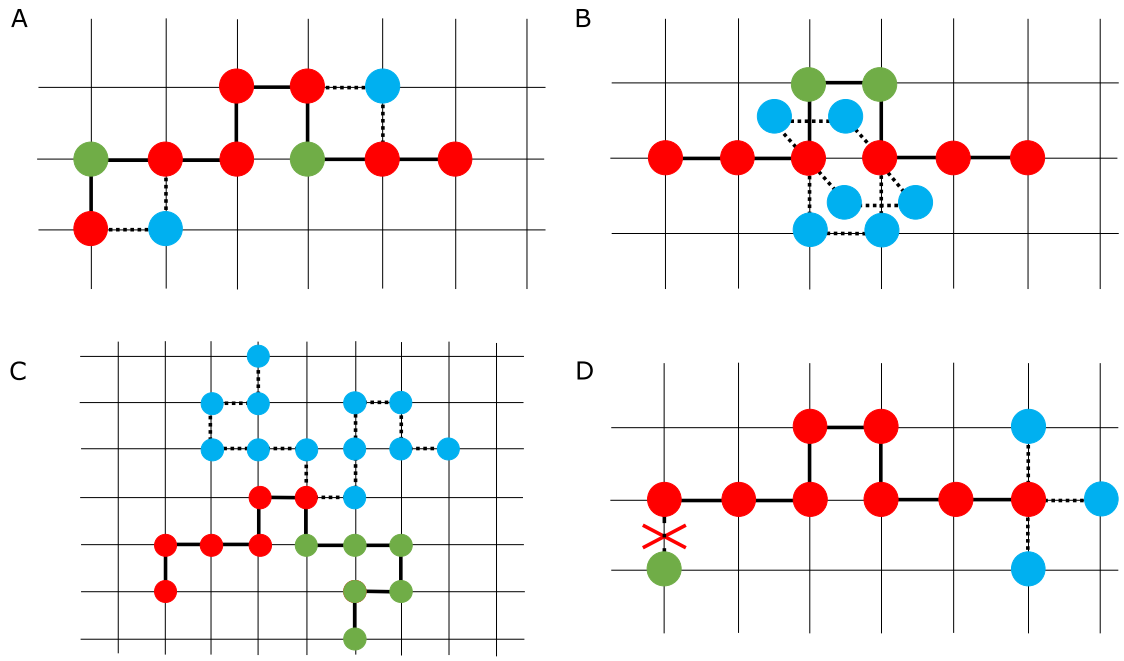


Figure 2.5: Images of individual polymer moves, with red and green beads connected by solid lines representing initial polymer configurations, and the blue beads connected by dotted lines indicating potential new configurations of the polymers, to replace the green beads) after undergoing one of the four moves of (A) right angle flip, (B) crankshaft rotation, (C) pivot or (D) reptation.

balance. Similarly, after the move type is chosen, the bead upon which the move is attempted is selected at random. Every move is proposed as a trial move, and the new positions are tested for any overlaps and to calculate the energy change of the move. The Metropolis algorithm, Equation 2.4, is then used to determine whether to accept or reject the move.

Implicit linkers require different initialisation and moves, since there is no series of adjacent linker beads connecting the binding beads. Therefore, some of the above procedures are neither possible nor efficient. Initialising implicitly linked beads involves stepwise positioning of the beads at a random displacement from the previous bead, such that the linkers and excluded volume rules are satisfied. One move deployed is a bead repositioning move, whereby a bead is selected at random and then displaced, Figure 2.6A. If the proposed site is unoccupied, then the linkers between the beads are checked to ensure that they do not exceed the maximum length. Following this, the move is either accepted or rejected using the Metropolis algorithm, as before. The other move is reptation, Figure 2.6B. This is similar to the explicit linker reptation move, with the main difference being the bead added to the end of the polymer is positioned at a random displacement (within the linker range) from

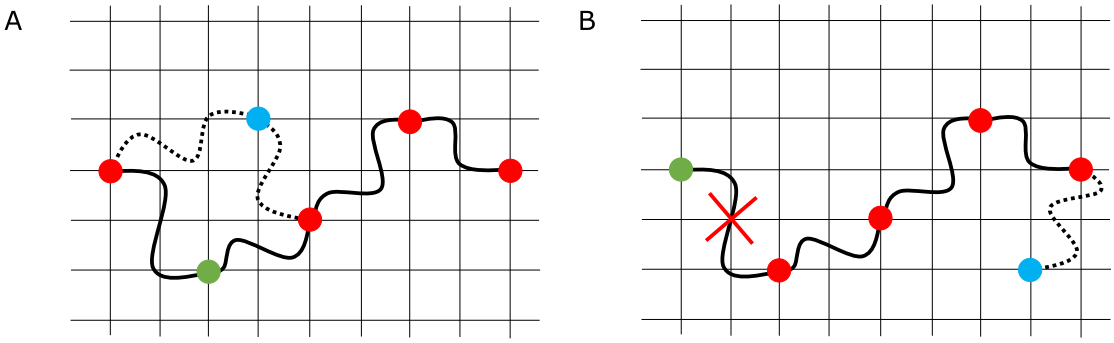


Figure 2.6: 2D representation of individual polymer moves. The polymers have implicit linkers which connect the red and green binding beads, representing the initial polymer configurations. The blue beads connected by dotted lines indicate potential new configurations of the polymers, after undergoing (A) a single bead displacement and (B) a reptation move.

the end bead.

For all of the linker types, polymer translation and regrowth are employed. Polymer translation is a simple process of moving the entire polymer by one lattice spacing in the x, y or z direction, whilst maintaining the polymer geometry. Polymer regrowth involves deleting the polymer and randomly selecting a point on the lattice from which to regrow the polymer. This regrowth occurs in the same manner as in system initialisation, where new beads are positioned randomly, within linker range, in a step-wise manner until the whole polymer has been grown. Any bead overlap results in the move being rejected.

Bond formation and breaking moves depend on the binding model implemented. For the vector model, two adjacent beads are considered bound if their binding vectors point towards each other. The vectors can change directions during some polymer moves, such as the pivot or crankshaft move, as well as during vector rotation move, which involve the random reassignment of the vector direction.

For the non-vector directional bonding, a list of adjacent beads and a list of bound bead are constructed. These lists are generated after every MC move. Bond breaking moves involve randomly selecting a pair of beads from the bound pair list. This bond is then broken with probability

$$P_{ab}^{\text{acc}} = \min[1, \omega \exp(-\Delta E/kT)]. \tag{2.7}$$

ΔE is the change in energy of the move and $\omega = N_{\text{bound}}/[N_{\text{unbound}} + 1]$, where N_{unbound} and N_{bound} are the number of unbound and bound pairs, respectively. This

weighting is required for the algorithm to obey detailed balance, as otherwise the probability of a bond being formed or broken would depend on the number of bound or unbound pairs (as pairs would be selected more or less frequently depending on the list length). Similarly, for bond formation, a pair of unbound beads are selected at random from the unbound pairs list and a bond is formed with probability determined by Equation 2.7. In this case $\omega = N_{\text{unbound}}/[N_{\text{bound}} + 1]$ [151].

No bond formation or breaking moves are required for the isotropic binding case.

2.3.3 Cluster moves

In systems containing multiple polymers capable of aggregating, local moves are insufficient for the polymers to fully explore the system. We observe a dispersion of small clusters forming which are spatially distant. This indicates that the polymers are unable to move between clusters without significant energy penalties, making such a process unlikely. To overcome this, cluster moves are used. The first step in this process is allocating polymers to clusters at each snapshot (the system configuration after a given number MC moves). The cluster determination is done on the basis of the bonds between the polymers; all polymers which are connected by a continuous pathway of linkers and bonds are in the same cluster. This process has the added advantage of allowing the number of clusters and the cluster population to be monitored simultaneously.

The two simplest cluster moves available are translational and rotational. The translation move involves the cluster moving in the x, y or z direction by one lattice spacing. The rotation move involves a rotation around the cluster's centre of mass by 90°, 180° or 270° around the x, y or z axis. The rotation move is computationally expensive without improving equilibration significantly. Therefore only the cluster translation moves are used in our simulations. As with the local moves, a polymer is chosen at random and the cluster that this polymer belongs to undergoes the cluster translation. One issue with this technique is that larger clusters will move most often, as there is an increased probability of one of its resident polymers being chosen at random. To overcome this, the probability of the cluster move being carried out is inversely proportional to the cluster size.

2.3.4 System exploration

Another consideration in running MC simulations is that at sufficiently low temperatures, the system can get 'stuck' in a potential energy well. This can be identified through monitoring system sampling and the convergence of measurable parameters. However, there are cases where exploring a system's configurational space becomes impossible, or is highly unlikely. Here, more advanced techniques must be used. Where sampling issues occur in this work, parallel tempering (replica exchange MC) [157] is used to overcome this. Parallel tempering involves running multiple simulations simultaneously at different temperatures and attempting to exchange the configurations of different temperatures. This technique allows for the full exploration of systems which would be impossible using single-temperature Monte Carlo method described earlier. Parallel tempering is discussed in greater detail in Section 5.1.1.

2.3.5 Benchmarking the code

The codes used within this work were written from scratch. The codes to run the Monte Carlo simulations were written in Fortran, with analysis being conducted primarily using Python scripts. The run times for the simulations included in this work ranged from minutes to several days. Due to the number of simulations run, the use of the Durham University Supercomputer, Hamilton, was essential. The codes remain with the author and will be shared, on request. As these codes were newly developed, extensive testing was required. One such benchmarking procedure is now described.

The lattice polymers used in this thesis are tested to ensure they scale with the exponents described in section 2.2.2. This involves separately simulating individual polymers, of lengths ranging from 100 to 1600 lattice spacings, as they explore a large simulation box using the moves described in section 2.3.2. The gradient of the best fit line in a log-log plot of $R_{\mathrm{G}}^2(N)$ and $\langle R^2(N)\rangle$ against polymer length gives the scaling parameter ν. An example plot is shown in Figure 2.7 for the radius of gyration of a polymer with ideal linkers and binding beads with excluded volume. The corresponding values of ν for all the linker types used in this thesis are shown in Table 2.1. Good agreement is seen between the scaling of our hard linkers and ideal linkers with the expected theoretical values. The inclusion of the

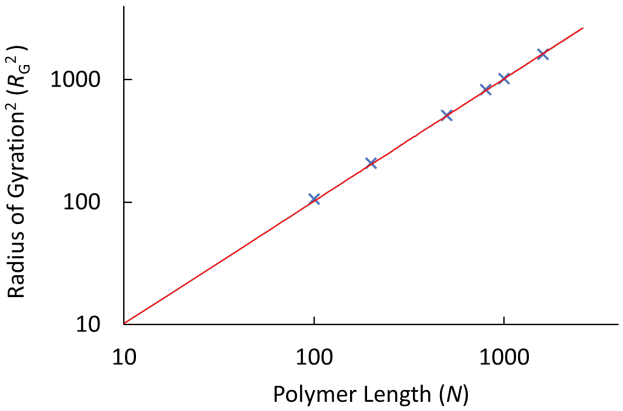


Figure 2.7: Log-log plot of the radius of gyration squared for a range of lengths of polymer, where the polymers have ideal linkers and the binding beads have excluded volume. The red line is the linear best fit of the data. Simulations for other linker types lead to a similar plot, albeit with different scaling exponents.

scaling of ideal linkers with binding beads, is to account for the fact that polymers with ideal linkers and binding beads with an excluded volume are not true random walks. Therefore their scaling properties are slightly different, largely resulting from the significant impact of the excluded volume of the binding beads on $R_{\mathrm{G}}^2(N)$ and $\langle R^2(N)\rangle$ for polymers with short linkers compared to a pure random walk. However, the agreement of the scaling of the ideal linkers (both with and without the binding beads) with the predicted scaling of a random walk model indicates that the code is correctly modelling these linkers.

Table 2.1: Scaling parameters of the different linker types in 3D, studied as part of the code benchmarking. Simulations results are shown alongside the theoretical scalings predicted by Flory and Volkenstein [158] and by le Guillou and Zinn-Justin [149].

Linker Type	ν ($\langle R^2(N)\rangle$)	ν ($R_{\mathrm{G}}^2(N)$)
Hard	0.594(1)	0.594(3)
Ideal	0.5055(6)	0.497(2)
Ideal (with binding beads)	0.5055(6)	0.505(5)
Implicit	0.588(2)	0.54(1)
SAW (Theory) [149]	0.588	0.588
RW (Theory) [158]	0.5	0.5

2.3.6 Efficiency and convergence

The efficiency of the code is of paramount importance as it determines the maximum system size that can be modelled. The acceptance rates of the moves aid in determining the code efficiency. This is because if the acceptance rates are too low then the system will not be exploring configurational space quickly enough, whereas too high an acceptance rate indicates that the attempted moves are too conservative (step size is too small). In general acceptance rates of 0.5 are targeted, but, there are several moves where this is not achievable for our systems. For instance, in order to be ergodic the crankshaft move is required, but the acceptance rate of this move is roughly 0.05. By restricting the crankshaft move to sections of polymer containing four beads — instead of picking any two beads on a polymer — the acceptance rate increases but still remains close to the 0.1 mark. Additionally, the pivot move acceptance rate is often far less than 0.5 (about 0.01), but due to the potential of this move to significantly change the polymer's path, it has been claimed by Madras and Sokal to be one of the most efficient moves for polymers [156].

As we are interested in the behaviour of equilibrated systems, statistics are only gathered in the production stage of the simulation. The length of the equilibration, prior to the production stage, is determined by monitoring the system energy, spatial exploration of individual polymers and the cluster size, where relevant. After this equilibration stage, the outputs from the simulation are used to produce the system statistics. This means that, for any given random number seed, the overall statistical behaviour of the system will be the same. Monitoring convergence involves looking at the system energy, maximum cluster size, cluster size distribution and spatial exploration of the system by individual polymers, as well as using different random seeds, to check the results are consistent. The number of MC sweeps in the equilibration stage and the production runs are presented in the description of each system studied, in Chapters 3, 4 and 5.

Chapter 3

Phase separation

3.1 Introduction

In this chapter, we investigate the influence of intrinsic polymer properties on the phase separation of associative polymers, motivated by open questions related to the formation of membraneless organelles inside biological cells. The properties we investigate are linker length, linker type, polymer valency and interaction strength between binding beads. Comparisons between the different means of modelling polymers, such as self-avoiding walks or random walks, are undertaken to determine the suitability of each for studying associative polymer behaviour.

3.1.1 Liquid–liquid phase separation (LLPS)

LLPS has been widely reported as a potential first step in the formation of membraneless organelles. Therefore, by probing LLPS, a deeper understanding of the initial stages of their formation can be obtained. The complexity and myriad of components involved in cellular systems obscures such conclusions in studies of real cells. We therefore utilise 'toy model' systems to study LLPS because this allows for more generic, universal findings, which can then be applied to specific biological systems.

A range of coarse-grained models previously used to study LLPS were introduced in Chapter 1, but the key ingredients of these models are also relevant for non-living polymer systems [5, 140, 159–164]. For example, the work of Semenov and Rubinstein shows that the phase separation of associating polymers can be disrupted by excluded volume (to form a gel and not a globule), and without excluded volume

phase separation is guaranteed for sufficiently strong interaction strengths [159]. Moreover, works involving oppositely charged species, known as complex coacervation give insight into the role of interaction types, such as electrostatic and van der Waals, on phase separation [165].

The aims of this chapter are to model multivalent polymers, thereby uncovering key parameters which drive phase separation, such as polymer valency, linker length and linker type. High polymer valency has been widely shown to promote phase separation, but the latter two variables have been studied less. We are interested to see whether changing the polymer linker type affects the phase separation, and crucially, how linker type and length alter the agreement between our computational model and the experimental results presented in the seminal paper by Li *et al.* [7]. We then turn our attention to studying the properties of the structures formed at a level of detail inaccessible in experiments. This includes investigating individual polymer geometries within the phase-separated bodies, alongside the interconnectivities and densities of these structures. An overarching purpose of this work is to compare the different linkers available when simulating polymers on a cubic lattice, and comment on their suitability for modelling biopolymer phase separation. Another fundamental goal of this work is to distinguish between the model-specific and universal behaviour of multivalent associative polymers. These findings will be be informative for future research of polymers on a cubic lattice, and will also be utilised in Chapters 4 and 5 of this thesis.

3.1.2 System set-up

Inspired by the model *in vitro* system studied by Li *et al.* [7], we envisage a system in which polymers of two different species, both with well defined valencies, can be simulated and the phase behaviour monitored. This allows for comparisons between our model and the *in vitro* Proline Rich Motif (PRM) — Simple Homology 3 (SH3) system central to Ref. [7]. However, our findings are applicable to other associative polymers and are not specific to the system used by Li *et al.* In this chapter we will use the labels SH3 and PRM interchangeably with A and B, respectively, to denote the binding beads, reflecting the generality of the approach. The polymers are represented as beads on a string, where the beads are constrained to lattice sites. The binding sites of the two polymer species interact with strength between

beads of type A and type B given by ϵ_{AB}, where $\epsilon_{\mathrm{AB}} > 0$ is attractive. The binding model used in this system is the vector binding model discussed in Chapter 2. In our simulations the interactions are those of excluded volume and an attractive bonding interaction between binding beads of polymer A and polymer B. There are no A-A or B-B interactions. Hence, the total energy of the system in a configuration with m bonds, is $-m\epsilon_{\mathrm{AB}}$. A two-dimensional schematic of the cubic lattice system is shown in Figure 2.1.

Initialising the simulations involves filling the simulation box with the requisite densities of SH3 and PRM binding domains. The packing fractions of SH3 and PRM beads ϕ_{packing}, defined here as the fraction of lattice sites occupied by a given species, are converted to concentrations in units of μM. This is to allow for more direct comparison with existing experimental results. The SH3 binding domain has an approximate volume of $v = 1.82 \times 10^{-26}\,\mathrm{m}^3$ [166], which can then be used to find the concentration [SH3] using [SH3]$= \phi_{\mathrm{packing}}/vN_{\mathrm{A}}$, where N_{A} is Avogadro's number. The PRM binding domain has a similar volume to SH3 [68]; therefore [PRM] is found using the same value for v. For future reference, 100μM is equivalent to a packing fraction of 0.0011. The equilibration was typically 5×10^8 MC steps or 10^5 simulation sweeps, where one sweep is defined as the number of MC steps required to attempt one move per bead in the system on average. The production run involved around 10^9 steps, or 10^6 sweeps of the system.

3.2 Modelling LLPS of associative polymers

In this section phase diagrams will be presented, where the colour at each point on the phase diagram indicates the percentage ϕ of protein domains located in the largest cluster, for SH3 and PRM binding site concentrations ranging from 100–400 μM (packing fraction of 0.0011–0.0044). This presentation of the results allows for rapid identification of the ϕ values, from which the degree of phase separation of a system can be identified. However, we are also interested in the structures of the clusters formed. We therefore also present the average packing fraction of SH3 and PRM beads, as a function of the distance from the cluster centre of mass, alongside our phase results.

Studying the structures of the clusters formed involves monitoring the connectivity and occupancy of the proteins in the system. Occupation, defined as the average

number of beads bound on any given polymer, allows us to probe how the number of bonds formed by the proteins is affected by the degree of phase separation. Connectivity, defined as the average number of polymers that any one polymer is bound to, should reveal the relationship between linker length or valency and the cross-linking of the clusters.

Here, unless stated otherwise, the interaction strength between the beads is chosen to be $7.95kT$. This was determined from the NMR titration value of the SH3–PRM interaction strength [7, 167].

3.2.1 Hard linkers

The first results we present are for proteins with hard explicit linkers, whereby binding beads are connected by a contiguous chain of linker beads, all with an excluded volume of one lattice site. The phase diagrams for this system, with interaction strength of $7.95kT$, are shown in Figure 3.1. There is almost no phase separation observed within this system. Instead, polymers pair up by forming multiple bonds with a single complementary polymer. This is enthalpically favourable, whilst maintaining some of the polymers' translational entropy, which would be lost if the polymers aggregated into a large cluster. The lack of interconnectivity of chains suppresses phase separation for chains of this linker type.

Since phase separation is observed in the experimental work of Li *et al.* , one can conclude that hard linkers are not a suitable model. Nevertheless, for the general case of associating complementary polymers, it is interesting to study the properties of the structures formed when polymers with hard linkers do form a phase-separated body. In order to do this, the interaction strength was increased to drive phase separation. Figure 3.2 indicates that increasing the interaction strength results in phase separation for systems with hard linkers, and that the propensity to phase separate increases with both valency and linker length. In fact, this is a dominant trend that we observe throughout this work.

The increase in phase separation with valency is due to the total interaction strength for each chain. The increase with linker length may result from longer linkers allowing the beads to explore the system more freely, thus making bond formation more likely. The greatest phase separation occurs along the diagonal of the phase diagram, $y = x$, which is promoted by the stoichiometry of the beads

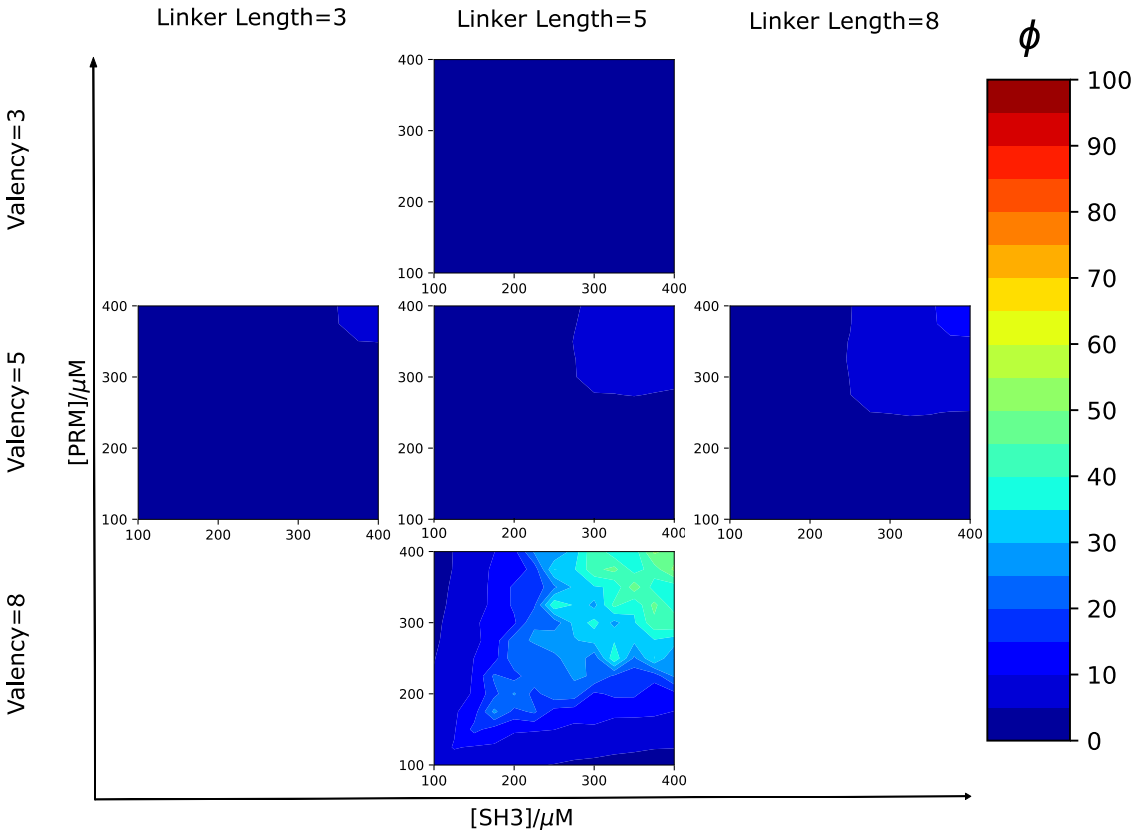


Figure 3.1: Phase diagrams of polymers SH3 and PRM indicating the percentage of polymers in the system that are located in the largest cluster of the system, ϕ. Here the polymers have hard linkers, and the interaction strength between SH3 and PRM beads is $7.95kT$. The colour bar, ϕ, indicates the percentage of the total chains (of both SH3 and PRM) that are located in the largest cluster. Five separate phase diagrams are shown, for chain linker lengths of 3, 5 and 8 and chain valencies of 3, 5 and 8. The phase diagrams illustrate the impact that changing the concentration of SH3 or PRM binding sites has on the cluster size.

of each type being the same. Going off the diagonal puts one species of chain into excess which causes the overall phase separation to be reduced.

One of the crucial findings of this investigation is that the density of binding beads in clusters is suppressed by the linker beads. This is illustrated in the density plots for various polymer concentrations along the $y = x$ diagonal, as shown in Figure 3.3. The packing fraction of receptor beads plateaus at a value of $0 \leq d \leq 0.01$. The overall binding bead density in the system (density of SH3 and PRM beads) is 0.0022 for 100μM and 0.0088 for 400μM, these correspond well with the values at which the densities level off. The levelling off of the density curves at $d \leq 0.01$, where large clusters form, indicates that the polymers connect to form an expansive network of approximately uniform density. This is in agreement with the work of the Pappu group, who found that hard linkers result in low-density, system-spanning networks

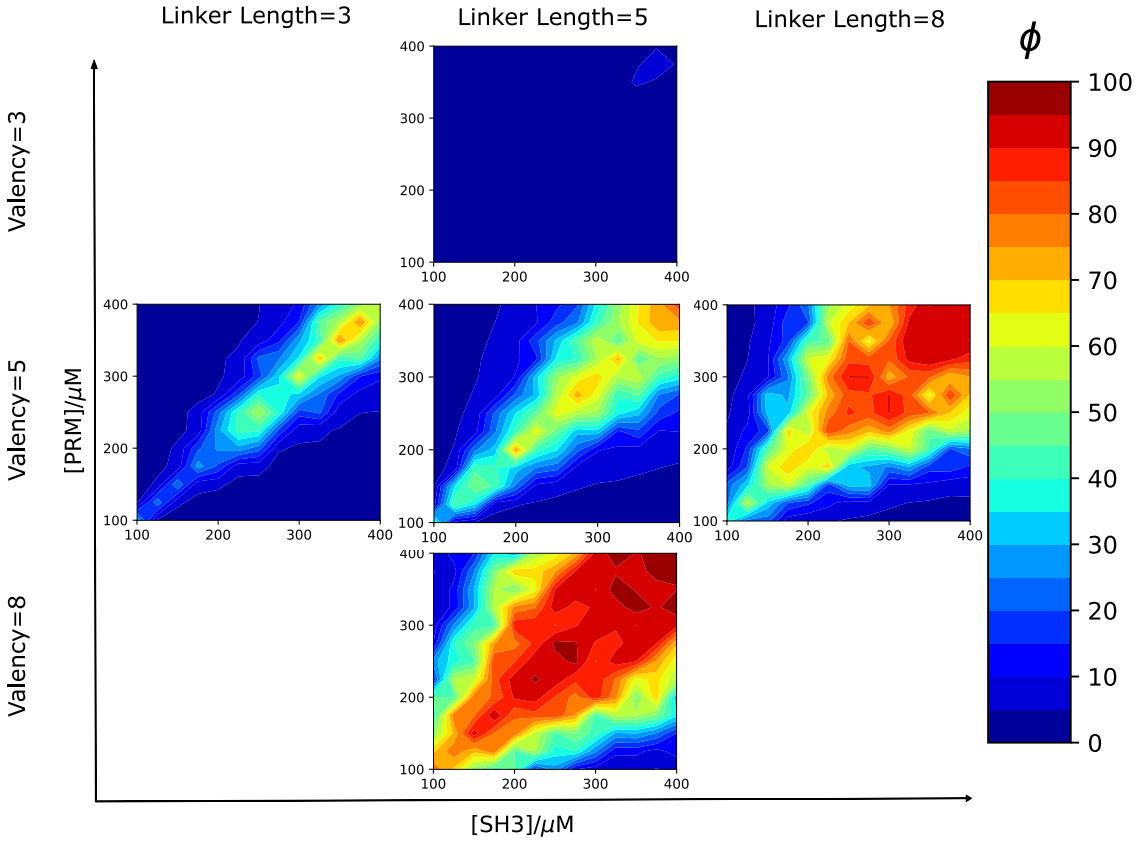


Figure 3.2: Phase diagrams of polymers of SH3 and PRM indicating the percentage of polymers in the system that are located in the largest cluster of the system, ϕ. The polymers have hard linkers, and interaction strength $\epsilon/kT = 12$.

forming [168]. This means that phase separation may be a misleading description, as the concentrations and densities do not change significantly; instead the chains cross-link to form a sparse network. Because of this, Harmon *et al.* described the process as gelation without phase separation [68].

Low-density clusters are much more likely to form as the interaction strength increases. As the concentration of binding beads increases, the average cluster size increases, but the density and structure of the clusters are similar regardless of the binding bead concentration. This allows the properties of the largest cluster in each simulation along the [SH3] = [PRM] diagonal of any given phase diagram, to be averaged together. The occupancy and connectivity of the polymers in the largest clusters for these simulations are shown in Table 3.1. They show that the occupancy increases with increasing valency but that the connectivity is largely insensitive to valency and linker length. Although occupancy increases with increasing valency, the fractional occupancy actually decreases. However, the more pertinent finding is that connectivity is approximately unchanged by the valency or linker length of the polymers. The entanglement and excluded volume of the polymers are likely to be

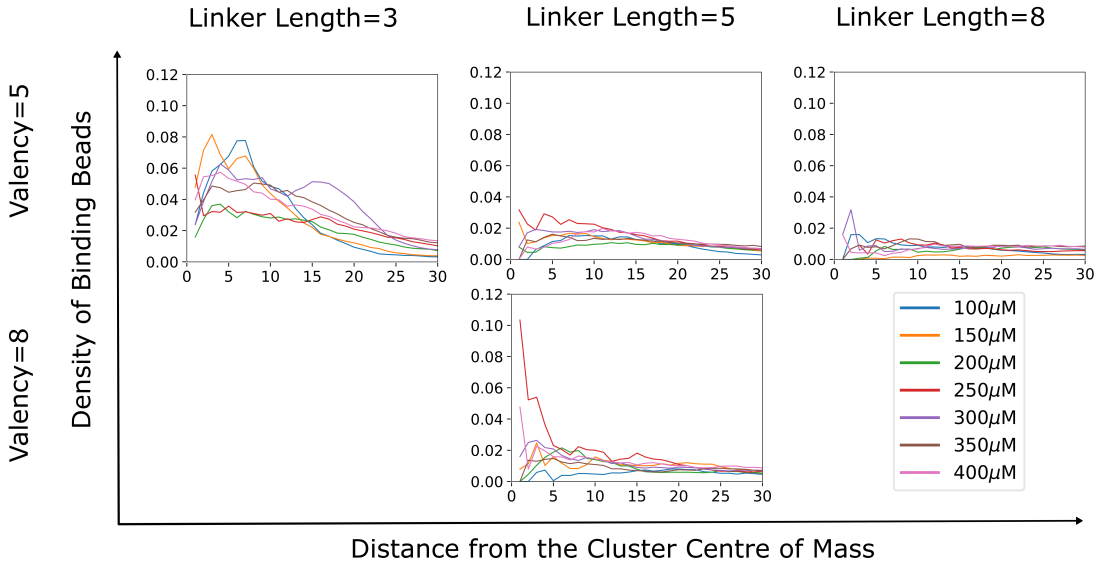


Figure 3.3: Densities, as packing fractions, of the largest cluster at selected polymer concentrations, with respect to the distance from the cluster centre of mass.

disrupting the ability of the polymers to cross-link with multiple different polymers.

Following on from the study of occupancy and connectivity, we turn to the radius of gyration R_g of the polymers, given by Equation 2.3. The results in Table 3.2 indicate that the radius of gyration increases with increasing linker length and valency. Combining the results from Tables 3.1 and 3.2, we therefore find the radius of gyration of the polymers increases without an increase in polymer interconnectivity.

3.2.2 Implicit linkers

Next we turn to a system where the polymers consist of binding beads connected by implicit linkers. These simulations remove the obstacles present for hard linkers,

Table 3.1: Occupancy and connectivity of polymers with hard linkers, where these values are averaged over 7 equally spaced points on the phase diagram where the concentrations of beads SH3 and PRM are equal. The interaction strength between SH3 and PRM beads is $12kT$.

Valency	Linker Length	Occupancy	Coordination
3	5	2.836(5)	2.041(7)
5	3	4.747(9)	2.35(1)
5	5	4.647(9)	2.33(1)
5	8	4.57(1)	2.43(2)
8	5	7.15(2)	2.56(1)

in terms of entanglement and computational expense. More importantly, they are likely to be more representative of the proteins involved in biopolymer LLPS as their solvation volumes match most closely to these proteins [68].

The results in Figure 3.4 show more significant phase separation for proteins with implicit linkers compared to those with hard linkers. As for the hard linkers, here the phase separation also increases with increasing protein valency and linker length. The reasons for these relationships between valency, linker length and phase separation are the same as those given for the hard linkers.

The simulation phase diagrams are of significantly higher resolution than those of Li *et al.*, partly due to the relative ease of running multiple simulations compared to experiments, and also due to the level of detail accessible in simulations. Despite the difference in resolution, we can see general agreement between Figure 1.8 (experimental data of Li *et al.*) and Figure 3.4, where clearly there is no phase separation for polymers of valency 3 but extensive phase separation for polymers of valency 5, with sufficiently long linkers. Calculating the precise linker length of the real polymers used in the experimental work was not possible. Therefore we used the finding of Holehouse *et al.* from their atomistic simulations that the linkers are approximately 7 times the size of the binding domains, leading us to choose a linker length of 8 lattice edges [68]. Given the phase behaviour observed, these implicit linkers show suitability for modelling of biopolymer phase separation as observed in the experiments of *Li et al.*

On probing the clusters that formed, we find that the density of the clusters is only weakly sensitive to linker length, as shown in Figure 3.5. Therefore, linker length has a pivotal role in determining the degree to which phase separation occurs,

Table 3.2: Radius of gyration of polymers located within the cluster, in the case of hard linkers. These results are averaged over points on the phase diagram where concentrations of beads SH3 and PRM are equal. The interaction strength between SH3 and PRM beads is $12kT$.

Valency	Linker Length	Radius of Gyration
3	5	2.57(1)
5	3	2.61(1)
5	5	3.93(2)
5	8	5.40(1)
8	5	5.44(4)

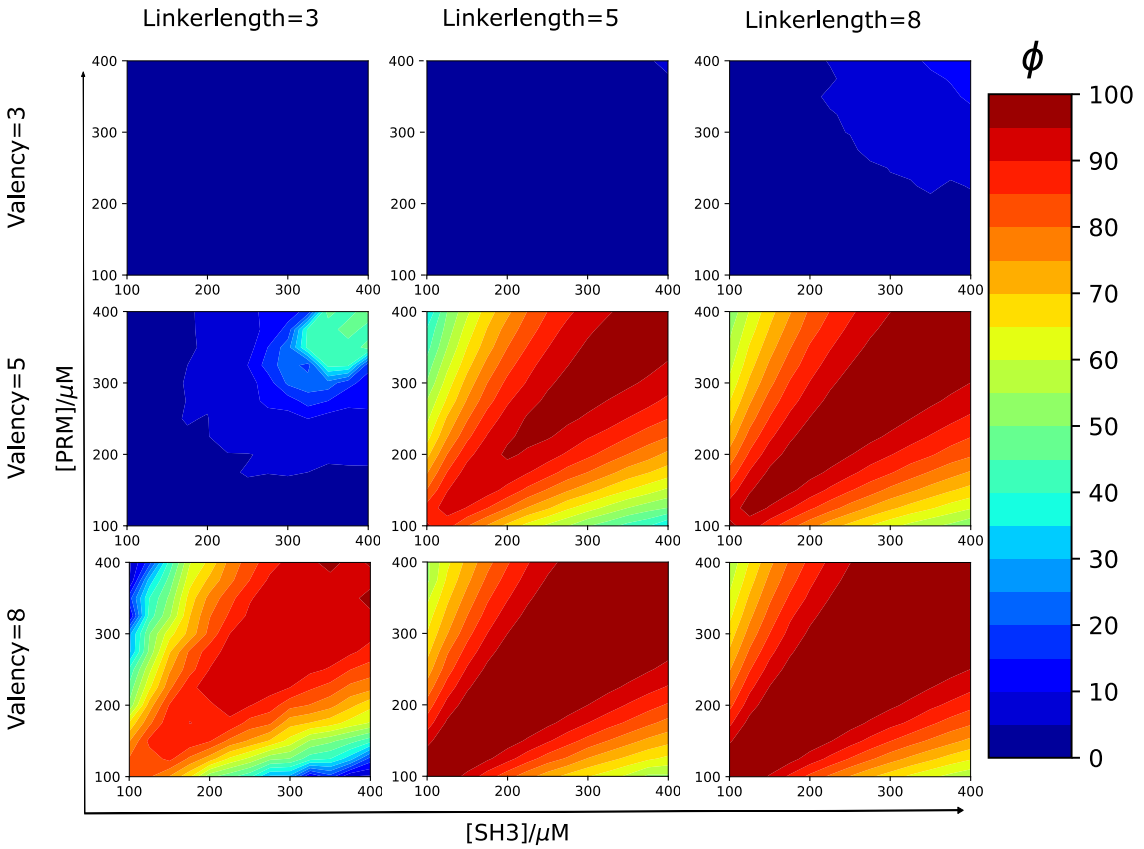


Figure 3.4: Phase diagrams of polymers indicating the percentage of polymers in the system which are located in the largest cluster of the system, ϕ. The polymers have implicit linkers and the interaction strength between SH3 and PRM beads is $7.95kT$

but the structures of the clusters formed are largely independent of this parameter.

Increasing valency enhances phase separation and has a more notable effect on the cluster density than linker length, due to each chain being able to form more bonds and be more interconnected. However, the lack of major density change suggests that there are other significant factors at play in addition to the energy of the system and the entropy of the linkers. A potential factor in the high densities of clusters is the entropic increase in the system due to beads having multiple binding beads adjacent to them, resulting in bond degeneracy. The density insensitivity suggests that bonding entropy S_{bonding} is more significant than linker entropy, and TS_{bonding} is of comparable size to binding energy at the temperature studied.

The densities are high, with the most dense clusters having a packing fraction of roughly $d = 0.55$. This is an order of magnitude greater than those of the hard linkers. Although these clusters appear to be very dense for biological droplets, this is in agreement with experimental observations of membraneless organelles [17, 169] and simulation results from a similar system [168]. It is also worth noting that the

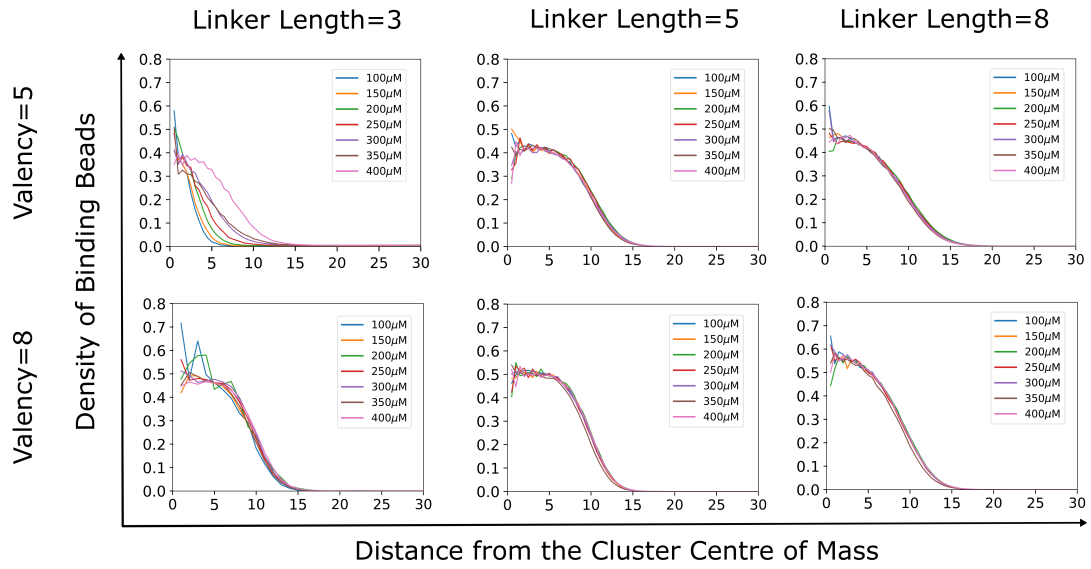


Figure 3.5: Density of beads within the largest cluster in the system, at a variety of polymer concentrations, with respect to the distance from the centre of mass of the cluster for polymers with implicit linkers and the interaction strength is $7.95kT$.

linker properties have previously been shown to have important influence over the phase separation and structures formed [1, 68, 69].

The bound states of the polymers in the largest cluster are characterised in Table 3.3. Linker length has limited impact on the number of beads that are bound on any one polymer. However, the connectivity (average number of distinct chains each chain is bound to) increases with linker length. This result indicates that having a longer linker length allows the chains to become more interconnected, without increasing the number of bonds each chain makes. The reason for this, as we shall shortly illustrate, is the entropic costs of forming highly interconnected structures are smaller for implicit linkers than for hard linkers. It is this interconnectivity that drives the formation of a single large cluster of protein, i.e. phase separation.

By measuring the radii of gyration of the chains in the cluster and comparing them to free chains, as given in Table 3.4, it can be seen that the chains stretch out in the cluster compared to when they are free. Although the standard deviation of the radii of gyration for the free chains is large, we can conclude that the chains do not become more compact in the gel. This shows that the high density of the clusters is not due to individual chains becoming more compact. Instead, the drive for chains to maximise their bonding entropy is facilitated by stretching the chains out so that they can connect to as many different chains as possible. This forms a highly interconnected cluster that allows for each bead to have options of forming

a number of different bonds. This stretching also explains why phase separation and connectivity are enhanced with increasing linker length, as by increasing the linker length, the beads are able to explore more space and form interactions with more distant chains. Combining this effect with the observation that connectivity improves with linker length and that density increases with linker length, it is apparent that the high density is achieved by the chains stretching out and forming a mesh like structure, where the linkers overlap significantly.

Table 3.4 also includes data for the direct linker lengths – the average separation in space of two adjacent binding beads. This is of limited use in isolation, but allows for more complete comparison of radii of gyration of different linker types, to be discussed in Sections 3.2.3 and 3.2.4.

3.2.3 Implicit linkers with entropic spring

Having postulated that the significant phase separation observed for the implicit linker is in part enabled by the lack of entropic penalty for stretching out the polymer chains, we now test the importance of this omission by adding an effective intra-chain potential $F_{\text{path}}(r)$, between adjacent polymer beads, to account for the change in the path entropy which is emitted by the implicit representation. For a random walk (chain) of N segments, where each segment has a length a equal to one lattice spacing, the probability of the chain having a given end-to-end separation $P(r)$ is

$$P(r) = \left(\frac{3}{2Na^2}\right)^{-3/2} \exp\left[-\frac{3r^2}{2Na^2}\right]. \tag{3.1}$$

Table 3.3: Occupancy and connectivity of polymers with implicit linkers, averaged over 7 equally spaced points on the phase diagram where the concentrations of beads SH3 and PRM are equal. The interaction strength between SH3 and PRM beads is $7.95kT$.

Valency	Linker Length	Occupancy	Coordination
5	3	4.25(2)	2.83(8)
5	5	4.331(1)	3.930(3)
5	8	4.229(2)	4.083(3)
8	3	7.101(1)	4.668(6)
8	5	7.039(1)	5.971(4)
8	8	6.897(3)	6.470(4)

Table 3.4: Radius of gyration of polymers and the direct linker length of polymers with implicit linkers found within the largest cluster of the system, alongside the radii of gyration of equivalent polymers in isolation. These results are averaged over points on the phase diagram where concentrations of beads SH3 and PRM are equal.

Valency	Linker Length	Radius of Gyration	Direct Linker Length	Radius of Gyration (Free)
5	3	1.68(7)	1.487(1)	1.6(3)
5	5	2.63(2)	1.869(1)	2.3(5)
5	8	4.06(3)	2.3431(3)	3.0(7)
8	3	2.15(2)	1.4958(2)	2.0(3)
8	5	3.22(4)	1.8744(1)	2.9(6)
8	8	4.86(5)	2.3480(5)	3.8(9)

This equation arised from the distribution of the number of different paths to achieve a bead separation of r for a linker of length Na in the limit of long chains [170].

The entropy of a chain is given by Boltzmann's equation,

$$S(r) = k \ln \Omega, \tag{3.2}$$

and this can be multiplied by $-T$ to give the associated free energy contribution $F(r) = -kT \ln \Omega$. Here, Ω is the number of unique ways to connect the two chain ends separated by r, and it is proportional to the probability distribution $P(r)$ given in Equation 3.1. This can then be used to calculate the effective potential required to account for the entropic contribution of the linker,

$$F_{\text{path}}(r) = \frac{(3r^2 kT)}{(2Na^2)} + C \tag{3.3}$$

where C is a system-dependent constant, which changes with T, N and a but is constant in any given simulation and can therefore be set to zero. This potential is then added to the energy change of the trial move, before the move is accepted or rejected using Equation 2.4.

The resulting phase diagrams are depicted in Figure 3.6, where there is a significant suppression of phase separation, compared to the purely implicit linker results in Figure 3.4, upon adding the entropic spring term. As with the implicit linkers with no entropic spring term, the concentration threshold at which phase separation occurs decreases with increasing valency and linker length.

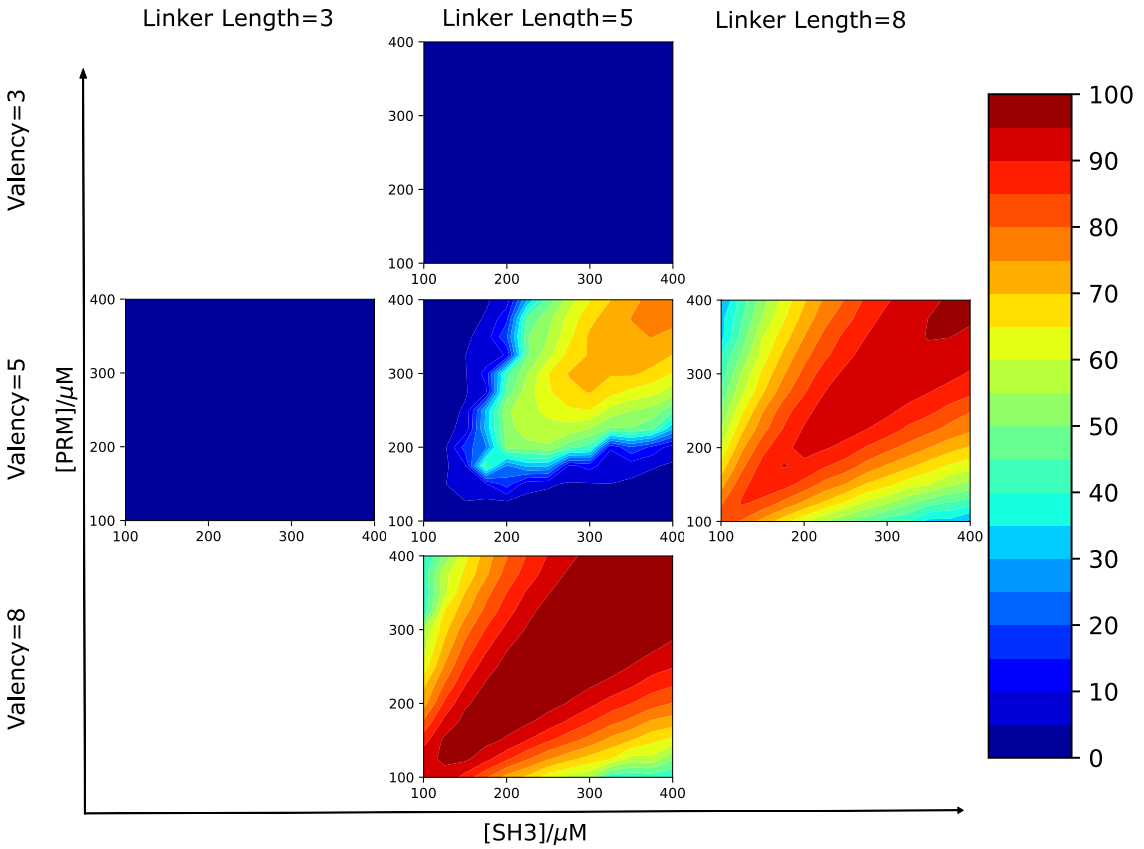


Figure 3.6: Percentage of chains ϕ in the largest cluster for the polymers with implicit linkers and an entropic spring correction. The interaction strength between SH3 and PRM beads is $7.95kT$.

The densities of the clusters sit in the region of $0.4 \leq d \leq 0.5$, with the exception of the clusters of chains with linker length of 3. This indicates that although the entropic spring reduces phase separation, it does little to alter the actual structure of the clusters formed. The clusters formed for chains with linker length of 3 are very small and dense, and are not of particular interest as they do not represent the global phase separation of LLPS. Looking more closely at the structures of the clusters, presented in Table 3.5, for systems where phase separation occurs, the occupation and coordination are very similar to the clusters formed from chains with implicit linkers and no entropic spring.

The final comparison between the implicit linker and the implicit linker with an entropic spring term involves Tables 3.4 and 3.6. The entropic spring reduces the radius of gyration of the chains, which is verified by the documented reduction in the direct linker length – the average separation in space of two adjacent binding beads. Both the radii of gyration and the direct linker lengths indicate a slight contraction of the linkers when the entropic term is introduced, with the most pronounced effect for long linkers.

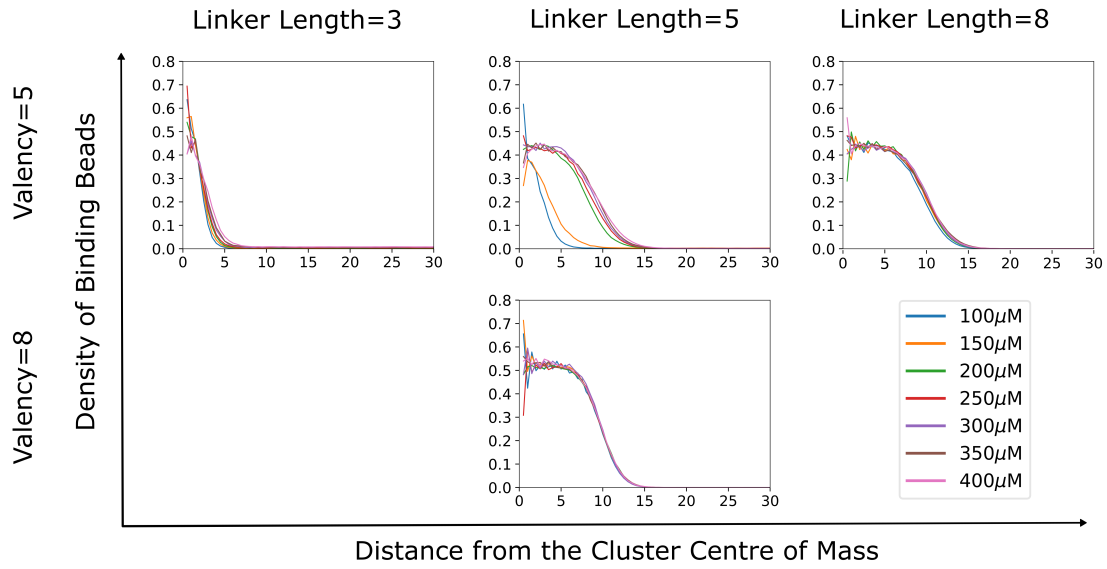


Figure 3.7: Density of binding beads in the largest cluster as a function of distance from the cluster centre of mass, for proteins with implicit linkers and an entropic spring correction.

3.2.4 Ideal linkers

The final linker type studied is ideal linkers. These consist of binding beads connected by a contiguous chain of linker beads, but where the linker beads have no excluded volume. This representation of linkers is appealing as it allows for entropic contribution of the linkers to be accounted for, but without causing the polymers to expand too much due to excluded volume effects. They also allow for simulation of polymers amongst obstacles where the linkers can overlap with each other but not with the obstacles, whereas implicit linkers can pass through obstacles in an unphysical way because their path through space is not considered.

On moving from implicit linkers to ideal linkers, phase separation becomes less likely, as shown in Figure 3.8. This shows that the linkers resist the formation of

Table 3.5: Occupancy and connectivity of polymers with implicit linkers and an entropic spring correction. These values are averaged over 7 equally spaced points on the phase diagram where the concentrations of beads SH3 and PRM are equal. The interaction strength between SH3 and PRM beads is $7.95kT$.

Valency	Linker Length	Occupancy	Coordination
5	3	4.21(1)	2.53(5)
5	5	4.345(1)	3.879(2)
5	8	4.258(2)	4.089(2)
8	5	7.057(3)	5.844(7)

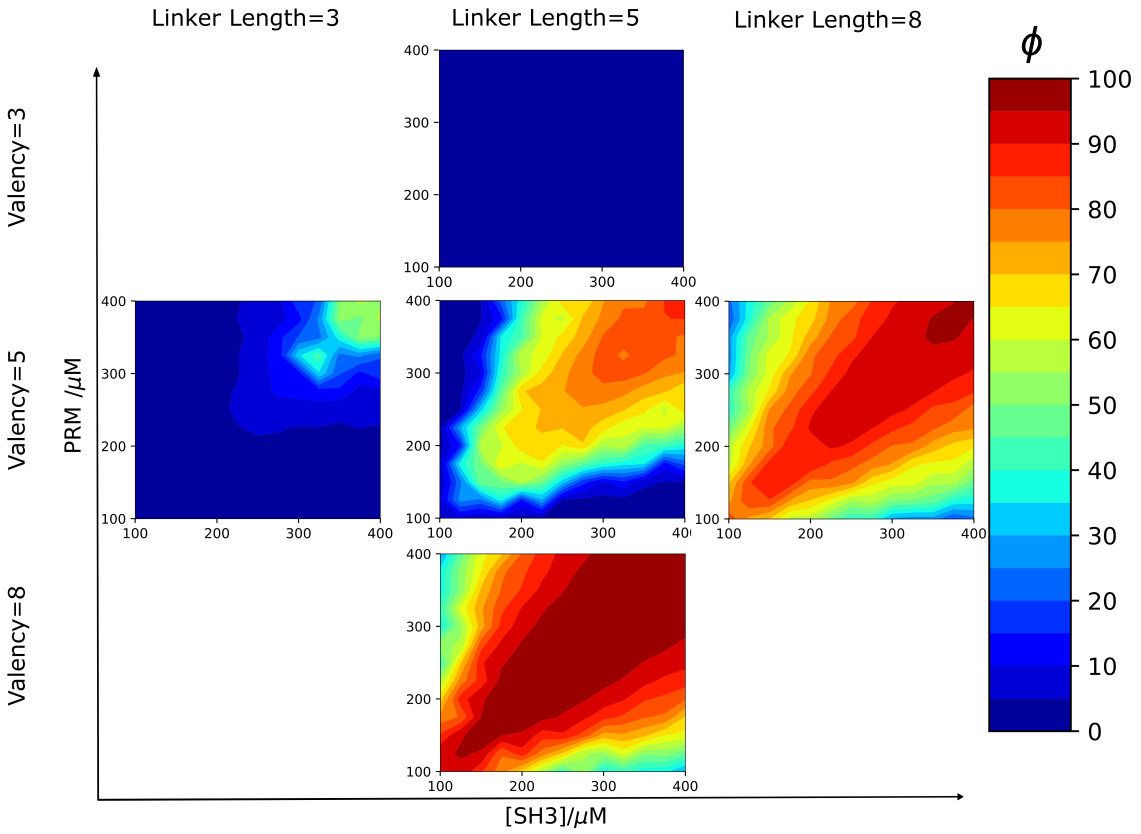


Figure 3.8: Phase diagrams of polymers of SH3 and PRM indicating the percentage of polymers in the system which are located in the largest cluster of the system, ϕ. The polymers have ideal linkers and the interaction strength between SH3 and PRM beads is $7.95kT$.

dense, interconnected structures. Notably, the phase separation for ideal linkers is in good agreement with the implicit linkers adjusted with an entropic spring term (Figure 3.6) revealing that the configurational entropy of the linkers can affect physical properties even if the linkers occupy negligible volume.

The densities of the largest clusters along the diagonal of the phase diagram are largely insensitive to the linker length, as shown in Figure 3.9. The densities sit around $0.4 \leq d \leq 0.55$, which is similar to that of the implicit linkers. Furthermore,

Table 3.6: Radius of gyration and direct linker lengths of polymers within the largest cluster. These results are averaged over points on the phase diagram where concentrations of beads SH3 and PRM are equal.

Valency	Linker length	Radius of Gyration	Direct Linker Length
5	3	1.578(4)	1.4012(7)
5	5	2.4730(6)	1.7592(2)
5	8	3.817(1)	2.1980(4)
8	5	3.0531(8)	1.7636(2)

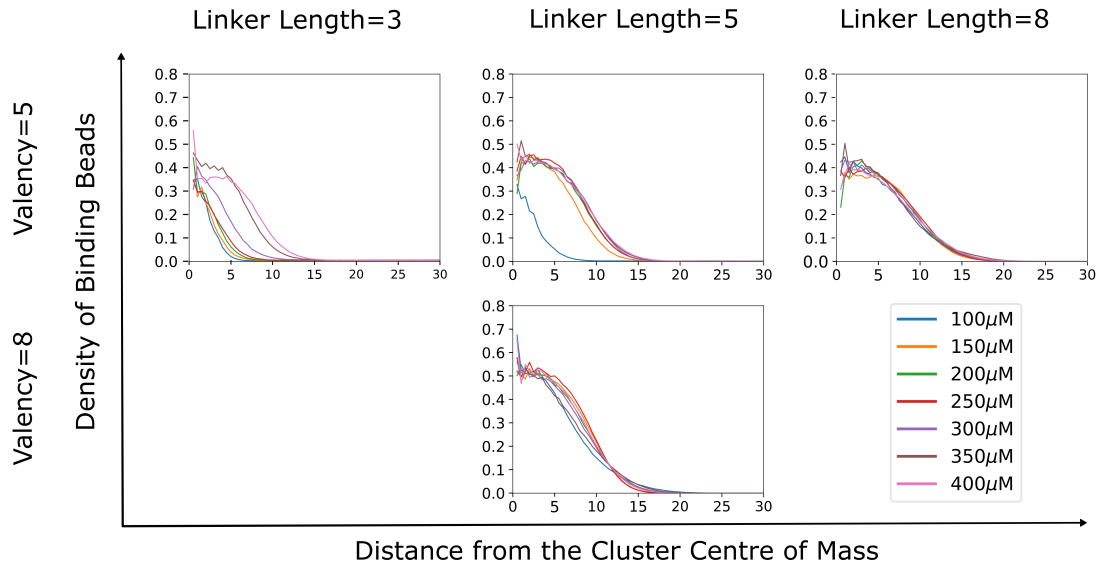


Figure 3.9: Density of binding beads with respect to the distance from the cluster centre of mass for polymers with ideal linkers.

they do not depend heavily on bead concentration, as shown by the strong overlap between the density curves within each plot.

As with the implicit linkers, the connectivity of the largest clusters increases with linker length and with increasing valency, as shown in Table 3.7. Additionally, the occupancy of the polymers is consistent with that of the polymers with implicit linkers. This connectivity is, however, lower than that of the implicitly linked polymer chains, including those with implicit linkers and an entropic spring. These observations may be a result of the ideal linkers occupying less stretched configurations, which are entropically costly, thereby reducing extensive inter-polymer interactions, and favouring the formation of multiple bonds with the same chain, as opposed to forming cross-links with many polymers. Polymers with ideal linkers have lower coordination than those with implicit linkers and an entropic spring and also have lower direct linker length. The lower coordination number is attributable to the slight difference in binding entropy between the ideal and implicit linkers; implicit linkers can form a bond with any adjacent bead on the neighbouring 6 lattice sites, whereas it would be unphysical for bond vectors to lie on top of the backbone of the chain, so such configurations are excluded when the linkers are represented explicitly. Therefore, the binding beads are limited in the number of directions they can bond. The end beads can only form bonds to neighbouring beads on 5 of the neighbouring lattice sites, and mid-chain beads can only have 4 directions available to its bond vector, and thus binding direction.

We now compare polymers with ideal linkers to those with both hard and implicit linkers. As the binding models of the hard and implicit linkers differ, due to the excluded volume of linker beads directly connected to binding beads, the bonding model of polymers with ideal linkers cannot simultaneously match both of these models. The slight difference in binding model between the implicit and ideal linkers should have a very limited impact on the overall phase behaviour of the system but does explain the slight discrepancy between ideal linkers and implicit linkers with an entropic spring. The shorter direct linker length will be discussed more fully in Section 3.3, where the scaling properties of the different linkers are compared. Briefly, the shorter direct linker length is unlikely to be due to lattice restrictions, and instead result from the reduced entropy of binding being compensated for by more compact (and therefore more entropically favourable) linkers.

The radius of gyration data in Table 3.8 show that the chains are less compact than those with implicit linkers for linkers of length $l \leq 5$, but then become slightly more compact at longer linker lengths. The greater radius of gyration at lower linker lengths is due to geometric restraints of having to connect the two beads by a random walk along lattice edges. The behaviour of the polymer at linker lengths of 5 and above is better illustrated by the direct linker lengths. These reveal the polymer linkers to be very compact, despite more modest reductions in the radius of gyration. This further supports our hypothesis that the entropic penalty for stretching out explicitly modelled linkers reduces phase separation by reducing the interconnection of the polymers. This clearly illustrates the entropic battle within these systems of maximising linker entropy by compacting the linkers, whilst maximising the binding entropy by lengthening the chains.

Table 3.7: Occupancy and connectivity of polymers with ideal linkers. These values are averaged over 7 equally spaced points on the phase diagram where the concentrations of beads SH3 and PRM are equal. The interaction strength between SH3 and PRM beads is $7.95kT$.

Valency	Linker Length	Occupancy	Coordination
3	5	2.394(9)	1.95(2)
5	3	4.26(3)	2.76(9)
5	5	4.34(2)	3.44(7)
5	8	4.527(2)	3.951(5)
8	5	7.020(3)	4.81(2)

Ideal linkers therefore facilitate study of polymer LLPS, with trends consistent with the work of Li *et al.* For example, we see increasing phase separation with increasing linker length and valency, and show that dense structures form in such systems. The results are also instrumental in determining the success of the entropic spring added to the implicit linkers. They will be used in future chapters where physical obstacles, which implicit linkers cannot account for, are present.

3.2.5 Stoichiometry and cluster properties

At this point, we introduce a schematic illustration of the nature of the structures formed at various regions of the phase diagram. Figure 3.10 is representative of the clusters formed for polymers of valency 5, linker length 5 and with implicit linkers. Similar behaviours are observed for other valencies, linker lengths and linker types. Up to this point we have focused largely on the cluster properties along the diagonal of the phase diagram, as this is the region where phase separation is most likely, and thus allows us greatest insight into the impact of polymer properties on the clusters formed. However, qualitative information can also be obtained by deviating from this diagonal.

On moving off the diagonal, one species becomes in excess with respect to the other. This immediately reduces the maximum cluster size (population) within the system, but the rest of the system also changes. When one species is in excess, instead of the free polymers in the rest of the system pairing up to maximise the enthalpy, very small aggregates form, with their stoichiometry dependent on the relative polymer concentrations. The pairing-up process of species not in large clusters is energetically favourable, as it maximises the number of bonds of the two chains. The decrease in cluster size continues with the increasing degree of

Table 3.8: Radius of gyration and direct linker length of polymers with ideal linkers, found in the largest cluster of the simulation. These results averaged over points on the phase diagram where concentrations of beads SH3 and PRM are equal.

Valency	Linker length	Radius of Gyration	Direct Linker Length
5	3	2.084(6)	1.261(2)
5	5	2.82(1)	1.423(2)
5	8	3.605(3)	1.5946(2)
8	5	3.711(7)	1.4264(4)

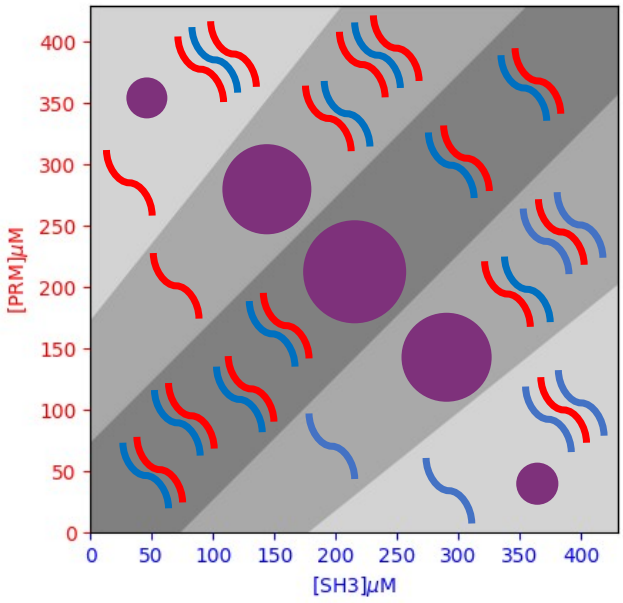


Figure 3.10: Schematic of the clustering behaviour corresponding to the phase diagrams for implicitly linked polymers of SH3 and PRM.

imbalance between the concentrations of the polymer species. We define the bulk as the system surrounding the large cluster(s) in the system. When the disparity in concentrations is sufficiently large, the bulk consists exclusively of single chains and triplets (of two of the excess polymers and one of the limiting polymers) as shown in the top left and bottom right corners of Figure 3.10. This is a result of balancing the minimisation of the loss in translational entropy of the chains with the maximisation of the entropy associated with the number of binding combinations. The nature of the bulk is important, as the relationship between the cluster and bulk helps identify the stability and behaviour of the phase-separated bodies.

This level of detail is inaccessible to experiments, but our findings go some way to explaining why, by having a deficit of one of the species, much smaller droplets formed [32]. This accounts for the lack of phase separation observed in experiments for low-valency polymers, at low concentrations of SH3 or PRM, shown in Figure 1.8.

3.3 Comparison of linker types for a single chain

The different linker types employed within this work are representative of a range of linker structures within proteins. To qualitatively summarise the three cases, ideal linkers have neither attraction nor excluded volume and are therefore neither

expanded nor collapsed. Hard linkers occupy an expanded configuration which is representative of a polymer in a good solvent [171]. As we shall see, the statistics of individual implicit linkers contrast with those of both ideal and hard linkers.

In order to uncover some of the reasons for the differences between the phase behaviour of the different linkers, a study of single chains, with each linker type, is useful. The probability distributions of the end-to-end separation from simulations of individual chains with a binding bead at each end, are shown in Figure 3.11. We also present the expected probability distribution for a freely jointed chain (ideal linkers) for comparison, using Equation 3.3, which is accurate in the limit of long linkers.

The end-to-end distributions for the hard and implicit linkers are shown for comparison, but are not expected to agree with the distribution for freely jointed chains due to excluded volume effects and the absence of path entropy, respectively. The remaining comparisons will be between the theoretical end-to-end distribution and the distributions for implicit linkers both with and without an entropic spring term and also the ideal linkers.

The linker lengths used in the phase separation investigation are quite short and therefore the end-to-end distributions are impacted strongly by the cubic lattice. This is shown by the limited agreement between the theoretical end-to-end distributions and the observed distributions for short linkers in Figures 3.11A–C. The reasons for these discrepancies are both the restrictions that a lattice applies to bead distribution (the number of lattice sites a bead can occupy within a radius r of a given position does not increase smoothly as $\frac{4}{3}\pi r^3$, unlike for an off-lattice system), and because the theoretical distribution is only accurate for long linkers. The restrictions imposed by the lattice are most clearly displayed by the red and green lines in Figure 3.11B, where there are no instances of end-to-end separations of 2.5 lattice spacings. This is not due to convergence issues, but rather because there are no possible ways to have an end-to-end separation of 2.5 for a random walk of length 5 on a cubic lattice. Whereas the hard and ideal linkers are restricted by both the available end bead sites and the path along lattice vertices that they can take to connect the two end beads, implicit linkers are only impacted by the first of these. The curve for implicit linkers in Figure 3.11B is substantially smoother than for the hard and ideal linkers, indicating that the non-uniform radial distribution of binding sites has a small impact on end-to-end distribution, but that the path

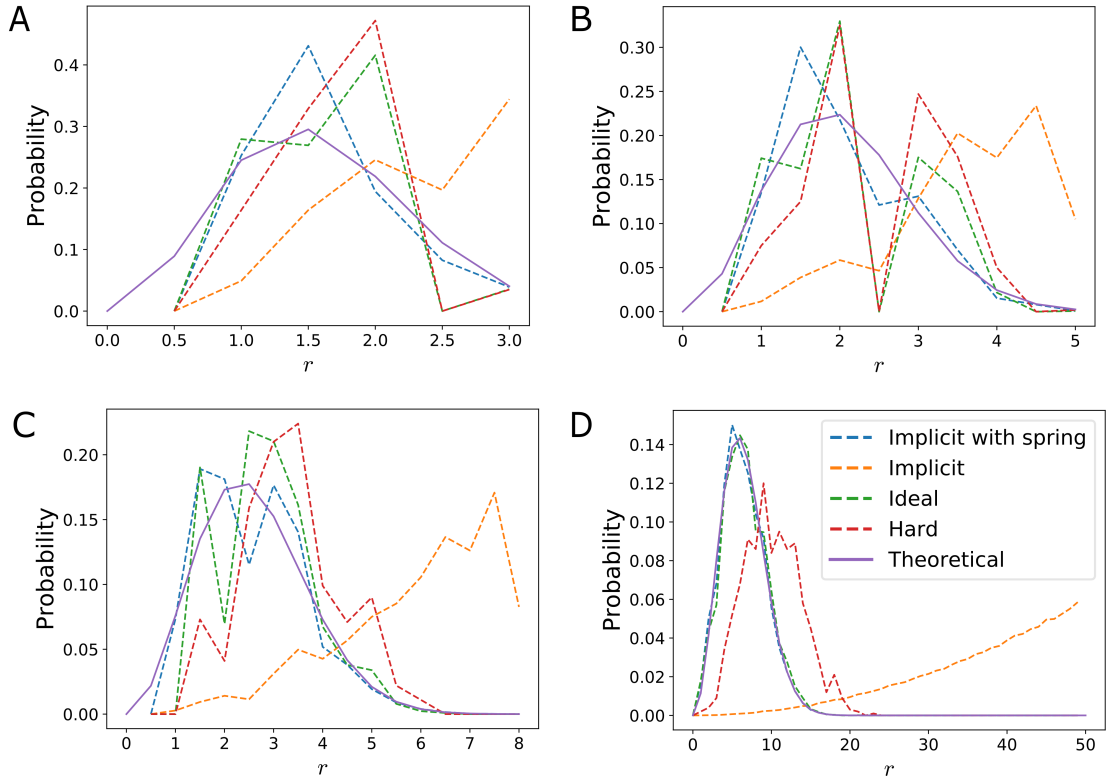


Figure 3.11: Probability distribution function of the end-to-end separation of chains consisting of one linker capped by a binding bead on each end. The linker lengths are (A) 3, (B) 5, (C) 8 and (D) 50.

restrictions imposed by the lattice are more significant. The differences between the ideal and the entropically adjusted implicit linkers show the impact of finite chain effects only, as they are under the same lattice restrictions. As the linkers get longer, the cubic lattice and finite linker length effects become negligible. This explains the good agreement between the theoretical distributions and the observed distributions for ideal and implicit linkers with an entropic spring, see Figures 3.11C and D.

The entropic spring applied to the implicit linkers causes greater reduction in end-to-end separations compared to the ideal linkers, for a linker length of 3. This would explain why no phase separation occurs for polymers of linker length 3 and valency 5, with entropically adjusted implicit linkers, but limited phase separation is observed for ideal linkers with the same linker length and valency. This is because the end-to-end distribution function used to calculate the entropic adjustment is only accurate for long linkers. As the linker length increases, the agreement in end-to-end distributions between the ideal and entropically adjusted linkers improves, which is borne out in the good agreement of the phase diagrams for these two linker types for systems with linkers of length ≥ 5. The entropic spring correction is shown

to be accurate at longer linker lengths, as shown by the excellent agreement between the end-to-end distributions of implicit linkers with an entropic spring and both the ideal linkers and the theoretical end-to-end distributions in Figure 3.11D for linker length 50.

3.3.1 Comparison with a previous lattice model for LLPS of biopolymers

As mentioned in Chapter 1, computational studies of a similar system to the one we are interested in were conducted by Pappu *et al.* [68]. The results we present are in agreement with many of the findings in their work. The main areas of agreement are: high density of clusters forming for polymers with implicit linkers [168], suppression of phase separation for polymers with hard linkers, and the correlation between phase separation and polymer valencies and linker lengths as discussed previously [68]. One of the key differences between the two studies is the binding model. Our primary binding model is one where the binding vectors are explicitly modelled, unlike the study by Pappu *et al.*, where beads have limited valencies but there is no explicit modelling of the binding direction. Furthermore, translating between these two schemes is non-trivial, as the bonding model adds a binding entropy term which cannot be accounted for or corrected by changing the interaction strength; therefore these two different binding schemes are not equivalent. The research of Pappu *et al.* has also involved studying the role of linker type in the overall cluster structure, with findings that a combination of hard and implicit linkers may facilitate the formation of clusters with a core-shell structure [168]. Our simulation of ideal linkers provides further information on the role of linkers on phase separation, as well as being an attractive linker for simulations where linkers should not occupy physical space but must avoid obstacles, such as in Chapters 4 and 5 of this thesis.

3.4 Conclusions

This exploratory work into the phase separation of model proteins has helped to illustrate the importance of linker type, linker length and polymer valency in the study of polymer phase separation. In combination with existing work on some of these systems [68, 168], we show that through having a large solvation volume —

represented by the hard linkers — phase separation can be suppressed, limiting the polymers to forming a sparse network of binding sites. On removing the excluded volume of the linker beads the structural behaviour of the polymers changes and phase separation becomes much more likely, with the formation of dense clusters. Much work has been done in this area on beads connected by implicit linkers [1, 68, 168], but we show that modelling the linker path, through our ideal linkers, accounts for linker entropy and alters the phase diagrams slightly, reducing the propensity of associative polymers to phase separate. Moreover, we show that similar phase diagrams can be achieved, at much lower computational cost, through the addition of a term to account for the entropy of the linker. There are some discrepancies between the results for ideal and implicit linkers with an entropic spring but the source of this is the effect of short linkers on a lattice and on the slight difference in binding entropies for the polymers of each linker type. Despite these differences, many of our findings have a universality, especially surrounding the observed cluster densities and trends relating to polymer lengths and valencies.

Our results indicate that hard linkers are unsuitable for the simulation of biopolymer phase separation, unless the simulations are of phase separation of polymers in a good solvent. In contrast, both implicit and ideal linkers show strong potential for such studies. The results for implicit linkers with an entropic spring and ideal linkers are in good agreement. This is pertinent as the computational expense of implicit linkers is significantly lower than that of ideal linkers, therefore this observation should facilitate much more expansive research into this phase separation. However, due to the approximations of the entropic spring at very short linker lengths and limitations of implicit linkers in the presence of obstacles, in the subsequent chapters we continue to utilise the ideal linkers to accurately account for the entropy of linkers with no excluded volume.

Finally, the extension/swelling of polymers in order to form a dense structure is a balance between the loss in configurational entropy in stretching the linkers, and the enthalpic and entropic benefits of binding beads being in close proximity to many other binding sites. For sufficiently strong interactions, the entropic cost of linker stretching can be overcome.

Chapter 4

Superselectivity in 3D: A compositional control mechanism for membraneless organelles

4.1 Introduction

The client-scaffold model of membraneless organelles would suggest that environmentally induced changes to the scaffold, such as by post-translational modifications (PTMs), temperature or pH, may result in significant changes in the composition of the organelle [34]. Furthermore, *in vitro* experiments have shown that changes to the scaffold can result in complete expulsion of particular clients and strong recruitment of others, indicative of a sharp binding transition with changing environment [32]. The mechanism of this highly specific recruitment behaviour is unknown but, given that superselective binding at a surface results in a very sharp binding transition with respect to receptor density [88], it is plausible that a similar effect could be present in membraneless organelles. Small changes in the cell environment can alter the number of available binding sites or the strength of interactions between the client and scaffold molecules, thereby shifting the binding site concentration above or below the transition density at which clients go from being predominantly bound to predominantly unbound. Our aim in this chapter is to demonstrate that superselective binding to a 3D host is possible, as to date it has only been studied for multivalent species binding to a surface of receptors [88, 90, 91, 93, 116].

4.2 System set-up

To study such an effect, a simplified model of a membraneless organelle is envisaged; the scaffold consisted of beads with excluded volume, such that each lattice site has a maximum occupancy of one, and the client is modelled as a single chain of beads. Some of the scaffold beads are capable of binding to the client beads, these are referred to as receptors. Scaffold beads incapable of binding are termed crowders, as they constitute the scaffold without interacting, beyond excluded volume effects, with any other beads in the system. In order to translate this into a configuration suitable for studying the recruitment and expulsion of a client, a cuboid system consisting of a scaffold structure adjacent to empty space devoid of binding receptor sites for the client (representing the cytosol) is constructed. This is illustrated in Figure 4.1, where the cuboid dimensions for the majority of the simulations are $100 \times 50 \times 50$ lattice sites, with the scaffold occupying one half of the cuboid and the other half being empty. The long dimension is the x axis, and by varying the x dimensions, the ratio of scaffold volume to free volume can be controlled. This set-up allows fair comparison of the client binding behaviour at different scaffold structures, as the cytosol remains constant despite changes in the scaffold (with the exception of systems where free volume is varied, as we will discuss later in this chapter).

The decision to simulate a static scaffold and mobile client is based on experimental observations of a number of different research groups. Firstly, membraneless organelles have been shown (using FRAP) to rapidly exchange material with the cytoplasm whilst maintaining their shape for timescales of hours [34, 50, 172, 173]. Additionally, the retention time and mobilities of different organelle components have been shown to vary significantly, with clients being much more mobile and transiently located in the condensates than their scaffold components [32, 39, 174, 175]. Finally, *in vitro* studies have been undertaken, in which stable condensates form, and are then capable of recruiting other proteins from the system [63, 176].

Generation of the scaffold cube initially involved filling a box with two species of associating polymers and allowing them to equilibrate and explore different configurations, as was done in Chapter 3. These gave a representative scaffold for a system containing two protein species. However, it was found that very similar results could be obtained by simply distributing receptor beads randomly within the

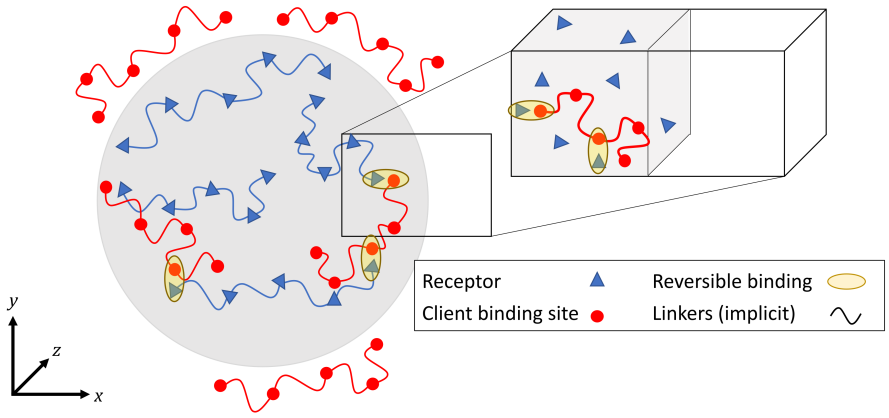


Figure 4.1: Schematic of the simulation set-up. Three types of molecules are simulated, corresponding to the client, receptor and crowder beads.

cube, as shown in Figure 4.2 where the probability of the client being bound, θ, for linker-length ≥ 10 agree well with those for a scaffold of randomly distributed beads. This is beneficial in that it allows us to have a much more general model, and reduces the chances of any artefacts arising due to the scaffold setup. Furthermore, whilst the client components are often enzymes and shorter proteins, scaffolds components consist largely of RNA and RNPs which tend to be longer than the clients [32]. Therefore, we assert that the longer linker-lengths in Figure 4.2 are more akin to actual biological scaffolds. As a final comment on the scaffold construction, where necessary, crowders are introduced to the system in the same manner as the receptors — by random distribution within the cube. This will be discussed further in due course.

After scaffold construction, the simulations involve the generation of the client chain and subsequent exploration of the system, using the chain growth procedure and the binding and displacement moves described in the Chapter 2. The two types of linkers used within this work are implicit linkers (occupy no volume and just act as invisible tethers) and ideal linkers (a contiguous random walk of linker beads with no excluded volume).

As with many biological systems, membraneless organelles consist of multivalent species with well defined binding regions [7] in combination with intrinsically disordered proteins or regions of proteins with weak but widespread non-specific interactions [53, 147]. To mimic these, we consider both specific directional and isotropic interactions between the receptor and client beads in this work, as introduced in the Chapter 2.

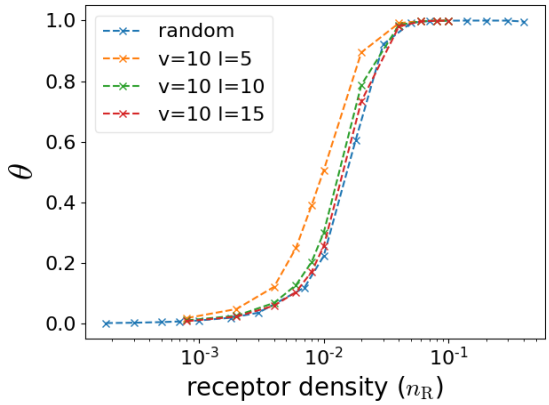


Figure 4.2: Equivalent results for the binding probability, θ, as a function of the receptor density, n_{R}, for cases when the scaffold beads are randomly distributed and when they are connected as polymers and are allowed to explore different configurations. For the polymer scaffolds, v is the number of binding beads on each polymer chain and l is the length of linkers connecting the binding beads along the polymer backbone.

4.3 Results

To probe the binding behaviour of the client, the probability of client being bound θ is measured across a range of receptor densities n_{R}. The steepness of the binding transition, from being predominantly unbound to predominantly bound, with increasing receptor density determines whether superselectivity is at play. In order to make the steepness of this transition more clear, we will use α [88],

$$\alpha = \frac{d \ln \theta}{d \ln n_{\mathrm{R}}}. \tag{4.1}$$

The key difference between the employment of this parameter in our 3D work, compared to that done previously on a 2D surface by Martinez-Veracoechea and Frenkel [88], is that n_{R} is the packing fraction of receptor beads, not the surface density of receptor. Despite this, in both definitions, an α value greater than 1 indicates that the binding is superselective, with the magnitude of α determining the degree of superselectivity.

Initial investigations into superselective binding to a 3D host indicate a lack of superselectivity in this system. As previous work on superselective binding at surfaces has shown that low interaction strengths are often conducive to superselective binding [90], a wide sweep of interaction strengths is necessary to ensure we are sampling the relevant parameter space. The results can be found in Figure 4.3, where interaction strengths ranging from $f = 0.01kT$ to $4kT$ are plotted. However, no

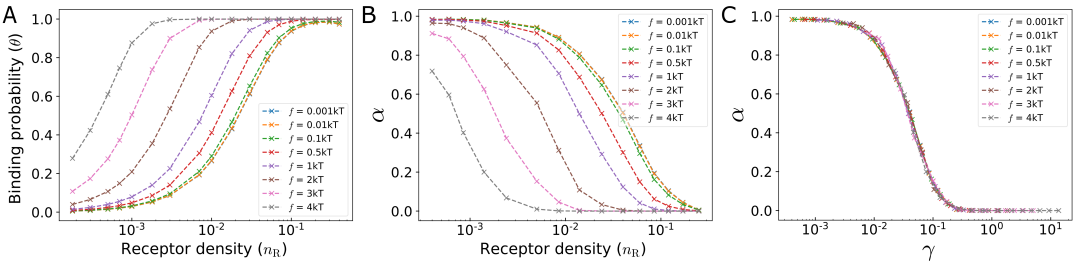


Figure 4.3: Simulation results for a decavalent client exploring a system of receptors in the absence of crowders. The strength of the directional interaction between the client and receptor beads range from $0.01kT$ to $4kT$. (A) Client binding probability θ as a function of receptor density γ. Superselectivity parameter $\alpha = d\ln\theta/d\ln n_R$ against (B) receptor density n_R and (C) scaled receptor density $\gamma = n_R e^{-f/kT}$.

superselective binding is observed for this system; indeed, the curves for the various interaction strengths all collapse on top of one another on the rescaling of n_R to γ, where $\gamma = n_R \exp[-f/kT]$, as shown in Figure 4.3C .

We then carry out a survey of various other parameters of the client in order to promote superselectivity and to see whether superselective binding is actually possible in a 3D host. As the superselective effect depends on there being a high entropic cost for initial binding followed by lower cost for stronger binding with a larger number of bonds, the alterations to the client are designed to maximise the entropic cost of initial binding. The parameters altered here are client valency and client linker length. Variation in these parameters do not lead to observations of superselective binding.

It was then reasoned that, by reducing the number of available sites in the scaffold box, the client would have to lose significantly more entropy in order to enter the scaffold. This can be achieved through the addition of inert crowders into the scaffold.

Although the addition of crowder beads to the simulation raises the receptor density at which the binding transition occurs, the significant difference is that the steepness of the transition increases, as evidenced in Figure 4.4A. When α is calculated for this system a clear peak above 1 is observed, see Figure 4.4B, indicative of superselectivity. This change highlights a crucial point about superselective binding; it does not necessarily increase the probability of binding at a given receptor density, in fact it often suppresses it, but it does increase the sharpness of the binding transition. The effects of crowders are significant — crowder packing fractions as low as 0.1 result in superselective binding.

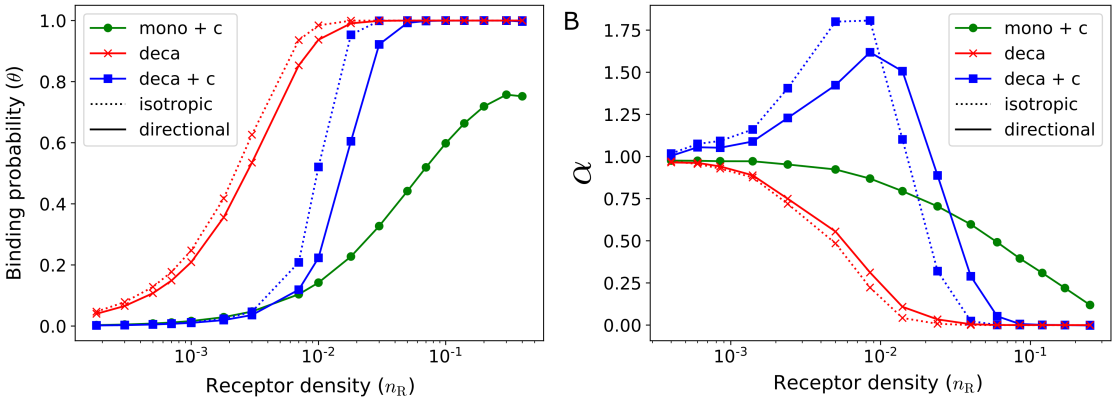


Figure 4.4: Dependence on receptor density n_R of (A) the probability of chain binding θ, and (B) the variation in α ($d \log \theta \,/\, d \log n_R$) for a sample of systems. The interaction strength is $f = 4kT$ and +c in the figure legend indicates that crowders are present in the system (at packing fraction of 0.4). Mono and deca refer to polymer valencies of 1 and 10, respectively.

The addition of crowders is appealing as it is also more representative of a membraneless organelle, which are relatively dense structures with a much smaller density of available binding sites to which the clients can bind [32].

4.3.1 Role of multivalency in superselectivity

In order to understand the origins of this superselective binding in the polymeric nature of the client, the behaviour of v monovalent clients is compared to that of a v-valent client. For instance, a decavalent client can be compared to ten monomeric clients. This comparison is achieved by generating histograms of the number of beads on a multivalent client that are bound at any one time and comparing that to the predicted number of monovalent clients that would be bound. This prediction is found from the binomial expansion of the probability of a single bead being bound. Explicitly, to calculate the probability of m independent monomers being bound, p_m, the binomial expansion of:

$$p_m = C_m^v {p_1}^m (1 - p_1)^{v-m} \tag{4.2}$$

is used, where p_1 is the probability of binding from the single monomer simulations and $C_m^v = [v!]/[m!(v-m)!]$. As monovalent clients do not exhibit superselectivity, this comparison elucidates the reasons behind the superselectivity of the multivalent clients. In Figure 4.5, the comparison is made between a decavalent client binding to

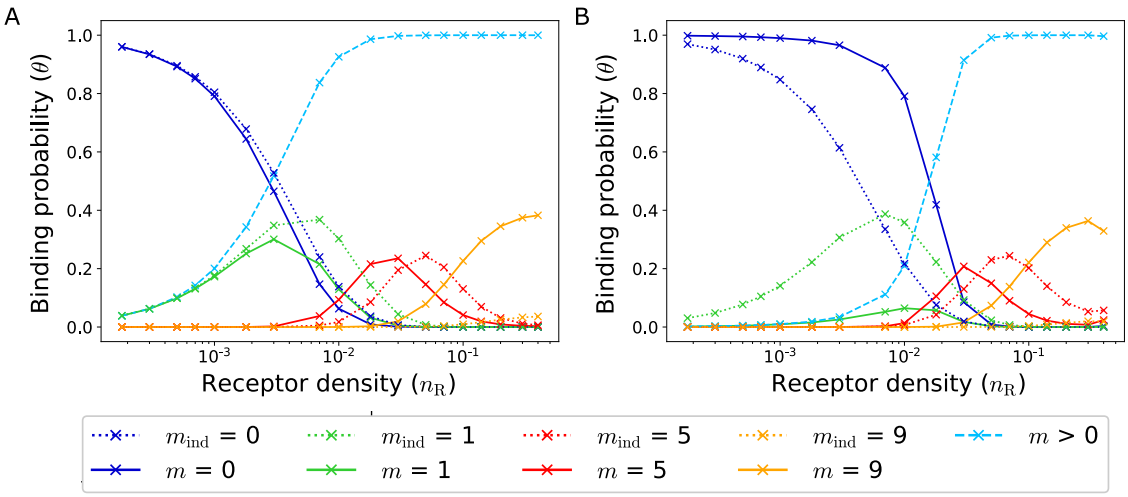


Figure 4.5: The probability of a decavalent client of linker-length $l = 5$ being bound to receptors by m beads for a range of receptor densities. To highlight the cooperative binding effect, results are also shown for ten independent beads binding to a host, where the dotted lines show the probability of m_{ind} of these beads being bound. These probabilities are shown for a system (A) without crowders and (B) with crowders at packing fraction 0.4. The receptor-client interactions are directional, with strength $f = 2kT$.

a receptor structure containing crowders at packing fraction of 0.4 and one consisting only of receptors.

For the system without crowders the probability of the multivalent client being bound by m beads maps closely onto the probability of m monovalent beads being bound. For higher values of m, cooperative effects come into play, meaning that it is significantly more likely that the multivalent client will be bound extensively than the equivalent number of monovalent clients. However, these cases concern receptor density values where all clients are bound and hence do not affect the binding transition. For the system containing crowders, on the other hand, there is a clear suppression of the probability of a low number m of polymer beads being bound compared to monovalent clients. As m increases, the probabilities become similar before the multivalent client becomes more likely to be bound, due to the same cooperativity argument used in the system with no crowders. It is therefore this suppression of initial binding followed by an increase as the number of bonds, m, rises which results in superselectivity. In other words, by narrowing the receptor density range in which the increase in number of bonds occurs, superselectivity is enhanced.

Evidence of superselectivity is also found in the change in free energy on binding. The free energy is calculated by taking the negative logarithm of the probability of m

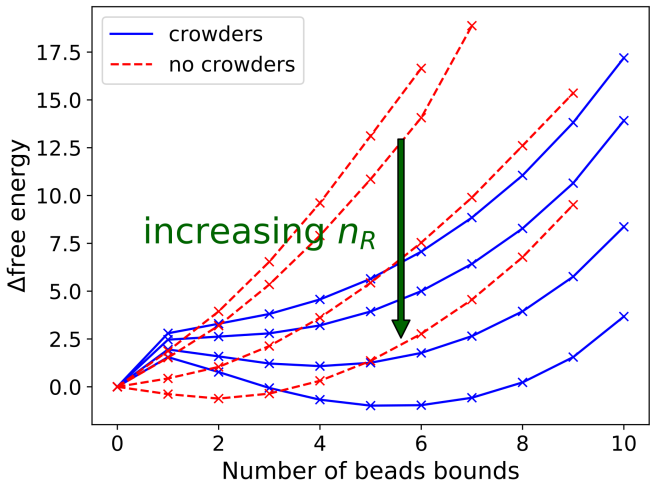


Figure 4.6: Free energy change, compared to an unbound chain, for a decavalent client of $l = 5$ forming m directional interactions to receptor beads, for a system with crowders at packing fraction 0.4 and without crowders. The interaction strength between clients and receptors is $2kT$.

beads being bound. The probability p_m of m beads being bound follows a Boltzmann distribution, therefore the relative free energy for a system with m beads bound is: $F_m = -kT \ln p_m$. The change in energy is then simply: $\Delta F = F_m - F_{unbound}$, which when applied to our system becomes: $\Delta F = kT(\ln p_{unbound} - \ln p_m)$. To facilitate comparison between different systems, the probabilities of binding are divided by the probability of the client being unbound, as this sets the unbound free energy to zero. Figure 4.6 shows that for the system containing crowders there is a free energy barrier to forming the first bond, followed by a minimum (at sufficiently low temperatures). In contrast, without crowders the free energy change increases with increasing degree of binding, indicative of an insignificant drop in entropy loss as the number of client beads bound increases. In combination, these results support the idea that the crowders increase the entropic loss on initial binding, thus increasing the superselectivity of the system.

4.3.2 Optimising superselectivity

We now turn our attention to optimising the system for maximum superselectivity. This involved altering the crowder density, client valency, client-receptor interaction strength, and client linker length, the results of which can be found in Figure 4.7.

Beginning with density, it is evident from Figure 4.7A that superselectivity in-

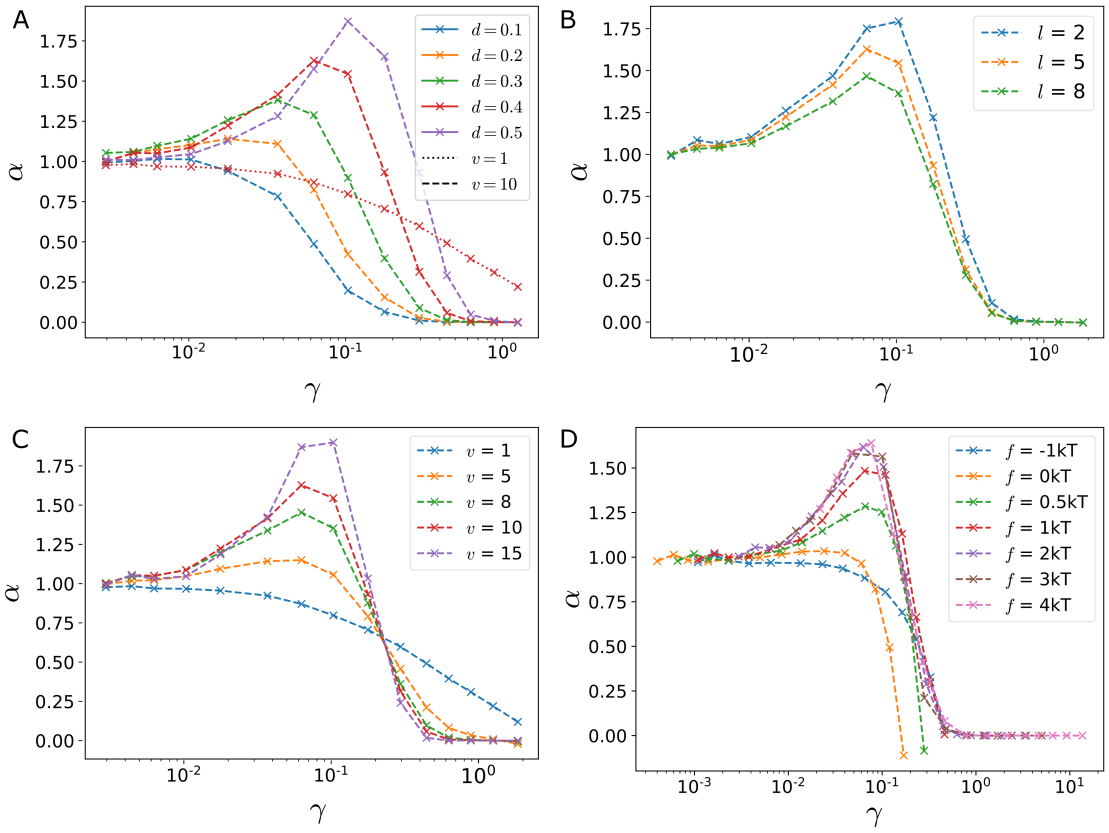


Figure 4.7: Dependence of α on receptor density (n_R) for a range of client parameters: (A) the density (packing fraction d) of crowders in the receptor cube, (B) the linker length of the client l, (C) the valency of the client v, and (D) the client-receptor interaction strength f. In each of these plots the variables are $v = 10$, $l = 5$, $f = 2kT$ and $d = 0.4$, unless stated otherwise. These results are all for simulations with directional binding.

creases with increasing density. This is attributed to the greater loss of client entropy on entering more dense structures. As previously discussed, entropic loss on initial binding of the client dictates whether the binding will be superselective. There are, of course, limits to this trend; for crowder density greater than 0.7 the client struggles to enter the scaffold due to excluded volume effects.

The next parameter to be studied is that of the client linker length. In arguments similar to the previous chapter, beads connected by shorter linkers are more constrained by and correlated to their neighbours, therefore on initial binding lose a greater proportion of their entropy than a similar client with longer linkers.

The client valency also plays a very important role in the degree of superselectivity of the binding. Increasing valency results in greater superselectivity, which is likely due to longer chains having greater initial configurational entropy when unbound, which is then reduced significantly on formation of the first bond. The increased number of client beads also permits the client to form multiple bonds

simultaneously, which again strengthens the superselective effect.

Finally, and somewhat surprisingly, the client-receptor interaction strength appears to have no effect, in the range $f = 1$ to $4\,kT$, on the superselectivity of binding. As the interaction strength f drops sufficiently low, and for repulsive interactions, superselectivity is reduced and eventually destroyed. The main reason for this is that the excluded volume effects become stronger than the enthalpic payoff for binding. On increasing the interaction energy this entropic barrier is overcome. The relative temperature independence is a result of two connected but opposing effects. On one hand, the increase in binding strength reduces the receptor density at which binding begins to occur. On the other hand, this means the receptors are, on average, further apart thereby reducing the probability of the client forming multiple bonds. This lack of availability of receptors and increased entropic loss on stretching the client chain to connect such dispersed receptors means there is a cap on the magnitude of superselectivity of any system of fixed linker length l, valency v and receptor density d, regardless of the binding strength. This is particularly interesting as it provides a counterexample to the arguments made by Tian, Angioletti-Uberti, and Battaglia [107] that the weak binding 'required' for superselective binding would be easily overcome by various supramolecular interactions such as van der Waals, aromatic, hydrophillic and hyrophobic, in addition to hydrogen bonding. This, however, appears to overlook the relative independence of superselectivity on temperature — temperature only acts to shift the receptor density of the binding transition. Therefore the superselective interaction strengths could be increased, such that they are dominant over the other interactions at play.

To fully optimise the superselectivity of the system, an extensive parameter sweep would be required. This was prohibitively computationally expensive for this work, therefore this research focussed on identifying trends in the binding behaviour. From the finding presented earlier in this chapter, maximising the superselectivity would most likely involve using a polymer of short linker length and high valency, and a 3D host with moderate to high packing fractions of crowders. The exact combinations of these parameters would depend on their inter-dependancies, and the limits of the ranges in which the trends we have uncovered are accurate/applicable.

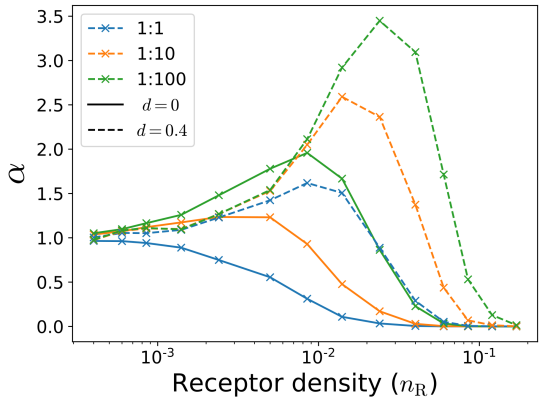


Figure 4.8: Simulation results for a decavalent client of linker length 5 and interaction strength $2kT$ binding for a scaffold with (dashed lines) and without (solid lines) crowders for three scaffold to free volume ratios.

4.3.3 Free space dependence

As previously discussed, superselectivity relies on a significant loss in entropy on initial binding. This can also be achieved by increasing the free volume in the system. In our simulations this is done by increasing the volume of the empty box adjacent to the scaffold. The results of this are shown in Figure 4.8, where it can be seen that by increasing the ratio of scaffold volume to free volume to 1:10, superselectivity can be induced in systems without crowders. Further increase in free volume results in even greater peak α values, indicating an even sharper binding transition. This phenomenon can be explained by the increase in total translational entropy for an unbound client in a larger free volume, which is then lost on the client binding to the scaffold. The degree of superselectivity also increases with free volume for scaffolds containing crowders.

We now turn to a comparison between the two scaffold systems – with and without crowders – for increased free volume relative to scaffold volume. This is achieved through the study of the relationship between multivalent and monovalent client binding.

The findings are shown in Fig. 4.9, where the scaffold to free volume ratio is 1:10 and the scaffold contains no crowders and crowders at density 0.4 in Figure 4.9(A) and Figure 4.9(B), respectively. A suppression of binding by low m beads for the multivalent client compared to monovalent clients, followed by increase for larger m, is indicative of superselective binding, and this is present in both cases. Crucially, the effect is more pronounced for the system containing crowders. Therefore, increasing

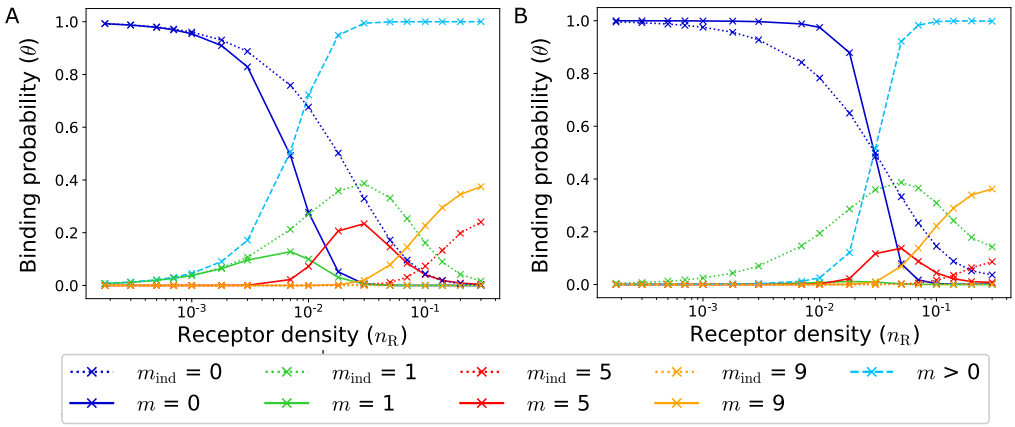


Figure 4.9: Probability of m beads being bound in a system of parameters $l = 5$, $f = 2kT$ and with scaffold to free volume ratio of 1:10. The dashed lines are for ten monovalent clients and the solid lines are for a single decavalent client. These plots are for systems (A) without crowders and (B) with crowders at density $d = 0.4$.

the free volume and having crowders can act in concert to promote a steep binding transition – otherwise known as strong superselectivity.

In order to investigate the limits of the free volume effect, simulations are carried out for scaffold to free volume ratios up to 1:1000. This is the largest system which is computationally economical. However, we can go beyond this volume limit using simple statistical mechanical arguments. The partition function of the client can be de-constructed in to three separate terms:

$$Q = Q_{\mathrm{b}} + Q_{\mathrm{ref}} + (V - V_{\mathrm{ref}})Q_{\mathrm{u}}. \tag{4.3}$$

Here Q_{b} is the partition function for client configurations with at least one bead bound to the scaffold, Q_{ref} is for unbound configurations in a reference free space volume V_{ref} from the scaffold, and Q_{u} is for unbound configuration per unit volume far from the external faces of the scaffold. Therefore, Q_{b} and Q_{ref} account for all client configurations affected by the excluded volume and interactions of the scaffold, and as such, V_{ref} must be large enough to encompass all client configurations which could be impacted sterically by the scaffold. This also means that these two terms are independent of the free volume V, provided that it is larger than V_{ref}. In contrast, Q_{u} is independent of the scaffold structure and accounts for the client configurations within the free volume, beyond V_{ref}, and therefore depends on the volume V of the system.

The decrease in binding probability P_b, at any given receptor density is linear

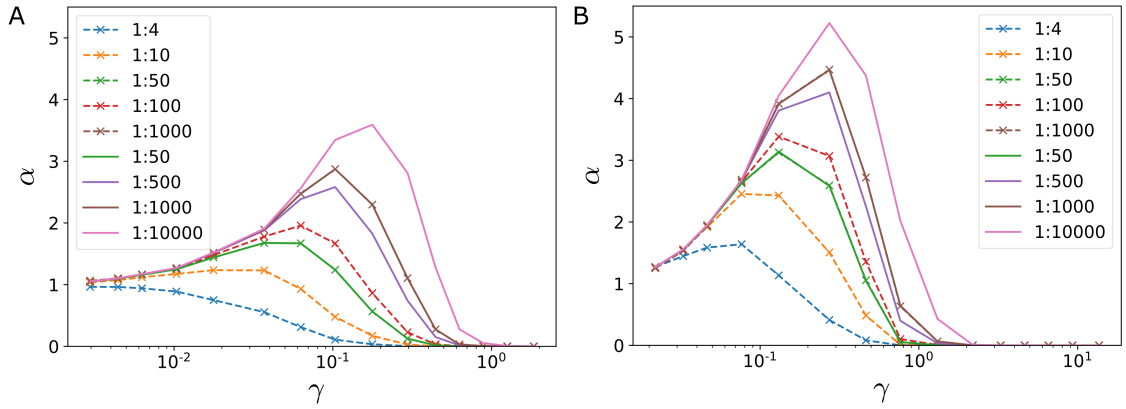


Figure 4.10: Superselectivity parameter α for a single client in simulation systems with various ratios of scaffold volume to free space. Dashed lines are simulation results, and solid lines are predicted curves.

with respect to the increase in volume from a reference value $V - V_{\text{ref}}$. The probability of binding can be described by the following relationship between partition function Q and the system free volume:

$$P_{\text{b}} = \frac{Q_{\text{b}}}{Q_{\text{b}} + Q_{\text{ref}} + (V - V_{ref})Q_{\text{u}}}, \tag{4.4}$$

This can be rearranged to:

$$\frac{1}{P_b} = 1 + \frac{Q_{\text{ref}}}{Q_{\text{b}}} + (V - V_{\text{ref}})\frac{Q_{\text{u}}}{Q_{\text{b}}}, \tag{4.5}$$

which explains the linear relation observed in the simulations.

By simulating clients binding to scaffolds for several free volumes the values for $Q_{\text{ref}}/Q_{\text{b}}$ and $Q_{\text{u}}/Q_{\text{b}}$ can be found. This must be done for each receptor density, but this then allows extrapolation of the binding probability, using these values, to produce curves of binding probability against receptor density for significantly larger free volumes. The agreement between the predicted curves and the simulation results for volume ratios of 1:50 and 1:1000 demonstrate the accuracy of this method. Therefore, extrapolation of the binding probability provides a straightforward, accurate and inexpensive method of predicting the binding curves of clients for systems with large free volumes. From these predicted curves, Figure 4.10 showing superselectivity parameter α against γ is produced. Superselectivity increases with increasing free volume but, crucially, crowders continue to significantly enhance superselectivity, even at very large free volumes.

Linking these findings back to membraneless organelles, rough measurements

of organelles' nearest neighbour separation, in *in vivo* systems, show a volume ratio around 1:1000, but with significant variance [2, 18, 50]. This was found through analysis of fluorescence spectroscopy images, using both cross-section images of cells, and reconstructed 3D images, by stacking cross-sectional images of cells *. This allowed the organelle volume to be calculated along with the volume of the cytoplasm, providing a mechanism by which to find rough volume ratios. Our results show that crowders are not strictly necessary for superselective binding to a 3D host, but they do significantly improve the fine compositional control of the 3D host.

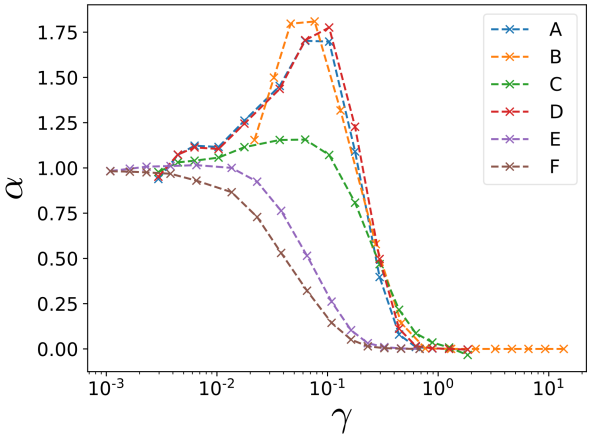
4.3.4 Ideal linkers

To understand the entropic effects of explicitly modelling the beads which constitute the linkers, as opposed to the ideals linkers used up to this point, simulations are run at several points in parameter space for polymers with ideal linkers in the directional binding regime. The results from this work are shown in Figure 4.11 where, despite the change in linker altering the quantitative results, the trends observed for the implicit linkers are still present. Superselectivity remains independent of interaction strength, and is correlated to both crowder density and polymer valency. The impact of linker length is slightly weakened, which may be accounted for by a greater entropic loss for ideal linkers compared to implicit linkers when the binding sites enter the dense structure and become more constrained. This is likely exaggerated by the high entropy of the ideal linkers when the polymer is unbound. These results provide further evidence that the key observation of superselectivity enhancement by the introduction of crowders is robust with respect to different ways of modelling the linkers.

4.3.5 Multiple clients

To this point, our simulations have involved simulating a single client binding to a scaffold. This significantly lower client concentration relative to scaffold is in line with previous studies on client-scaffold systems, including the pioneeering work by Banani *et al.* [32] and Jo and Jung [174], among others. Typically these studies involve client concentrations two orders of magnitude lower than the scaffold.

*These images were provided by our collaborators Dr Carl Jones and Dr Sushma Grellscheid from the University of Bergen.



	d	v	l	f / kT
A	0.4	10	5	2
B	0.4	10	5	4
C	0.4	5	5	2
D	0.4	10	2	2
E	0.1	10	5	1
F	0	10	5	1

Figure 4.11: Superselectivity results for clients with the ideal linkers in six systems.

To probe when this dilute limit assumption breaks down, we simulate multiple clients within each scaffold system. Full binding curves are shown for 5 clients, Figure 4.12A and B, and 50 clients, Figure 4.12C and D. There is strong agreement between the binding curves of single clients and those 5 and 50 clients, corresponding to client densities of $\phi = 8 \times 10^{-5}$ and $\phi = 8 \times 10^{-4}$ respectively. On further increase in client number beyond 50, the results deviate from the single client curves – with a clear deviation appearing once the client population reaches 100. For these higher client concentrations, the competition between clients starts to have a significant impact on the binding behaviour of individual clients, where previously they were inconsequential.

There is limited data on the exact *in vivo* concentrations of clients and scaffolds within real membraneless organelles [175], but, as mentioned before, seminal work done by Banani *et al.* on client scaffold systems focus on low client concentrations. Higher concentrations of clients relative to scaffolds would blur the distinction between clients and scaffolds, resulting in physically and structurally different scaffolds and organelles [42]. Studying the impact on client concentration on the propensity for phase separation and on the structure and dynamics of such organelles would be a very interesting avenue for further research.

4.3.6 Isotropic binding

To investigate the robustness of superselective binding to a 3D host, especially considering the range of interaction types in membraneless organelles [147], isotropic binding is implemented within the system. This is defined in more detail in Chapter

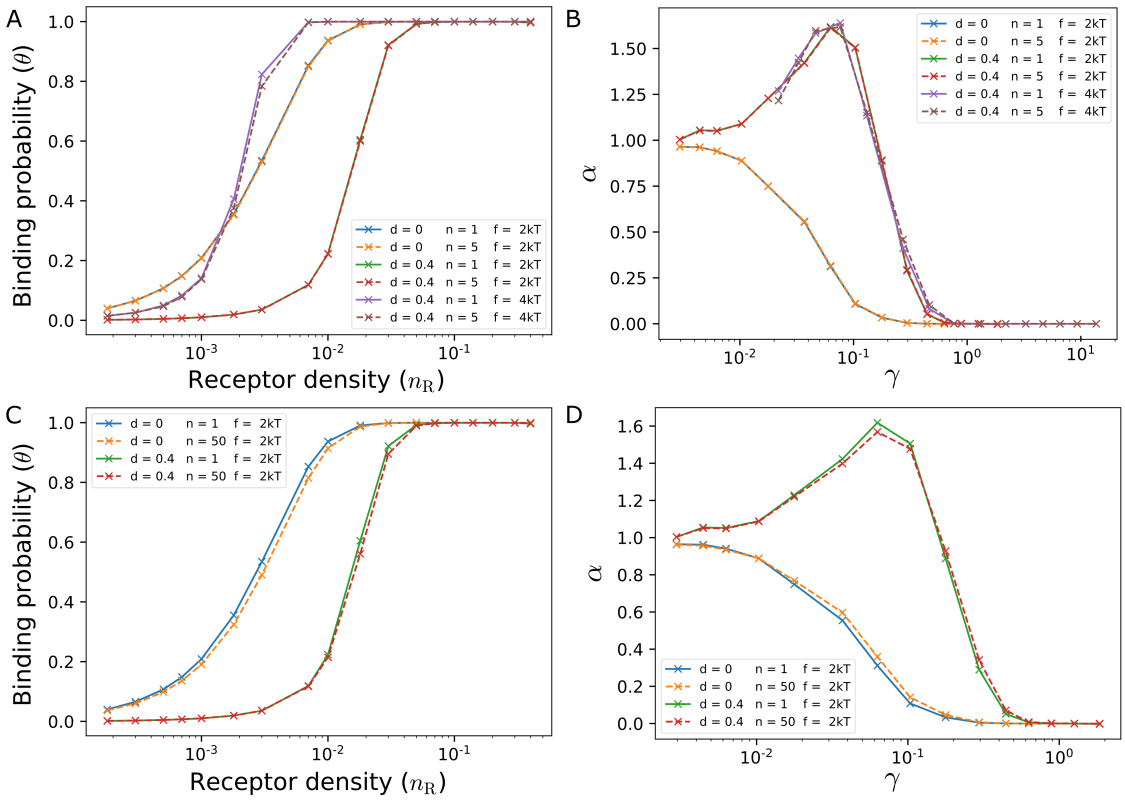


Figure 4.12: Binding probability θ of a given chain and superselectivity parameter α for multiple client simulations, where $n = 5$ clients were used in (A) and (B) and $n = 50$ clients were simulated in (C) and (D). In each case, the results for a given crowder density d and binding strength f are practically identical.

2, but briefly, all clients on adjacent lattice sites to receptors are automatically bound and all beads can form multiple bonds. This is in contrast to the directional case, where beads are limited to a maximum valency of one, and binding is determined probabilistically.

As can be seen from Figure 4.4, the isotropic results align relatively closely with those of the directional binding regime. In this binding regime, systems without crowders still exhibit no superselectivity, provided that the free volume is sufficiently small. In systems with crowders, isotropic binding results in some enhancement of superselectivity for the decavalent client relative to the directional binding case, for reasons which will be discussed shortly.

Through studying the nature of the client binding, shown in Figure 4.13, the mechanisms behind the binding behaviour can be exposed. As with the directional binding case, there is a clear suppression of binding with a small number of beads for the multivalent client compared to monomers followed by cooperative effects for forming multiple beads binding simultaneously. This is characteristic of superselective binding.

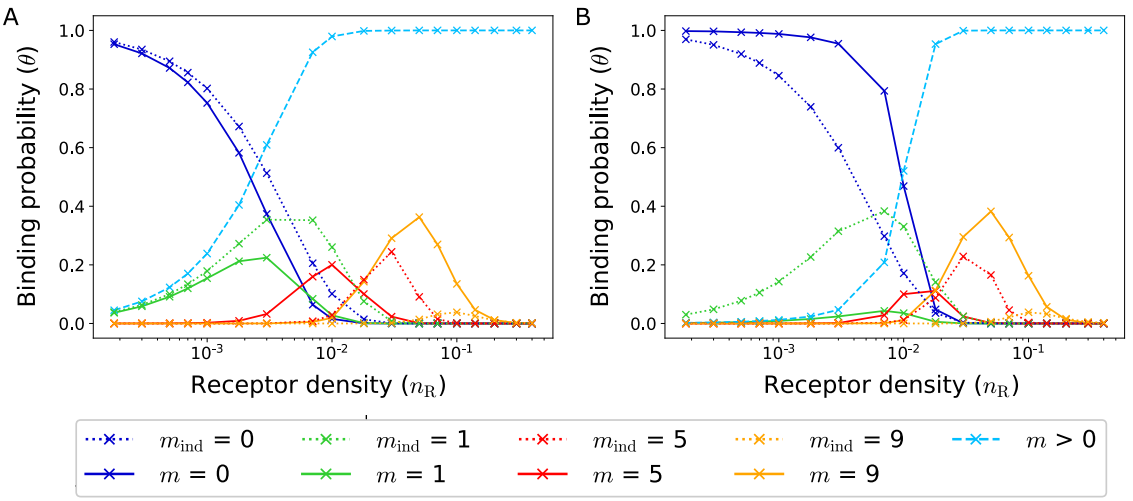


Figure 4.13: Plot showing the probability of a decavalent client of $l = 5$ being bound to receptors by m beads for a range of receptor densities. This is for a system (A) without crowders and (B) with crowders at packing fraction 0.4. The interactions present are isotropic, with strength $f = 2kT$.

This is further supported by Figure 4.14, where the change in free energy is plotted against the number of beads bounds. The superselective binding again results from an initial barrier to forming a single bond followed by an energy minimum for forming multiple bonds (blue lines in the figure). Where superselectivity is not at play there is no peak or discontinuity, just a continuous curve.

Permitting client and receptor beads to form multiple bonds has the potential to change the binding behaviour of, and configurations exhibited by, the client. This could be borne out in the client surrounding one or more receptors to maximise the number of interactions it has when receptors are scarce, a phenomenon we term as 'encircling'. To explore this more thoroughly, a parameter sweep is exhibited in Figure 4.15.

Initially the results follow similar trends to those in Figure 4.7 with superselectivity being correlated to the crowder density and the client valency, and being largely independent of the interaction strength. In contrast to the directional binding system, however, there appears to be relative independence of superselectivity with respect to linker length.

In order to find the origin of the difference between the binding behaviour in the two regimes, a more thorough analysis of the client's binding configurations is required. We monitor the number of client beads bound to each receptor (receptor occupancy w). The results of this study are shown below, where Figure 4.16 is for a decavalent client in systems with various interaction strengths and Figure 4.17 is for

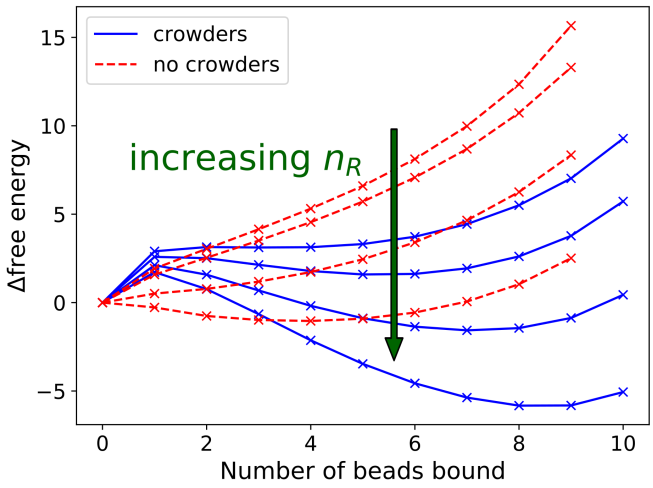


Figure 4.14: Free energy change for a decavalent client of $l = 2$ forming isotropic interactions to receptor beads through m beads, for a system with crowders at packing fraction 0.4 and without crowders. The interaction strength between clients and receptors is $2kT$.

clients of several linker lengths. It should be noted that each client bead can only bind to any given receptor once, therefore the occupancy gives the number of unique client beads that are bound to any single receptor. Another important consideration is that the plots are normalised by the number of binding events, irrespective of the number of unbound configurations observed.

It is apparent when looking from left to right of Figure 4.16, that the probability of receptors having multiple occupancies increases with increasing interaction strength. Another observation is that the highest occupancies are only likely at very low receptor densities. The occupancy trends can be explained as follows. At low receptor densities, the client is unable to bind to many different receptors due to the geometrical restraints of the linkers, but it still faces the entropic barrier to entering the 3D host. Therefore at weak interaction strengths binding is not common and receptors are seldom bound by more than 2 beads. However, for sufficiently strong interaction strengths the client can overcome the entropic penalty by binding to one receptor with many beads and thus gaining significant energy benefits. This 'encircling' behaviour requires the client to occupy a 'collapsed' or constrained configuration which results in a significant configurational entropic penalty. Despite this entropic penalty, Figure 4.16C shows that this does not prohibit client binding as it can be overcome by the favourable interaction energy from multiple bonds. This 'encircling' behaviour is least entropically costly for short linkers. As the density of receptors increases, the clients are capable of binding to multiple unique receptors.

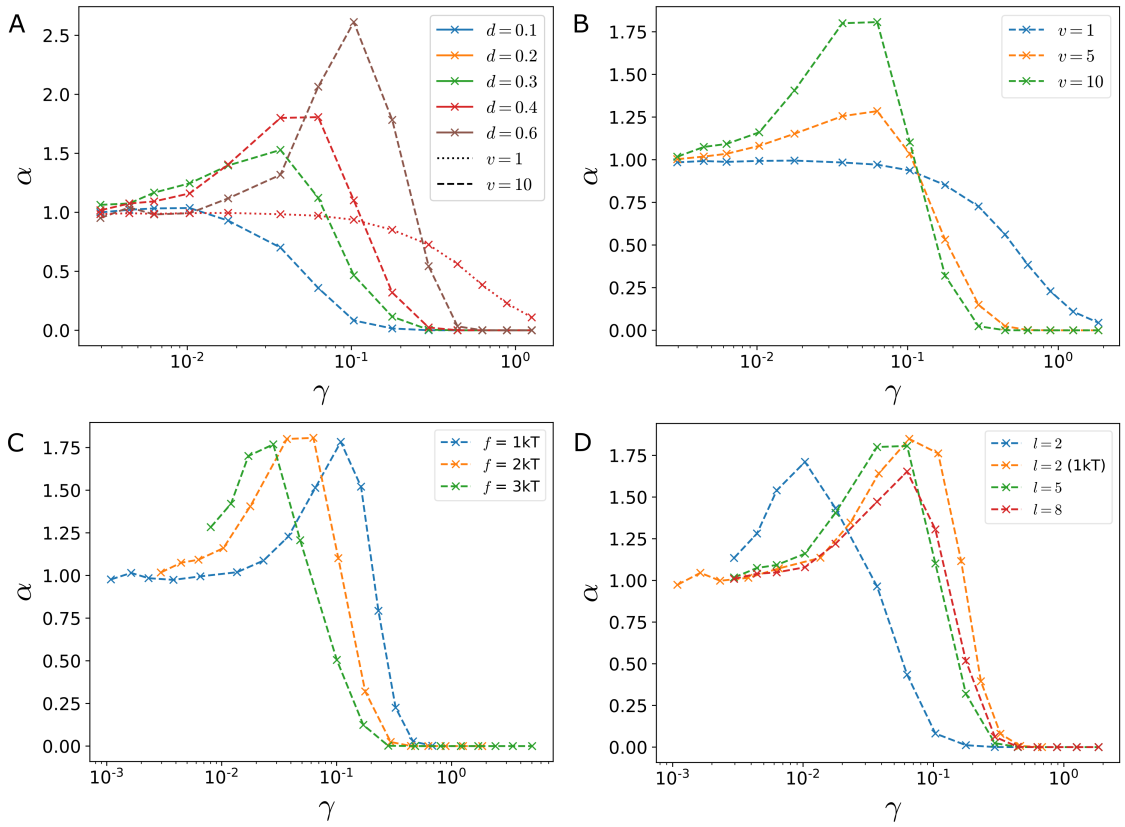


Figure 4.15: Plots of α against receptor density (n_{R}) for a range of client parameters: (A) the density (packing fraction d) of crowders in the receptor cube, (B) the valency of the client v, and (C) interaction strength f between the client and receptors and (D) the linker length of the client l. In each of these plots the variables are $v = 10$, $l = 5$, $f = 2kT$ and $d = 0.4$, unless stated otherwise. These results are all for simulations with isotropic binding.

Therefore, the 'encircling' behaviour ceases as it is entropically very costly to be in a such compact configurations. This explains the significant rise in probability of receptors having single occupancy in Figure 4.16B and C.

The pinching in Figures 4.16A, 4.17B and 4.17C is a result of plotting the normalised occupation numbers. Figure 4.17D shows the unnormalised occupation numbers used to produce Figure 4.17B; the probability of binding is very low until a receptor density of around 10^{-2}, before increasing significantly. This is not accounted for in the normalised figures, as these just show the relative probabilities of each value of w (number of bonds). Finally, Figure 4.17E shows the collapse of the client radius of gyration corresponds with the increase in probability of $w > 1$ shown in Figure 4.17D.

Linking 'encircling' back to Figure 4.15D, the low γ value of the peak for linker length 2 compared to lengths 5 and 8 for $f = 1kT$ is due to the 'encircling' behaviour which permits significant binding at much lower receptor concentrations. It

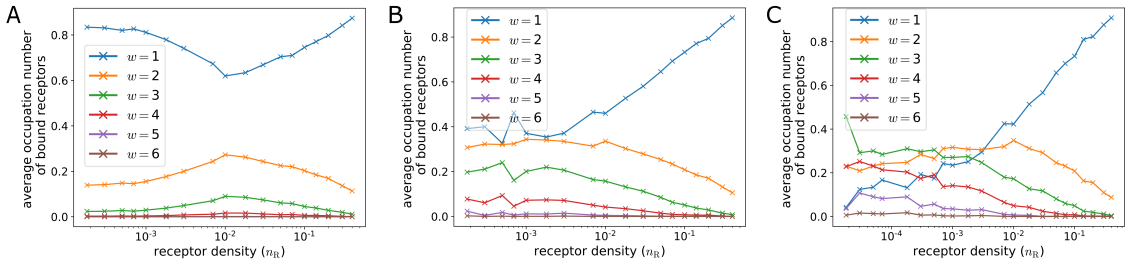


Figure 4.16: Normalised probabilities of bound receptor occupation numbers for a decavalent client of linker length 5, binding with an interaction strength of (A) $f = 2kT$, (B), $f = 3kT$ and (C) $f = 4kT$

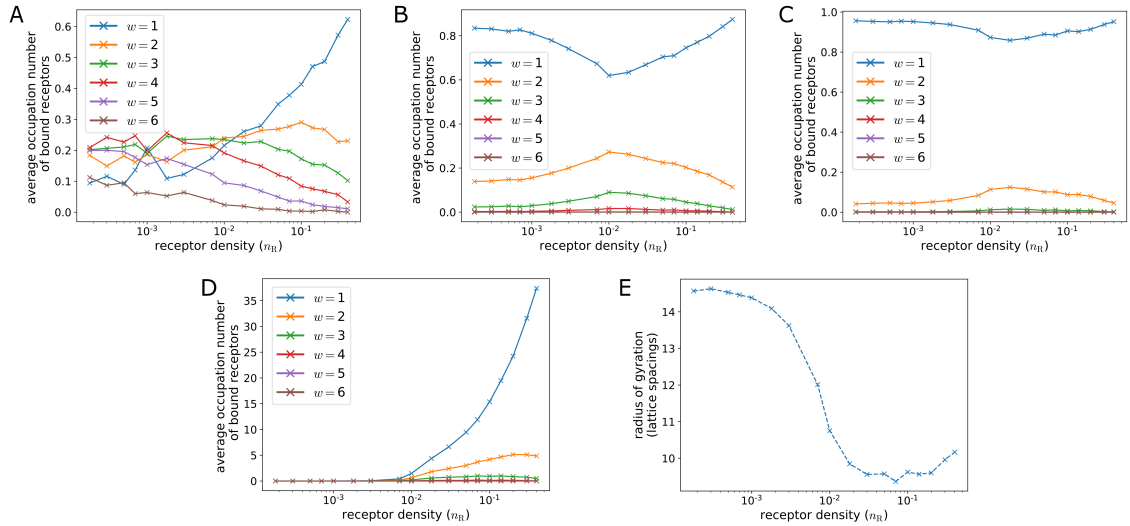


Figure 4.17: Plot showing normalised probabilities of occupation numbers for bound receptors against n_R for a decavalent client of linker length (A) 2, (B), 5 and (C) 8 with $f = 2kT$. (D) is the unnormalised probability of w for a system with $f = 2kT$ and $l = 5$, and (E) is the average radius of gyration of the client when bound.

is interesting that this mechanism change does not appear to significantly affect the magnitude of the α peak, indicating that this 'encircling' is a binding mechanism which can be superselective.

4.3.7 Biological relevance

A crucial finding of Banani *et al.* is that client recruitment varies non-linearly with increasing concentration of scaffold species (which the client binds to) [32]. Hence, simple stoichiometry arguments are incapable of accounting for this behaviour. Due to the strong expulsion and recruitment of membraneless organelles, indicative of a sharp binding transition, we propose that superselectivity could be the mechanism controlling this behaviour.

We have shown that receptor density controls the binding of a client to a 3D scaffold, and that changes in interaction strength can sharply switch the client from being bound to unbound or *vice versa*. Biologically, the changes in receptor density or interaction strengths could be achieved through alterations in the cellular conditions [18], PTMs [177] or by oxidative or reductive processes [178], all of which are already associated with membraneless organelle formation. PTMs act by altering interactions between protiens and between proteins and RNA; this is through configurational changes or chemical changes to binding sites. There are a wide range of PTMs including phosphorylation, glycosylation, acetylation and methylation [22, 34–37, 179], providing extensive means by which interaction strengths and scaffold receptor density could be tuned. Additionally, changes in the cell environment including pH, salt concentratiom, temperature and glucose availability could also act to alter the number of binding sites on the scaffold which could also result in a sharp transition in the location of the client. Changes in cellular environment have already been shown to alter organelle formation, structure and physical properties [22, 34, 179, 179, 180]. Furthermore, it has been borne out during this work that sufficiently high crowder densities are required for superselective binding. Support for this key role of crowders can be found in the presence of hundreds of molecular species within membraneless organelles [181, 182], which provide a potential source of crowders.

4.3.8 Potential applications

There are a number of potential application for this phenomenon in supramolecular multivalent structures, such as in diagnostics or drug delivery. Promising studies into hydrogels for tissue engineering [183], drug delivery vehicles [184] and biosensors [185] have been carried out, and the polymeric nature of the hydrogel could act a scaffold which can expel or recruit materials. There is active research into the use of nanoparticles for drug delivery [186], and superselectivity may provide a new mechanism for payload release at the target, i.e. environment controlled, rather than the light trigger release often deployed [187]. Therefore this research may allow development of drug delivery vehicles with different properties and the potential for new components and materials to be used in these devices. This could aid in one or more of functionality, cost and ease of production. In fact, synthetic membraneless

organelles capable of controlled recruitment and delivery of a payload within cellular cytoplasms has already been achieved, where protease was used to stimulate deployment of cargo [63].

A potential use of this finding outside of biology is in supramolecular chemistry, where formation of carefully tuned structures could be achieved within solution, instead of at a surface [77]. It could also potentially used in solutions where aggregates are used to distribute or remove a species from the fluid.

4.3.9 Conclusions

By studying the binding behaviour of a client to a 3D host of receptors in the dilute limit, we have found that superselective binding is only exhibited if there are crowders at a sufficiently high density within the structure, or if the free volume is much larger than the scaffold volume. Previous literature on superselectivity has focussed on surface binding [88, 90, 91, 91]. In contrast, here we expand the concept of superselective binding to porous 3D hosts. Crucially, we believe this to be a plausible mechanism by which membraneless organelles can control their composition. The key mechanism centres around the client paying a very significant entropic cost to enter the 3D host, thus suppressing initial binding, followed by forming multiple bonds. We have shown this to be a robust effect, which is present for two distinct binding cases and linker types.

It would be interesting to test this theory using an *in vitro* model system, similar to that of Li *et al.* [7], to see if it is observed in a physical system. This could involve comparison between the binding behaviour of a client to a scaffold constructed of a single species of protein and that of 2 or more species, to introduce crowders. Additionally, inducing PTMs to alter the availability of binding sites and monitoring the client recruitment could allow for greater control of receptor density than the stoichiometric methods used by Li *et al.* [7], where there was no indication of the concentration of 'available' scaffold binding sites.

This work was largely carried out in the dilute limit. Therefore further work would be required, using more coarse-grained systems or significantly parallelised code, in order to study higher client concentrations, where competition between clients would be a more significant factor. It would also be intriguing to discover the impact of higher client concentration on LLPS and scaffold properties.

Chapter 5

Utilising superselectivity to achieve polymer sorting at surfaces

5.1 Introduction

Superselective binding at a surface has been studied both experimentally and computationally [88, 90, 91], as reviewed in Chapter 1. The primary focus of these previous studies has been to maximise the superselectivity on a flat surface. In this chapter we probe the effect of the target surface topography on superselective binding and propose a mechanism by which this binding can be utilised to sort polymers based on their physical properties.

5.2 Surface simulations

We must first verify that the model we have chosen exhibits superselective binding for a multivalent client, as had been observed by Dubacheva *et al.* [90], and theoretically proposed by Martinez-Veracoechea and Frenkel [88]. An illustration of our system is given in Figure 5.1.

The system consists of a surface coated with receptors, at a range of surface densities. The receptors are positioned randomly on the surface and are immobile. To avoid the specific arrangement of receptors on the surface influencing the overall results a number of uncorrelated 'snapshots' of receptors on a surface are gener-

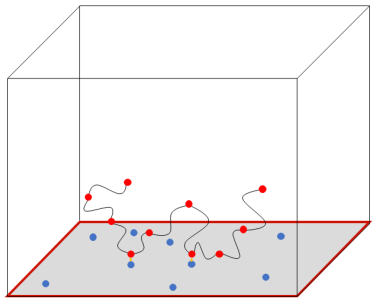


Figure 5.1: 3D illustration of the system. The receptor binding sites and the client beads are represented by the blue circles and the red circles respectively. The grey region is the surface.

ated at any given density and the results are averaged. Clients explore the system and the transition from the client being predominantly unbound to predominantly bound, henceforth referred to as the binding transition, is measured. The client is represented as binding beads connected by implicit linkers for this initial work, but different linker types will be discussed later in this chapter. In the vast majority of simulations the binding beads interact via the (non-vector) specific, directional binding model discussed in Chapter 2. There are several simulation studies where the isotropic interaction scheme is used, but this will be explicitly mentioned as we present results for these specific cases.

In order to characterise the binding behaviour during the simulations, the probability θ of the client chain being bound to the surface by at least one bead is monitored. The results for a single client binding to a surface can be seen in Figure 5.2A; the binding transition is much steeper for the decavalent client compared to the monovalent client, in agreement with the work of Dubacheva and Frenkel [90]. As in Chapter 4, we define

$$\alpha = \frac{d \ln \theta}{d \ln n_{\mathrm{R}}},$$

but with n_{R} being the surface density of receptors, rather than the 3D packing fraction of receptors. Recall from Chapter 4, that the binding is superselective if the increase in logarithmic binding probability is faster than the rate of logarithmic receptor density increase. Hence, as shown in Figure 5.2B, this means that α will increase from 1 to a peak, and then decay to 0 as θ levels off at 1. In subsequent plots of α where multiple interaction strengths are shown, such as in Figure 5.2,

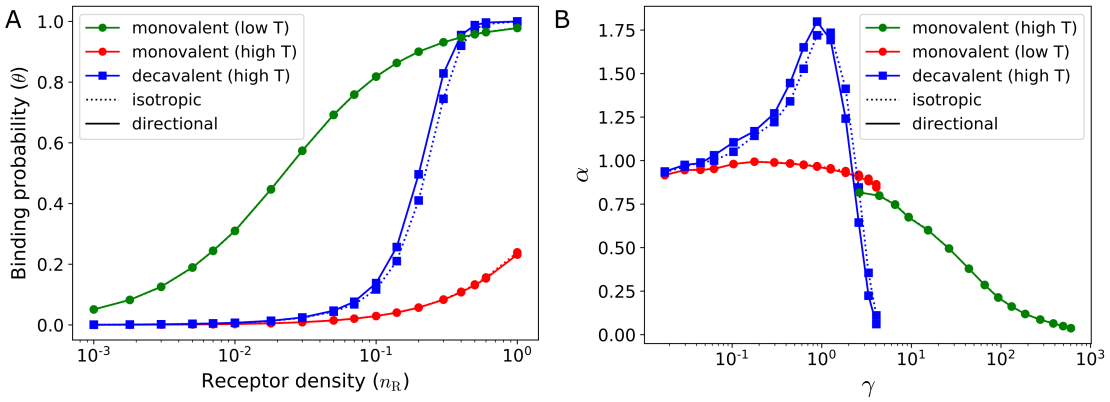


Figure 5.2: Binding probability for a single client chain (with binding bead valency shown in the legend) at a surface of receptor sites. Plot (A) shows the probability of at least one bead on the chain being bound against the density of binding sites on the surface. Plot (B) shows how α varies with γ, where both of these terms are described in the text above. High and low temperature refer to binding strengths of $2kT$ and $7kT$ respectively.

$\gamma = n_R e^{-f/kT}$ is used to allow the n_R coordinate of the α peak to be compared.

5.2.1 Role of the impenetrable surface

The initial simulations we discuss are those of a single client binding to an impenetrable surface decorated with receptors. Superselectivity relies on the client losing significant entropy on binding to a surface by a small number of bonds but a much smaller entropic loss for any additional bonds it makes with the surface. A solid surface coated with receptors allows for a polymeric client to bind superselectively because the client loses translational entropy on binding, but also loses configurational entropy as it binds. To investigate whether both the translational and configurational entropy loss are required for superselective binding, the impenetrable surface is replaced with a penetrable one, where the configurational entropy loss should be lower. A representation of this system is given in Figure 5.3. As with the impenetrable surface, receptor beads are distributed randomly on a single x-y plane, but with the key difference that the client beads can occupy sites on this plane, provided the site is not occupied by a receptor bead. A further consideration for the porous surface, is that the client can straddle the 'surface', unlike for the impenetrable case, but this has little impact on the binding transitions observed.

The binding transition for a client binding to a porous surface is still superselective, but much less so than for an impenetrable surface, as shown in Figure 5.4. A parameter sweep of the polymer valency, linker type, linker length and interaction

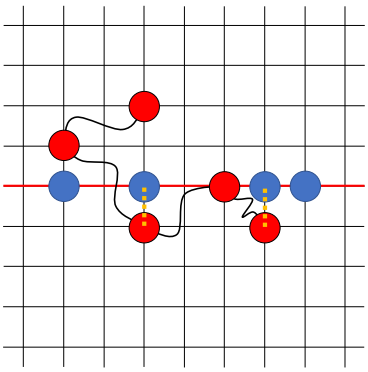


Figure 5.3: Schematic cross-section of the 3D system of a suspended plane (red line) of binding sites (blue beads) which the client can bind to.

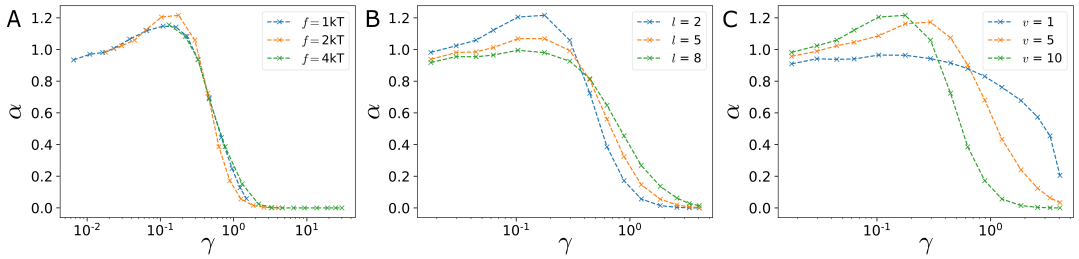


Figure 5.4: Superselectivity parameter α for a client with implicit linkers binding to a suspended plane of receptors at various densities. The impact of changing (A) interaction strength between client beads and receptors, (B) linker length l and (C) client valency v on the superselectivity are shown. Unless otherwise stated the interaction strength between the client beads and the receptors is $2kT$, the client linker length l is 2 and the client valency v is 10.

strength prove insufficient to increase α to values comparable with those in Figure 5.2. This can be largely attributed to the significant conformational entropy that the client maintains after binding to the porous surface with a single bond. It should be noted that the shift in the binding transition to lower γ values is a result of there being a greater probability of binding since the receptors can have up to six adjacent lattice sites where binding is possible, unlike in the solid surface case where there are only ever two lattice sites adjacent to each receptor where binding can occur.

The free energy change of the client with respect to the number of beads bound to the surface is shown in Figure 5.5 (calculated using the procedure detailed in Chapter 4). The impenetrable surface has a free energy barrier on the order of kT to forming one bond, followed by a minimum (at sufficiently large receptor densities n_{R}). In contrast, while there is a slight discontinuity in the gradient on going from $m = 1$ to $m = 2$ for the penetrable surface, no barrier is observed. The curves rise at large m because the client has highly restricted configurations when the majority of its beads are bound.

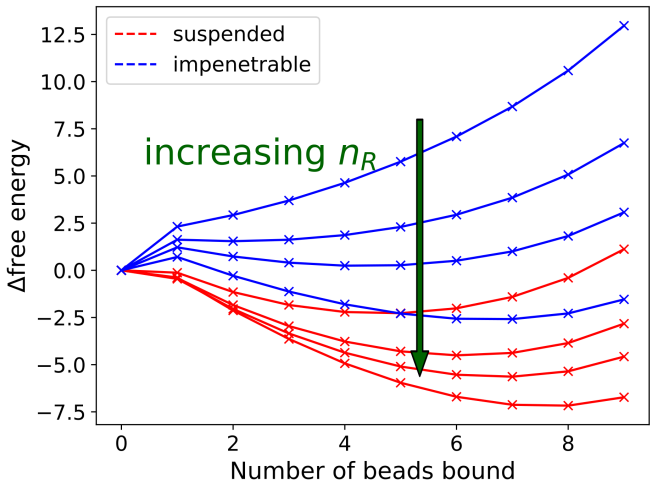


Figure 5.5: Change in free energy against number of client beads bound, m, for a decavalent client of linker length $l = 2$ binding to suspended and impenetrable surfaces with interaction strength $f = 2kT$.

Simulations of clients binding to plane of receptors with an impenetrable layer set a displacement z below, as shown in Figure 5.6A, allows us to move more continuously between the impenetrable and fully penetrable cases. The results of this impenetrable layer can be seen in Figure 5.6B; by increasing the vertical distance between the receptors and the impenetrable surface, we reduce the superselectivity. Therefore, a significant proportion of entropic loss on initial binding for the impenetrable surface is due to the reduced volume available for the client beads to explore, and corresponding reduction in configurational entropy. Increasing this volume, by increasing z, reduces the entropy loss on initial binding thus suppressing superselectivity.

Increasing the volume of the system, whilst maintaining the area of the surface, increases superselectivity due to increased entropy loss on initial binding, as discussed in Chapter 4. Crucially, however, superselectivity remains superior for the impenetrable surface case. The impact of increasing the volume to surface area ratio is noteworthy when designing superselective systems, but for our application of polymer sorting we will be searching for differences in binding behaviour at a given volume. The difference between the γ values at which the probability of binding sharply increases for different linker lengths remains constant with increasing volume, as expected. This is because, for a sufficiently large box size, the configurational entropy difference between the different clients remains the same and the translation entropy is independent of linker length.

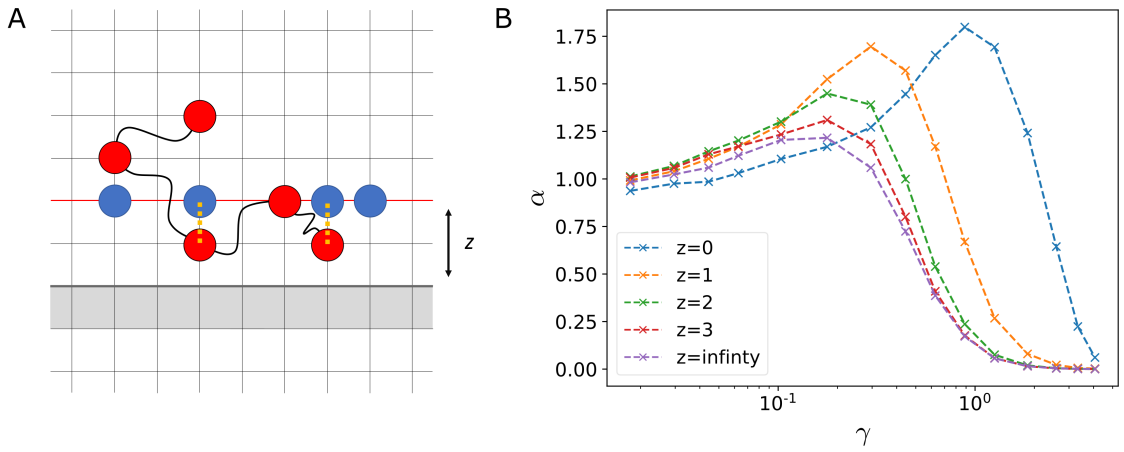


Figure 5.6: Figure (A) is a schematic cross-section of the 3D system of suspended receptor binding sites, with an impenetrable boundary (shown in grey) z lattice sites below the plane of receptor sites (shown by the red line). Figure (B) is a plot of α against γ for decavalent clients of $l = 2$ for various z values. $f = 2kT$ for these simulations.

5.2.2 Pitted surfaces

Given the difference between superselectivity at penetrable and impenetrable flat surfaces shown in Section 5.2.1, we expect that superselectivity can also be tuned by the topography of the surface. To investigate how significant an impact surface topography has, simulations of clients binding to a flat surface decorated with square wells (a "pitted" surface) are informative. A regular pitted surface is suitable for this, because it is simple to manufacture in order to test our findings experimentally [188–190]. It also allows for relationships between surface geometry and binding behaviour to be obtained — without the extensive sampling required at, for instance, a random surface — and should cause qualitatively different configuration changes to the client when binding compared to a flat surface. In order to maximise the superselectivity of this system, the receptor beads are restricted to sites on the base of the pores. This forces the chains to enter the pores in order to bind to the receptors, altering the shape of the bound conformations. The receptor densities quoted for this work on "pitted" surfaces are packing fractions of receptors with respect the the total area of pore bases.

The pitted surface is illustrated in Figure 5.7, with square wells of fixed depth, width and separation. Previous studies into the binding behaviour of polymers to a surface with square wells have focussed largely on the relationship between the curvature of the surface and the configurational entropy of binding at a patterned

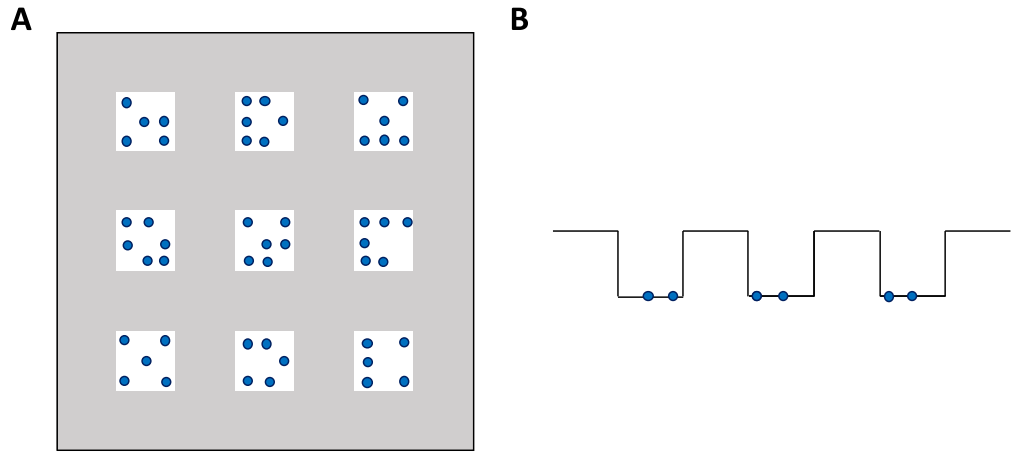


Figure 5.7: Schematic of the surface with square wells, where the white squares are the base of the pores, at a depth of 10 lattice spacings below the grey surface. The blue circles represent the random distribution of receptors across the base of the pores. The results we present are for a surface with geometry, in units of lattice spacings, of pore width 5, pore depth 10 and pore separation 5. (A) is the surface viewed from above and (B) is a 2D cross-section of the surface.

surface [191]. As we have square wells, and only the bases of these are occupied by receptors, the curvature on the relative attraction of the client to the surface does not play a role. This has the advantage of focussing the study on the impact that the confinement of the polymer by the pores has upon the binding behaviour. The polymers simulated in this work have either hard or ideal linkers; due to the convex curvature of the pores, implicit linkers could cut through the surface, therefore their length would not be conserved and as such they are not studied here. Modelling both hard and ideal linkers allows us to predict the effect of different polymer types or solvent properties on the polymer binding behaviour. It should be noted that the system of study here differs significantly from that of Chapter 3, where hard linkers were deemed to be inappropriate for modelling the formation of membraneless organelles, as we are now studying synthetic systems with different target polymers.

The pitted surface is harder for clients to navigate than the flat case, and clients are prone to getting 'trapped' in pores, therefore impairing the exploration of the surface. To overcome this, we use parallel tempering. Parallel tempering involves simultaneously running multiple simulations over a ladder of temperatures and periodically attempting to swap the configurations between replicas at adjacent temperatures [192, 193]. Any swap, from configuration a to b, is then accepted or rejected with probability

$$P_{ab}^{\text{acc}} = \min[1,\, \exp\left(-\Delta E_{ab}\left(\beta_a - \beta_b\right)\right)], \tag{5.1}$$

where $\beta_a = 1/kT_a$, $\beta_b = 1/kT_b$ and $\Delta E_{ab} = E_b - E_a$.

Parallel tempering overcomes the trapping issue as the high temperature simulations provide 'free' chain configurations, where the client is unbound, which the lower temperature chain configurations can swap with. This allows for good exploration of the surfaces. It also provides binding curves for multiple temperatures (equivalent to inverse binding strengths), which is also a potential mechanism for sorting polymers.

The temperature ladder used has 8 values. As our interaction strengths and temperature are intertwined, ϵ/kT, ϵ is varied to achieve the different temperatures. ϵ values are set at intervals of $1kT$, and run from 3 to $10kT$ for the ideal linkers and 5 to $12kT$ for the hard linkers. This choice of temperature ladder allows for full exploration of the surfaces without being prohibitively expensive. The lower range of interaction strengths for the ideal linkers, compared to that hard linkers, is to account for polymers with ideal linkers binding at lower interaction strengths than polymers with hard linkers. This allows us to observe the full range of binding behaviours for both linker types, without simulating polymers with hard linkers at interaction strengths where the probability of binding is vanishingly small, which would be uninformative for this work. Swaps of the client configurations between a client at a randomly selected ϵ on the ladder and the ϵ directly above it are attempted every 10^4 sweeps of the system. In total 10^5 swaps are attempted for every suite of simulations run at a particular surface (where a suite of simulations is the 8 simulations at various ϵ values on the ladder).

Implementing parallel tempering into the code necessitated the use of Fortran's Message Passing Interface (MPI). Ensuring the correct order in which messages were sent and received was challenging, and required extensive testing. Sample simulations of very simple systems were run, to check that the known probabilities of binding were reproduced at each interaction strength modelled. Implementing parallel tempering was time consuming, but without it, full sampling of the corrugated surface system would not have been possible.

Superselectivity is significantly enhanced at porous surfaces compared to flat surfaces, as illustrated for clients with hard linkers in Figure 5.8A and B. The

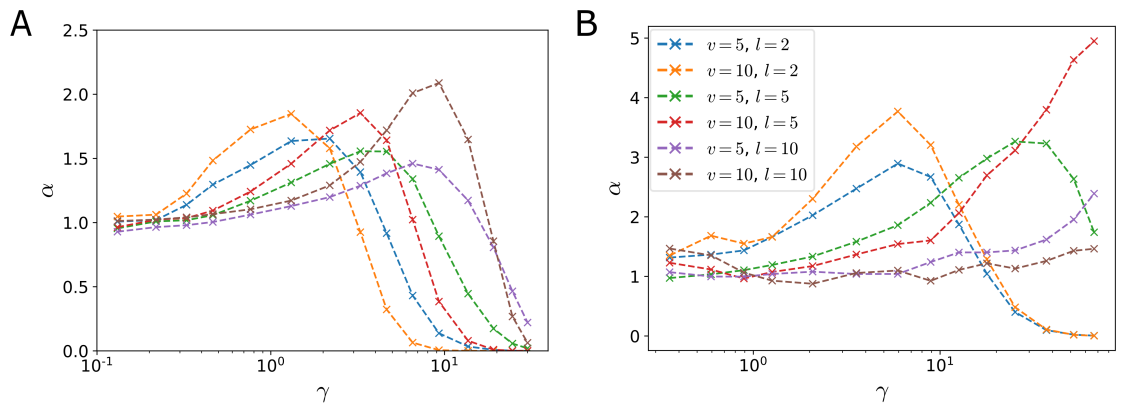


Figure 5.8: Superselectivity parameter α for penta- and decavalent clients with hard linkers of length 2, 5 and 10 binding at (A) a flat surface and (B) a pitted surface, with an interaction strength of $f = 4kT$.

enhanced superselectivity arises from a strengthening of the geometric constraints at an impenetrable surface, identified in Section 5.2.1. In order for a client to bind to the surface it must enter the pore, and thus loses a vast volume of space it can occupy. This is further compounded by the need to form multiple bonds to strongly bind to the surface, which means that a greater proportion of the chain must be in the pore, rather than having one end in the pore and the other protruding out into the free space above the surface. Therefore, we propose that the increase in superselectivity is due to the large loss of configuration entropy of the chain on binding to the pitted surface. It is the placement of the receptors at the base of the pores that forces the client to enter the pore at high entropic cost.

Comparing Figures 5.8A and 5.9A shows that linker type has a minimal impact on the superselectivity curves for binding at a flat surface. In contrast, the curves for pitted surfaces in Figures 5.8B and 5.9B differ noticeably. A major difference between the two plots for the pitted surface is the location of the maxima of the curves for species with linkers $l \geq 5$. For these longer chains, the larger total excluded volume of hard linkers becomes the dominant factor in determining whether binding occurs. This excluded volume effect is sufficiently large that the peak superselectivity of clients with linker length 10 is never reached. The magnitudes of superselectivity for the curves where comparisons are available are similar, indicating the linker type has little impact on the degree of superselectivity at pitted surfaces.

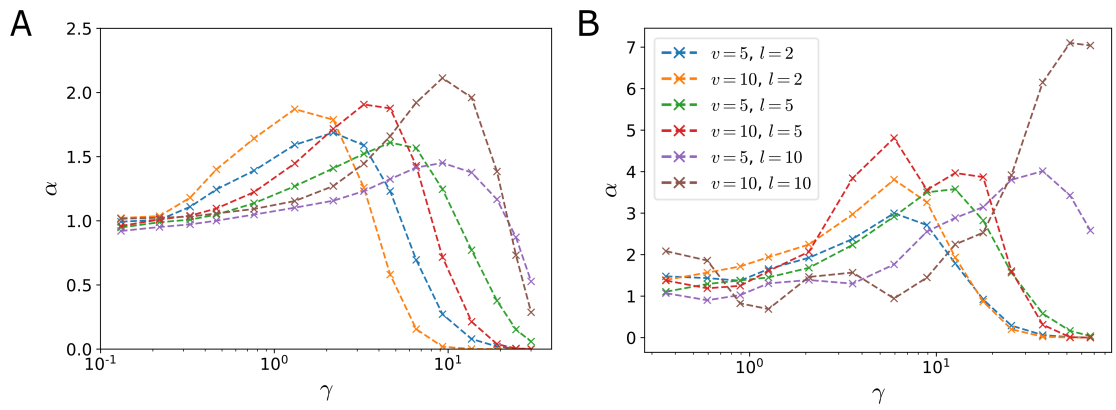


Figure 5.9: Superselectivity parameter α for penta- and decavalent clients with ideal linkers of length 2, 5 and 10 binding at (A) a flat surface and (B) a pitted surface, with an interaction strength of $f = 4kT$.

5.3 Sorting polymers using superselective binding

Superselectivity results in a characteristically sharp binding transition, as receptor density increases. This sharp, "on-off", binding behaviour could be utilised to control the adsorption of multivalent species to a surface. If the location of the binding transition is sufficiently sensitive to the polymer properties, then it should be possible to design a surface that will selectively bind certain polymers and not others, at given environmental conditions. This could have numerous potential applications including design of functional surfaces, polymer sorting and in controlling the self-assembly of supramolecular structures. The second half of this chapter is dedicated to determining the feasibility of, and the fundamental physics behind, such sorting behaviour.

5.3.1 Sensitivity of sorting based on polymer properties

To achieve sorting, we must maximise the difference between the receptor densities at which binding transitions occur for different polymeric client species. In this section, we examine the sensitivity of sorting with respect to both valency and linker length. The clients studied have valency 5 and 10, and linker lengths 2, 5 and 10. The binding behaviour of single polymers with ideal linkers to a flat surface is shown in Figure 5.10A. Overall, the spacing of the binding transitions is narrow and it would therefore be difficult to separate clients using such a surface. On moving to a pitted surface, there is a notable improvement in sorting potential. For example, there is a range of receptor densities at which clients with linker lengths of 2 would

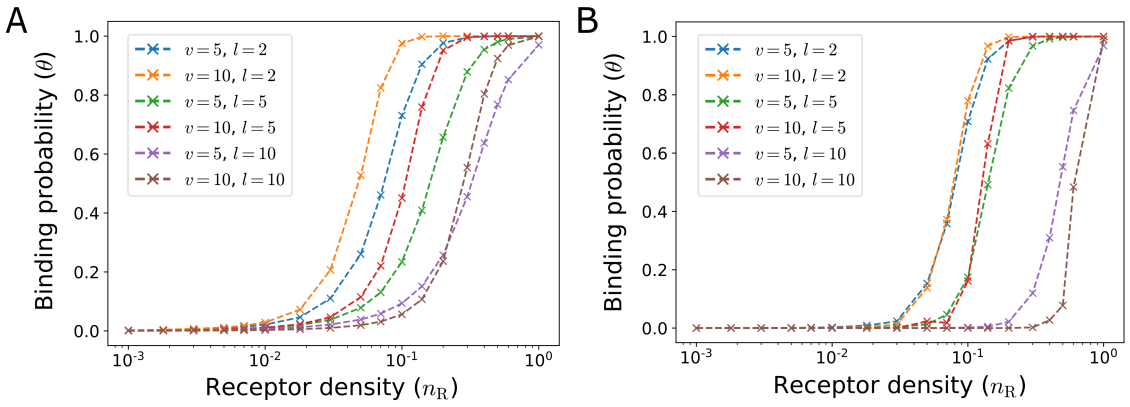


Figure 5.10: The probability of binding θ for a range of client valencies and linker lengths for binding at (A) a flat surface and (B) a pitted surface. The interaction strengths are (A) $4kT$ and (B) $5kT$. The clients in this case have ideal linkers.

be fully bound whilst those with linker length 10 would be unbound. Maximising this receptor range allows for the most reliable sorting, especially when considering the feasibility of experimental implementation. As higher interaction strengths are required to promote binding of the client at a pitted surface, the interaction strength between the client beads and the receptors is $4kT$ at the flat surface and $5kT$ at the pitted surface in all of the subsequent plots.

Enhanced sorting of clients with hard linkers compared to ideal linkers is shown in Figure 5.11B. Indeed, this system shows that a higher resolution of sorting is possible; for example, clients with linker length of 5 could be separated from those of linker length 2, a feat which is not possible for the ideal linker case on this surface. This result is promising in identifying a robust means by which polymers can be sorted based on both interaction strength and linker length. We note at this point that we do not study polymeric clients with valencies less than 5, as these have less steep binding transitions. Less steep binding transitions reduce the potential for there being a receptor density where one such polymer is bound whilst the other is unbound.

Having shown that sorting of polymers can be achieved by exploiting a pitted surface, and is most promising for polymers with hard linkers, our subsequent results are focussed on clients with hard linkers. This allows the points of difference between the flat and pitted surfaces to be illustrated most clearly.

The resolution of the sorting behaviour, with respect to linker lengths and valencies ≤ 10, is shown in Figure 5.12, where we find that the receptor density at which the binding transition occurs depends more strongly on the client linker length than

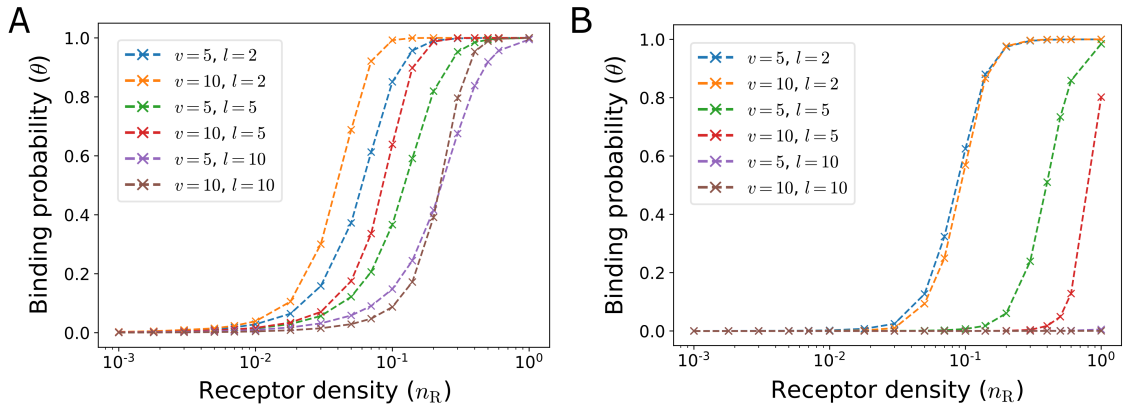


Figure 5.11: The probability of binding θ for a range of client valencies and linker lengths binding at (A) a flat surface and (B) a pitted surface. The interaction strengths are (A) $4kT$ and (B) $5kT$. The clients in this case have hard linkers.

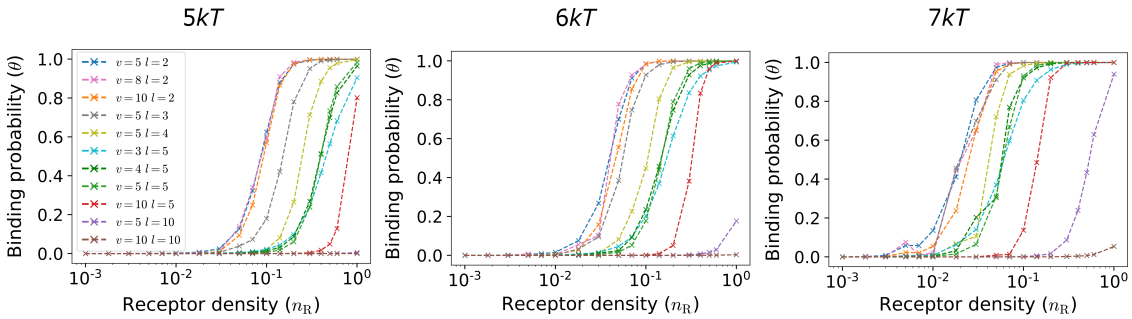


Figure 5.12: Client binding probabilities for multiple client valencies and linker lengths, where all the clients have hard linkers, to illustrate the resolution for sorting achievable in this system.

the client valency. The binding transition is only weakly impacted by valency, an effect most apparent when comparing the results for clients of linker length 5 and valencies 3, 4 and 5. In contrast, by comparing pentavalent clients with linker lengths in the range 2 to 5, there is significant separation between the binding transitions. Therefore, we can conclude that this sorting system is most sensitive to the polymer linker length.

Several interaction strengths are shown in Figure 5.12. This shows how interaction strengths can also be used to sort different polymer species. Again, the superselective binding transition facilitates separation, with high sensitivity to receptor density. The results for an interaction strength of $7kT$ illustrates one of the key limitations of this sorting mechanism; when the interaction strengths get sufficiently high, the binding transition ceases to be superselective. This is because it occurs at very low receptor densities, which do not permit polymers, of the length used in this study, to form multiple bonds. The pores cause this issue, as at very

low receptor densities there may be a maximum of one receptor in any one pore, therefore polymers would have to reach into multiple pores to form the multiple bonds required for superselectivity. This is therefore a significant consideration in designing a system to sort polymers. The receptor density limit at which superselectivity ceases depends on both the pore geometries and the polymer length, as these alter the receptor density at which the polymer is unable to form multiple bonds simultaneously.

5.3.2 Thermodynamics of binding

The free energy of binding can be used to interrogate the entropic and enthalpic drivers of polymer sorting at a pitted surface. This is done through the deconstruction of the free energy, and decoupling the different contributions. This allows for a fuller understanding of this binding process, thereby allowing the manipulation and optimisation of such systems for future studies and applications.

Examining the probability of a client binding by m beads, at various receptor densities, for both a flat and a pitted surface allows comparison of the rate at which the multiplicity of binding increases with respect to receptor density. To illustrate, the results for a pentavalent client are shown in Figure 5.13A, C and E and for a decavalent client in Figure 5.13B, D and F. There is almost complete suppression of binding by 1 or 2 beads to the pitted surface, followed by a significant increase in probability of observing $m > 2$ bonds between the client and surface. Indeed, the dramatic increase in binding by $m > 2$ beads is superior to that of bonding to a three-dimensional scaffold, Chapter 4, shown in Figure 4.5. Binding by low m is unfavourable at a pitted surface due to the large loss of configurational entropy of the polymer entering the pore. However, by forming many bonds polymer binding becomes favourable, due to the associated increase in internal energy magnitude.

The probability of m beads binding can be used to calculate the change in free energy on the formation of client-receptor bonds,

$$\Delta F_m = -kT \ln p_m/p_0. \tag{5.2}$$

At low receptor densities binding is never thermodynamically favourable (Figures 5.14A and B), but as the receptor density increases binding becomes thermodynamically favourable for the chains with short linkers (Figures 5.14C and D). For

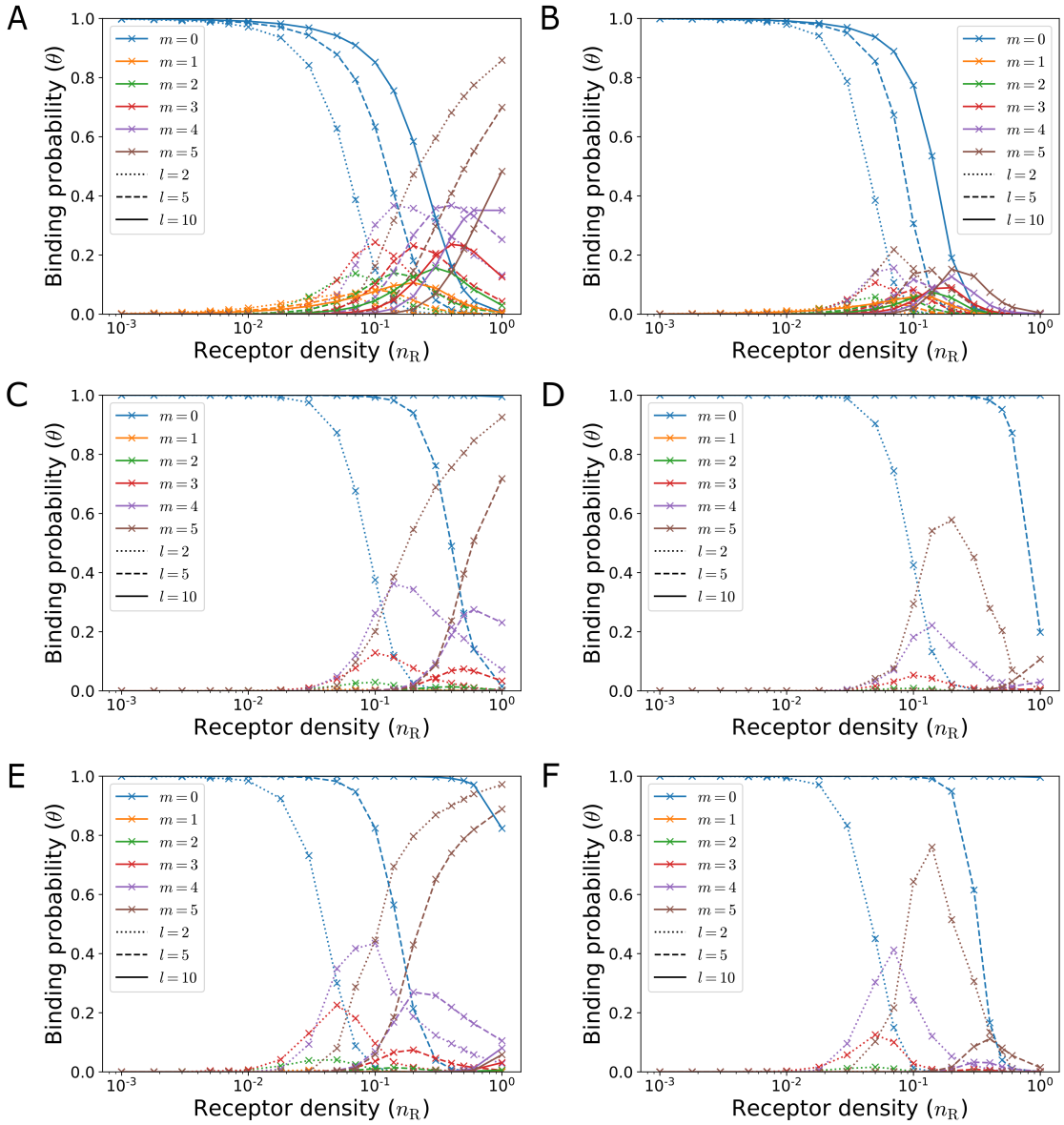


Figure 5.13: Probability of m beads being bound for a pentavalent client (A, C and E) and decavalent client (B, D and F) binding to a flat (A) and (B) and pitted (C - F) surface. The client-receptor interaction strengths are (A and B) $4kT$, (C and D) $5kT$ and (E and F) $6kT$. The dotted lines are for linker length 2, the dashed lines for linker length 5 and the solid lines for linker length 10.

both of these receptor densities the free energy penalty for forming a single bond is notably higher for the pitted surface compared to the flat surface, and acts as a free energy barrier for some of the shorter chains, where binding by large m becomes thermodynamically favourable. The results for the highest receptor density (Figures 5.14E and F) reveal a significantly higher free energy barrier to forming multiple bonds for the pitted surface compared the the flat surface. The barrier height for the clients binding to a pitted surface remains relatively unchanged on moving to different receptor densities. Therefore the barrier height can be controlled largely independently of the free energy of binding. This potentially allows for the kinetics of binding to be controlled. Up to this point, the work has largely focussed on the thermodynamics of binding, but by controlling the binding kinetics the "on-off" rate of polymer binding can be tuned. This binding control is likely to be utilised in biology, where prolonged binding is often required, such as in inhibiting enzymes [194]. It also has potential advantages in improving drug efficacy [194] and in the development of functional surfaces.

The larger free energy barrier to binding at a pitted surface explains the increased superselectivity observed at a pitted surface, but not the sorting behaviour. Instead, the range of free energy changes for the different polymer species is the key reason for enhanced sorting behaviour. The most obvious example of this is in Figure 5.14F, where binding to the surface is only strongly energetically favourable for clients with linker length 2. This demonstrates the key finding of this chapter, there is a significant free energy barrier to a client entering a pore, and this can only be overcome if the client is capable of forming many bonds. The entropic loss of a short chain moving from the unbound state to being bound by many bonds is much less than for a long chain, where such binding is not seen at low receptor densities and a low interaction strength. This is borne out in the deeper free energy minima of clients with short linkers compared to long linkers.

Comparison between the binding transitions and the complete sets of free energy curves indicates that whenever the change in free energy (for $m > 0$) becomes negative, the client is predominantly bound. Therefore these free energy curves can be beneficial in finding the receptor density at which θ approaches 1 but not the point at which the binding transition starts.

To uncover the cause of the greater free energy spread for clients binding at a pitted surface compared to a flat one, the free energy must be broken down further

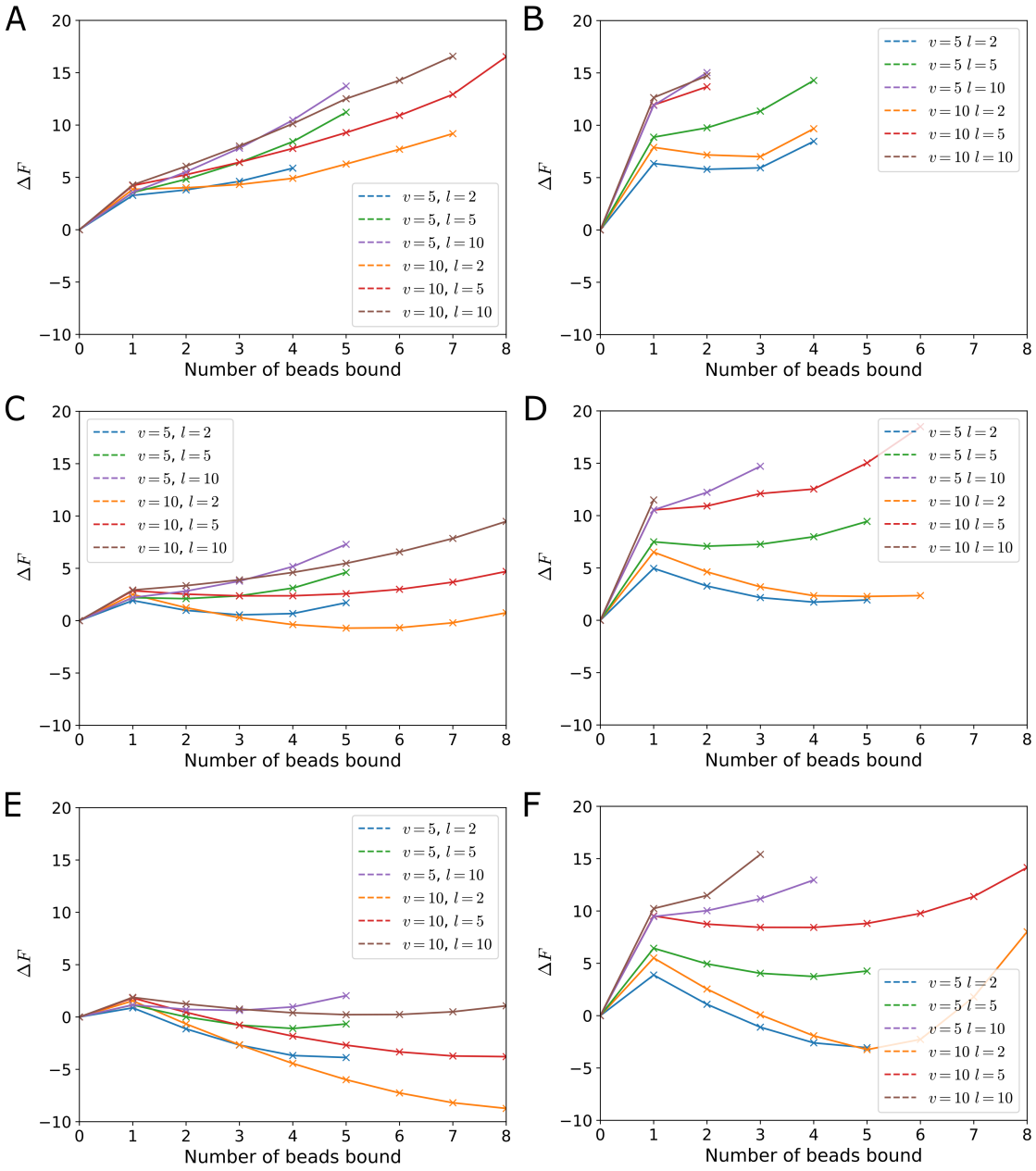


Figure 5.14: Free energy change on various clients, with hard linkers, going from the unbound state ($m = 0$) to being bound by m beads at receptor densities of (A and B) 0.018, (C and D) 0.07 and (E and F) 0.2. (A, C and E) are the results for flat surfaces and (B, D and F) are the results for binding at the pitted surface. The interaction strength between the client and receptor beads is $f = 5kT$.

into its constituent parts. The three components of the free energy are the entropy due to the number of ways of binding to the receptors, the chain entropy (configurational and translational) and the internal energy (total energy of the bonds formed). Firstly, the bond energy is removed from the free energy, given in Equation 5.2, to leave the change in system entropy,

$$\Delta S = k \ln \frac{p_m}{p_0} - m\frac{\epsilon}{T}. \tag{5.3}$$

In order to allow for fair comparison of entropy loss of chains binding at different receptor densities, we adjust the entropy to remove the contribution that arises from the number of ways that m bonds can be distributed over n receptors at the bottom of a pore. This entropic adjustment is calculated using Boltzmann's formula, $S = k \ln \Omega$, where Ω is the number of microstates for a client binding with m beads at a surface with n receptors. In the limit where there is no correlation between receptors and the client is capable of binding to all receptors, the number of microstates, $\Omega = n^m$.

This results in the final equation for the calculation of the change in chain entropy being:

$$\Delta S_{\text{chain}} = k \ln \frac{p_m}{p_0} - m\frac{\epsilon}{T} - km \ln n. \tag{5.4}$$

The entropic loss of a polymer chain moving from the unbound state to one where it is bound by three beads is shown in Figure 5.15. The results for the ideal (Figure 5.15A), and hard (Figure 5.15B) linkers are similar, with just a very slight increase in range of entropic loss across the different polymer types for the hard linkers. Note that the entropic loss for the chain is largely independent of receptor density, indicating that the different combinations of binding sites are indeed equally probable, as assumed in Equation 5.4.

As $\Delta F = \Delta U - T\Delta S$, and both ΔU and ΔS are negative, the free energy change is only negative when $|\Delta U| > |T\Delta S|$. We therefore see, in Figure 5.14, that the free energy rise, and by extension the entropy loss, on the client binding to the pitted surface is significantly higher than at the flat surface. To compare the relative magnitudes of the chain entropies in each system, the interaction strengths $(1/kT)$ must be accounted for, but even with such adjustments, the polymer entropy loss on binding is much greater at a pitted surface than a flat one.

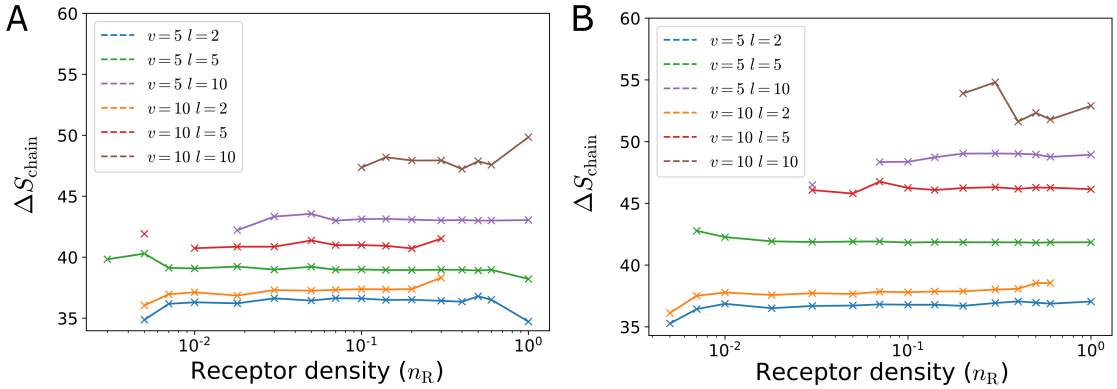


Figure 5.15: Entropy loss of polymer chain transitioning from the unbound state to being bound by 3 beads to the pitted surface. The polymers have (A) ideal and (B) hard linkers. The interaction strength between the client and receptor beads is $f = 5kT$.

Accounting for the entropy associated with receptor density shows that the change in chain configurational entropy on binding by m beads for a given receptor density is representative of all receptor densities. This entropy change upon binding at a receptor density $n_{\text{R}} = 0.1$ is shown Figure 5.16. Where lines are incomplete, this is a result of the clients never being bound by high m beads at an interaction strength of $5kT$. Again, this shows that clients with long hard linkers lose more entropy than those with long ideal linkers. The difference in entropy loss between short and long chains increases with the number of beads bound. This also explains why the free energy minima for the short chains are shallower than the long chains in Figure 5.14F. The difference in entropy loss for hard and ideal linkers allows for separation of polymers by their linker type — which in real systems translates as the solvation volume of the polymers.

Although we are able to compare the entropy loss for clients with ideal and hard linkers, we wish to compare the absolute difference in free energy of the chains, when they are bound to the surface. The thermodynamic cycle shown in Figure 5.17 illustrates this process we wish to quantify. The free energy change on binding is already known, therefore to discover the free energy difference between the hard and ideal linkers, we need to find ΔF_3, the change in free energy on going from ideal to hard linkers in the bound state. However, this is very challenging to calculate. Yet, calculation of ΔF_1 is possible through thermodynamic integration [151], therefore allowing us to complete the thermodynamic cycle to give ΔF_3. Thermodynamic integration involves the gradual transition of an unbound polymer's linkers from

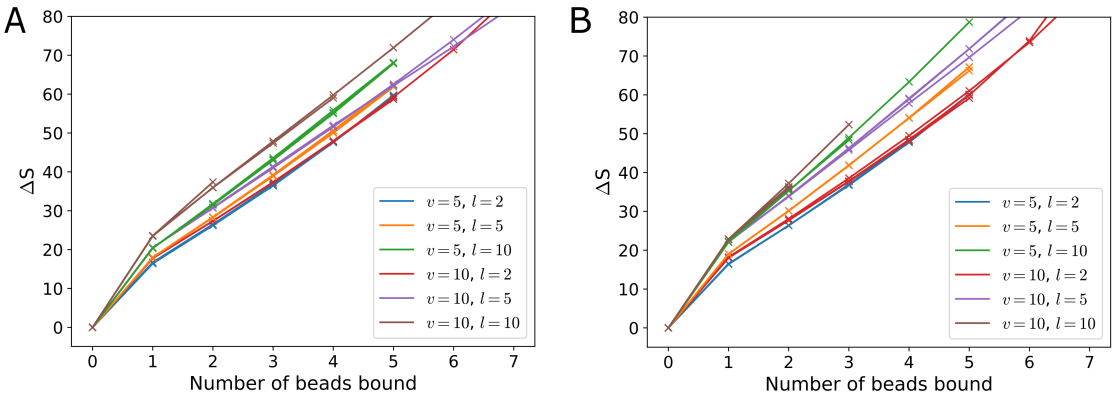


Figure 5.16: Configurational entropy change for clients with (A) ideal and (B) hard linkers binding by m beads to a pitted surface, after the entropy associated with the number of receptors has been removed. The interaction strength in this system is $5kT$ and the receptor density is 0.01.

ideal to hard, and a calculation of the associated free energy change.

Within the simulations required for the thermodynamic integration, hard linkers are simulated in the same way as ideal linkers, but with an overlap penalty, sufficiently strong that there are no bead overlaps. The energy penalty for a polymer in a configuration with n_{ov} linker bead overlaps, has the form

$$V = n_{\text{ov}}\chi\lambda^4, \tag{5.5}$$

where χ is the maximum overlap penalty (set to be sufficiently high that there are no overlaps when $\lambda = 1$), and λ is a coupling parameter over which the value $\partial V/\partial\lambda$ is integrated. The coupling parameter λ is raised to the power of 4 in Equation 5.5 to avoid the mean value of the derivative diverging at very low values of λ, where V is very large and fluctuates significantly. Different powers can be used but they must be sufficiently large. The free energy change on going from hard to ideal linkers is then given by,

$$\Delta F = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{5.6}$$

100 distinct simulations of the polymers at evenly spaced λ values in the range $0 \leq \lambda \leq 1$ are used for this thermodynamic integration. The integral, Equation 5.6, can be evaluated numerically using Simpson's rule to obtain the free energy difference on going from a polymer with hard linkers to ideal linkers. Throughout this process, binding beads are unable to overlap, unlike linker beads, which can overlap with both binding beads and other linker beads.

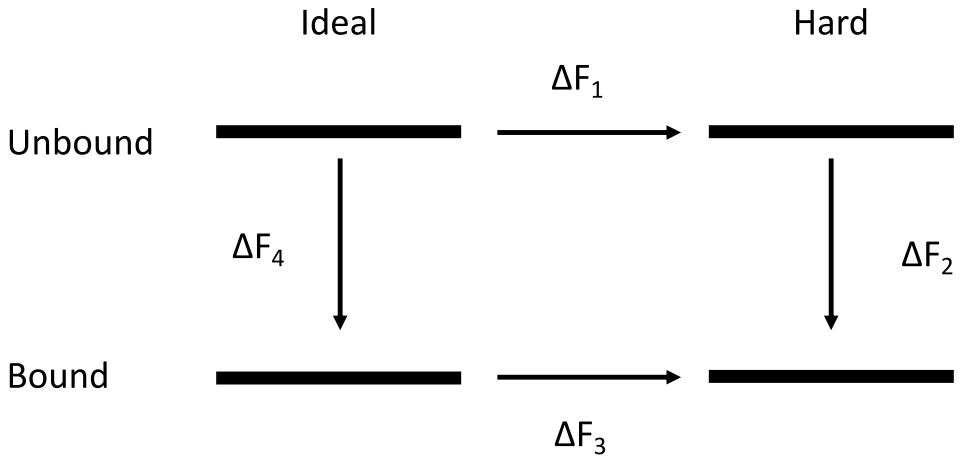


Figure 5.17: Thermodynamic cycle for the transition between ideal and hard linkers of a polymer. This cycle can be used to find the free energy difference between ideal and hard linkers, in both the bound and unbound states.

Table 5.1: Free energy change on going from unbound chains with ideal linkers to unbound chains with hard linkers.

Valency	Linker Length	$\Delta F_1/kT$(Unbound)
5	2	0.479(2)
10	2	1.309(7)
5	5	5.549(4)
10	5	9.07(2)
5	10	8.03(1)
10	10	19.61(3)

The free energy change of various polymers moving from ideal to hard linkers ΔF_1, calculated using the thermodynamic integration, are shown in Table 5.1. They indicate a significantly greater free energy for chains with ideal linkers, compared to hard linkers, for sufficiently long chains. Furthermore, as there is no internal energy of the system, it is possible to convert between the free energy change and the entropy change, as $\Delta S/k = -\Delta F/kT$ when $U = 0$. Therefore, the polymer loses entropy and gains free energy on moving from ideal linkers to hard linkers, with the values in Table 5.1 for $\Delta F_1/kT$ being equal to $-\Delta S_1/k$.

The values of ΔF_1 given in Table 5.1 allow for the calculation of ΔF_3, the free energy difference between polymers with hard and ideal linkers when bound. This is calculated using $\Delta F_3 = (\Delta F_1 + \Delta F_2) - \Delta F_4$, where Equation 5.3 is used to obtain

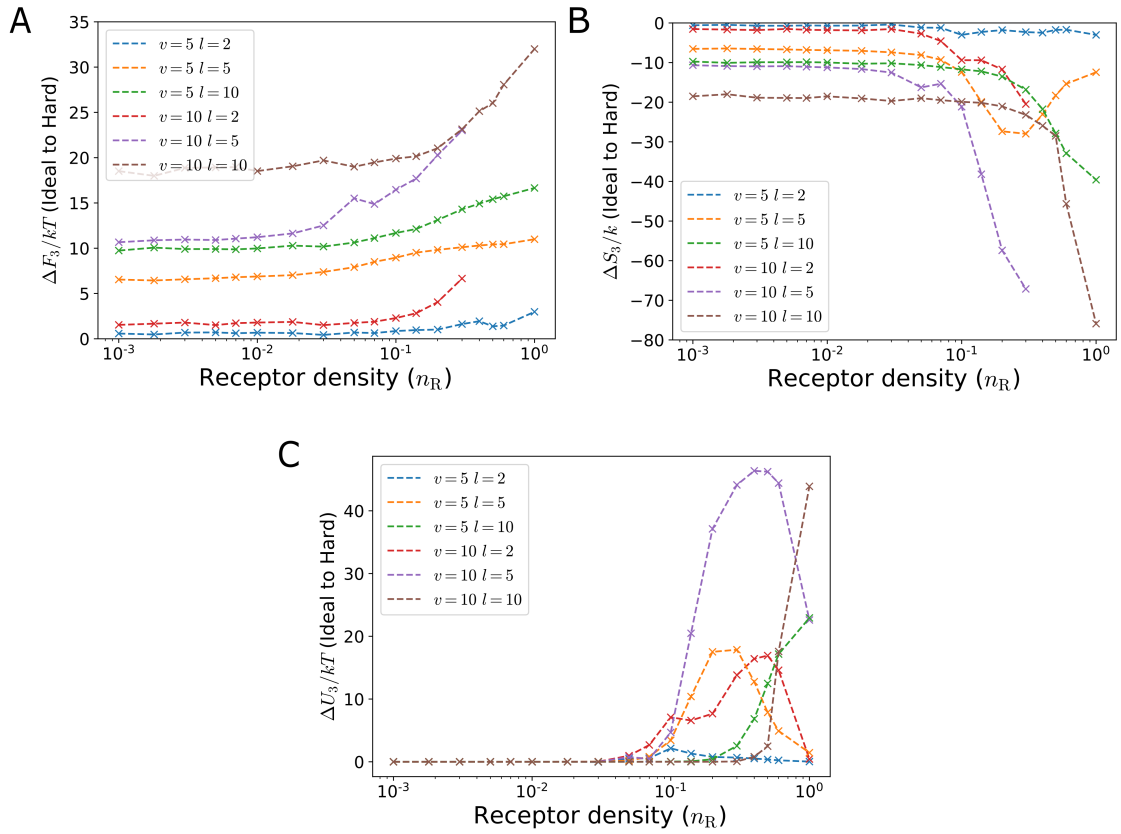


Figure 5.18: (A) The free energy, (B) entropy and (C) internal energy change on moving from ideal linkers to hard linkers when bound, for a range of receptor densities, where the interaction strength between the client beads and receptors is $5kT$.

ΔF_2 and ΔF_4. The difference between the free energy of the polymer on moving from ideal to hard linkers when bound are shown in Figure 5.18A alongside the corresponding plot for the difference in entropy in Figure 5.18B. The entropy ΔS_1 was calculated using Equation 5.3 and the values for $\Delta S_1/k$ are given in Table 5.1.

The free energy difference between the ideal and hard linkers, Figure 5.18A, is almost flat at low receptor densities, but as receptor density increases the hard linkers lose more free energy than the ideal linkers, resulting in a greater values of ΔF_3. This is pronounced for the long linkers, but very moderate for shorter linkers. Where the lines are incomplete, this indicates that the client is never unbound ($p_0 = 0$), therefore $\Delta F_3/kT$ cannot be calculated at these points. The entropy change on going from ideal to hard linkers $\Delta S_3/k$, remains roughly constant at low receptor densities before increasing in magnitude at high receptor densities, as shown in Figure 5.18B. Hard linkers suffer much greater configurational entropy loss on binding to pitted surfaces than ideal linkers. The reduction in entropy difference at very high receptor densities for the pentavalent client with linker length 5 is a

result of both the hard and ideal linkers losing significant configurational entropy when bound by a large number of beads. At this point the entropic differences between the two linker types are similar to those when the chains are unbound. This minimum in the entropic difference between hard and ideal linkers is also observed for polymer species other than that of a pentavalent client with linker length 5 in systems at lower temperatures. The decrease in $\Delta S_3/k$ is larger than the increase $\Delta F_3/kT$, therefore the differences in the internal energy of the polymers must be similar to those of $\Delta S_3/k$, in order to obtain the modest $\Delta F_3/kT$ observed in Figure 5.18A.

The difference in internal energy is shown in Figure 5.18C, where it can be seen that the clients with ideal linkers have greater internal energy than those with hard linkers (due to their increased binding propensity). The higher entropy when bound for long ideal linkers, compared to hard linkers, facilitates binding of such clients to the surface at receptor densities which clients with hard linkers cannot. The difference in internal energy reaches a peak when polymers with ideal linkers are bound more extensively than those with hard linkers. The polymers with hard linkers then start to 'catch-up' with the number of bonds formed by ideally linked polymers thereby reducing the difference in internal energy. This peak in $\Delta U_3/kT$ and the sharp decline in $\Delta S_3/k$ result in the relatively monotonic curves for $\Delta F_3/kT$. Therefore, the large loss in configurational entropy of clients with hard linkers binding by many beads to a pitted surface is crucial to the sorting behaviour of the polymers. The much larger entropy loss for long chains on going from ideal to hard linkers compared to short chains reveals why the sorting of hard linker clients is superior to those with ideal linkers.

5.4 Conclusions and discussion

Superselectivity is heavily dependent on the surface to which a client species binds. Firstly, the excluded volume of a solid surface acts to significantly increase the superselectivity of binding. Further to this, we have shown that the superselectivity is enhanced on moving from a flat surface to a pitted surface (with receptors on the base of the pits). This is due to the loss of polymer configurational entropy on entering the confined volume of pore. This large entropic penalty on entering the pore can be exploited to control the binding of polymers to a surface, either through

tuning interaction strength or receptor density. This tuning is most sensitive for polymers with hard linkers; the configurational entropy loss of polymers on binding increases at a greater rate with respect to increasing linker length for hard linkers, compared to ideal. Therefore the free energy penalty for polymers with long hard linkers prevents polymers binding to a surface at temperatures and receptor densities where polymers with short or ideal linkers can bind. Another key finding is that the kinetics of binding appears to be controllable by the pore geometry, independent of the polymer affinity.

Within biology, the diversity and complexity of surface structures is significant [195]. These interfaces have a vast range of compositions and topographies, which define how they interact with their surroundings [196]. It is likely that, within nature, examples of porous or pitted surfaces geometries which control the binding of species within their environment are already present. For instance, the heirarchical nature of biological structures, such as diatoms [197], may utilise the confinement of pores to selectively control binding, through configurational entropic penalties for pore entry. The pore geometries and distributions are finely controlled [198, 199], and are thus a promising avenue for investigating whether nature is already utilising superselective binding to control recruitment.

Moreover, our findings should inform the design of synthetic functional surfaces for supramolecular self-assembly, or for use within diagnostic/detection devices [200–202]. An interesting potential application is in polymer or protein separation or chromatography. Chromatography is a vast field [203], yet pitted surfaces have shown capacity to sort based on a combination of polymer properties, such as linker length, linker type, interaction strength and valency. This gives hope for a versatile analytical technique, which can sort based on a range of parameters, rather than a single one, as most current techniques are limited to [204].

Metal organic frameworks (MOFs) are another area whereby the polymer configurational entropy loss on entering a confined volume could be used to control binding. MOFs are highly porous, ordered, crystalline structures, which have very high surface area to volume ratios [205]. Due to the plethora of components from which MOFs can be created, to date there have been over 90000 different MOFs reported, these structures can be finely tuned to deliver the geometries and funtionality required [206]. They have shown promise for use in gas separation, sensors and catalysis [205, 207–209]. Many of these MOFs have pores on the order of angstroms

to nanometers [210], which is too small for the polymers we study. However, there has been significant promise in the development of hierarchically porous MOFs which can have pores of diameters of tens of nanometers [211]. These larger pores would be very suitable for the biopolymers which inspired this work. Therefore these structures could be tuned to allow for selective recruitment of polymeric species, which is important in controlling catalyst activity and performace, increasing their precision of the self-assembly of supramolecular species and vital for effective sensor function [212].

Our studies were limited to a single geometry of surface, therefore future work would involve the study of a range of pore geometries, varying pore depth, width and separation to discover how these alter the binding behaviour. The pore geometry is likely to have a significant effect on the free energy curves, with increasing pore depth increasing the entropic loss of a polymer binding, up until the limit where the polymer length is short on the length-scale of the pore. Conversely, widening the pore is likely to reduce the free energy barrier to binding, and reduce the disparity in entropy loss for client of different lengths upon binding. This would impair the sorting properties of a surface. Finally, experimental studies into comparable systems would be fascinating, and would verify the sorting behaviour of polymers at a pitted surface, and facilitate a wider parameter sweep. The first step in this process would be to manufacture the pitted surface (at the nano and microscale) [188, 189] and decorate the base of the pores [190]. Systematic studies probing both the degree of superselectivity and the range in binding transitions would then be compared to the trends we have uncovered.

Chapter 6

Conclusions and future work

6.1 Conclusions

In this section we shall highlight several common themes observed across the different systems studied in this thesis.

Firstly, this work highlights the pivotal role that multivalency can play in controlling the binding and phase behaviour of polymers. Multivalency can allow for either cooperative binding or its suppression, which in turn leads to numerous interesting phenomena. In Chapter 3, we have shown that valency strongly impacts the propensity for phase separation in a system of two associative polymer species, due to the cooperative binding of the polymer beads driving the phase separation. In Chapter 4 this multivalency was shown as a potential mechanism for the fine compositional control of binding of polymers to a 3D host. We observed superselective behaviour where the sharp binding transition is enabled by suppression of binding before the transition and by cooperativity after the transition. In Chapter 5 we presented results indicating that the reduced binding of polymeric clients at the base of pores on a pitted surface allows for the sorting of polymers based on their length and excluded volume.

From our studies, it is clear that the properties of the polymer linkers, including both their length and type, play a major role. Here, we have utilised three discrete linker types (hard, ideal and implicit), as well as a wide range of linker lengths, to represent a variety of polymeric species within biological systems. The impact of linker type is a key message of Chapter 3, where the three different linkers result in distinct phase separation behaviour, with the hard linkers diverging consider-

ably from the other two. This indicates that increased solvation volume suppresses phase separation. Moreover, as shown in Chapter 5, the natural radius of gyration of polymers with different linker lengths and types result in significantly different behaviour where there is an aspect of confinement. The impact of linker properties on the polymer behaviour helps explain the versatility of biological systems, where desired structures and behaviours can be controlled by the cellular constituents. This designability comes in the form of the intrinsic properties of the interacting biomolecules (proteins and RNAs), including their valency, rigidity/solvation volume, interaction strength, concentrations and linker lengths.

Many of the aforementioned behaviours can be explained by the different entropic considerations of the systems. The main entropic contributions are the chain configurational entropy, translational entropy, bond orientational entropy, and binding combinatorics. These often act in competition, both with each other and with the enthalpy of binding (divided by temperature), in order to determine the final system behaviour. For instance, in the phase separation work, there is a continual competition between chains 'pairing up' and forming a single large cluster. Both of these 'states' minimise the internal energy of the system, but whereas the paired up chains have very low bond degeneracy (entropy due to combinatorics of bonds) and high translational entropy, the opposite is true for the large clusters. From our work on polymers binding to a host, we also observe a competition between gaining internal energy through bond formation and losing translational and configurational entropy upon binding. This is of paramount importance in systems wherein significant configurational entropy can be maintained during binding, such as for polymers binding in wide pores or at flat surfaces, where bonding combinatorics or preserving much of the linker entropy increase the favourability of binding. Moreover, we have shown that crowders can be introduced to a 3D host in order to enhance the superselectivity of client binding. This is achieved by the crowders reducing the free volume in the 3D host, thereby increasing the entropic penalty of the client entering the structure.

6.2 Future work

There are numerous directions of study which could lead on from this work. In this section we appraise the scope of, and techniques used in, our studies, and

propose three of the most promising avenues of future research.

6.2.1 Polymer sorting

One such area of further research would be to do a wider survey of the polymer sorting surfaces discussed in Chapter 5. Our study involved the investigation of one corrugated surface, but by studying polymer binding at porous surfaces with different pore widths, depths and separations, relationships between the binding probabilities of various polymers and pore geometries could be obtained. Superselectivity is expected to increase with deepening and narrowing the pores, within limits, providing that there is a sufficient number of binding sites on the pore base to allow for the formation of multiple bonds. Therefore, similar studies could be undertaken to maximise the superselectivity of binding, and to find the corresponding upper limit of this phenomenon. Furthermore, determining a relationship between the entropy loss of a polymer on binding and the polymer properties and the pore geometries would be informative in the thermodynamic studies of polymer binding. It is likely that such a relationship would be complex, but principal component analysis could be utilised to focus on the most important factors. This would give a deeper understanding of what causes the large entropy loss on a long polymer binding at a corrugated surface; whether it is the geometrical confinement or, for example, whether there is a similar loss in systems where polymers can bind to receptors in multiple pores simultaneously. This situation — where a single polymer straddles pores — was not covered in Chapter 5, but is likely to further complicate the relationship between binding probability and surface geometry.

We also hope that our study will inspire experimental investigations into sorting polymers by exploiting pitted surface geometries. If an experimental proof of concept of an analytical device was achieved, predictive modelling would help with the rapid selection of a surface in order to extract the desired polymer, or protein, species from a given solution.

6.2.2 Off-lattice simulations

MC simulations on cubic lattices are highly effective for the simulation of large, complex systems — the discretisation of space by the lattice allows for efficient study [123]. Our coarse-grained simulations allow us to distil the complex biological

systems of interest down to simple building blocks, from which we can determine the fundamental causes of behaviours exhibited in nature. Therefore, the general results and trends obtained from our simulations of coarse-grained polymers are applicable to a wide range of polymeric species, rather than a single specific one. Yet, the approximations used in studying such a system come at the cost of some of the fine details, especially on short length scales, such as the interactions between beads. Lattices can also result in undesirable artefacts, for instance, very short linkers can result in un-representative distributions of binding-bead separations due to the discretisation of positions on a lattice. These considerations were taken into account in the designing of our system, where it was deduced that the impact of lattice discretisation on the overall phase behaviour would be acceptable for the investigations pursued in this thesis.

Moving the study of coarse-grained polymers (beads on a string) off-lattice is an attractive direction to increase the model complexity in the future [213]. This improves the resolution of the positional and orientational degrees of freedom, allows for the excluded volume effects to be accounted for in a more precise manner [214] and enables the transition from square well potential energy curves to more representative potential energy curves [215], incorporating contributions from hydrophobic, polar and electrostatic interactions [216]. Further to this, accounting accurately for polymer rigidity using cubic lattice polymers is not possible, so off-lattice simulations would be necessary in modelling the impact of polymer rigidity on phase separation [217]. Additionally, studies of the dynamics of LLPS and membraneless organelles would be possible off-lattice, whereas a dynamic interpretation of lattice models is ill defined [218].

An appealing off-lattice, coarse-grained model is one that combines patchy particles [219] and beads-on-a-string [121], such as that utilised by Espinosa *et al.* to model LLPS [1]. Using this model they were able to study systems with many components, and draw conclusions on the impact of connectivity on the physical properties of droplets formed. At this point, we must emphasise that the computational expense of moving off-lattice is notable, but does have significant potential in this field of work. Other such simulations successfully used to model phase separation include un-connected patchy particles [5], and Brownian simulations of independent beads [164].

It is important to recognise that the systems we study are intrinsically multi-

scale. Therefore, full insight into these systems is only possible through multi-scale modelling. One strategy is to run simulations at different resolutions whereby detailed findings from high resolution models of particular biological species, such as interactions and commonly occupied configurations, are then fed into coarse-grained models, in order to model large sytems of biopolymers as accurately as possible but at a much lower computational cost than fully atomistic simulations. This is known as sequential multiscale modelling [220]. High resolution techniques such as atomistic and quantum mechanical simulations can be used to obtain accurate interactions present between the proteins [214, 221]. Additionally, there is a wealth of protein secondary and tertiary structures, from both simulation [222, 223] and experiments [224, 225], which can be included within multiscale models [220]. Capturing the effect of atomistic changes to polymers accurately is important, as membraneless organelle formation and behaviour are very closely linked to post translational modifications, where the chemical structure of the polymer is altered [34]. Additionally, alterations of the chemical sequence of protiens can significantly affect the protein structure and behaviour [214].

Further improvement in sampling can be achieved by running simulations where the resolution can vary within each simulation, known as concurrent multiscale modelling [220]. One way to do this is via the adaptive resolution simulation scheme, AdResS, where the degrees of freedom of different parts of the simulation vary, depending on the resolution required [226, 227]. The promise of these techniques is notable, and will allow system-specific findings, for instance, the change in behaviour caused by polymer mutations or modifications, or how changing the protein sequencing can alter both LLPS and polymer recruitment to a 3D host [22, 34, 179, 179, 180].

6.2.3 Membraneless organelles and LLPS

In spite of the increased research into LLPS and membraneless organelles, there are a number of important questions that remain, especially on the relations between the organelle structure and function. From a biological perspective, their inextricable link to neurodegenerative disease increases the impetus to better understand how these organelles evolve with time, including the mechanisms by which fibril formation occurs [18] and the effects of mutations on their formation, function and constitution [31, 67, 228]. From a physics perspective, there are open

questions on the roles of active ATP-driven processes in suppressing phase separation [17, 229]. From a materials science perspective, there is great potential within synthetic chemistry and bioengineering to utilise the properties and functions of biomolecular condensates, for example in drug delivery [184] and tissue engineering [183].

In our work, the role of linkers has also been simplified significantly, including assuming they are inert. Yet, linkers are known to be able to interact with the binding domains and contribute to phase separation [69]. Very brief exploratory work was done to study how an isotropically attractive 'tail' connected to our multivalent polymers would change the phase separation, but these isotropic interactions were so abundant that they drove phase separation at much lower interaction strengths than those required for the phase separation of polymers with only modular binding sites. A more structured approach to such linker interactions would be useful in further understanding these complex biological condensates.

Another potentially fruitful direction is to study the competition between binding species. The systems investigated here were simple, with only two species. In contrast, membraneless organelles have been reported to contain hundreds of species [181, 182]. While it is not yet practical to study such problem in its full complexity, a good step forward is to analyse a system of two associating polymer systems interacting in the presence of a competitor species. A particularly interesting angle of research would be to titrate the interaction strength between the polymers and the competitors. This will shed light into how competitor species may alter the phase separation process and the structure of the clusters formed in Chapter 3, as well as how these competitor species influence the compositional control discussed in Chapter 4. For example, the study by Espinosa *et al.* [1] found that low-valency clients reduce the connectivity and stability of the clusters formed. This is supported by other computational studies of high-valency polymers and low-valency clients, where the polymers provide the stability for the cluster [6], and the clients are rapidly exchanged and occupy more dynamic regions on the periphery of the droplet [5]. Additionally, simulating the binding of a multivalent client at a 3D host, where there is a competitor species present, would indicate the robustness of the superselectivity observed. Interestingly, studies of such competition at a flat surface of receptors have shown that selectivity can even be enhanced by the presence of competitors [107].

Modelling polymer binding and LLPS using cubic lattice polymers has uncovered multiple significant findings, especially with regards superselectivity and the role of linkers. This works lays the ground-work for more complex, and detailed, system-specific studies of LLPS and superselective binding.

Bibliography

[1] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components, Proceedings of the National Academy of Sciences **117**(24), 13238–13247 (2020).

[2] J. R. Wheeler, T. Matheny, S. Jain, R. Abrisch, and R. Parker, Distinct stages in stress granule assembly and disassembly, Elife **5**, e18413 (2016).

[3] S. Strome and W. B. Wood, Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of Caenorhabditis elegans, Proceedings of the National Academy of Sciences **79**(5), 1558–1562 (1982).

[4] D. S. Protter and R. Parker, Principles and properties of stress granules, Trends in Cell Biology **26**(9), 668–679 (2016).

[5] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, Valency and binding affinity variations can regulate the multilayered organization of protein condensates with many components, Biomolecules **11**, 278 (2021).

[6] J. A. Joseph, J. R. Espinosa, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, Thermodynamics and kinetics of phase separation of protein-RNA mixtures by a minimal model, Biophysical Journal **120**(7), 1219–1230 (2021).

[7] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, et al., Phase transitions in the assembly of multi-valent signaling proteins, Nature **483**, 336–340 (2012).

[8] Y. Shin and C. P. Brangwynne, Liquid phase condensation in cell physiology and disease, Science **357**, eaaf4382 (2017).

[9] E. Voronina, The diverse functions of germline P-granules in Caenorhabditis elegans, Molecular Reproduction and Development **80**(8), 624–631 (2013).

[10] J. C. Schwartz, E. R. Podell, S. S. Han, J. D. Berry, K. C. Eggan, and T. R. Cech, FUS is sequestered in nuclear aggregates in ALS patient fibroblasts, Molecular Biology of the Cell **25**(17), 2571–2578 (2014).

[11] V. Verdile, E. De Paola, and M. P. Paronetto, Aberrant phase transitions: Side effects and novel therapeutic strategies in human disease, Frontiers in Genetics **10**, 173 (2019).

[12] T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, and A. J. Baldwin, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles, Molecular Cell **57**(5), 936–947 (2015).

[13] S. Kroschwald and S. Alberti, Gel or die: Phase separation as a survival strategy, Cell **168**(6), 947–948 (2017).

[14] A. Rojas-Fernandez, A. Plechanovová, N. Hattersley, E. Jaffray, M. H. Tatham, and R. T. Hay, SUMO chain-induced dimerization activates RNF4, Molecular Cell **53**(6), 880–892 (2014).

[15] S. Alberti and D. Dormann, Liquid–liquid phase separation in disease, Annual Review of Genetics **53**(1), 171–194 (2019).

[16] C. Mathieu, R. V. Pappu, and J. P. Taylor, Beyond aggregation: Pathological phase transitions in neurodegenerative disease, Science **370**(6512), 56–60 (2020).

[17] S. Jain, J. R. Wheeler, R. W. Walters, A. Agrawal, A. Barsic, and R. Parker, ATPase-modulated stress granules contain a diverse proteome and substructure, Cell **164**(3), 487–498 (2016).

[18] S. Kroschwald, S. Maharana, D. Mateju, L. Malinovska, E. Nüske, I. Poser, D. Richter, and S. Alberti, Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules, Elife **4**, e06807 (2015).

[19] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, and M. K. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, Molecular Cell **63**(1), 72–85 (2016).

[20] S. Boeynaems, E. Bogaert, D. Kovacs, A. Konijnenberg, E. Timmerman, A. Volkov, M. Guharoy, M. De Decker, T. Jaspers, V. H. Ryan, et al., Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics, Molecular Cell **65**(6), 1044–1055 (2017).

[21] M. Mammen, S.-K. Choi, and G. M. Whitesides, Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors, Angewandte Chemie International Edition **37**(20), 2754–2794 (1998).

[22] C. P. Brangwynne, P. Tompa, and R. V. Pappu, Polymer physics of intracellular phase transitions, Nature Physics **11**(11), 899–904 (2015).

[23] X. Y. Ling, D. N. Reinhoudt, and J. Huskens, Reversible attachment of nanostructures at molecular printboards through supramolecular glue, Chemistry of Materials **20**(11), 3574–3578 (2008).

[24] D. Bracha, M. T. Walls, M.-T. Wei, L. Zhu, M. Kurian, J. L. Avalos, J. E. Toettcher, and C. P. Brangwynne, Mapping local and global liquid phase behavior in living cells using photo-oligomerizable seeds, Cell **175**(6), 1467–1480 (2018).

[25] O. D. King, A. D. Gitler, and J. Shorter, The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease, Brain Research **1462**, 61–80 (2012).

[26] S. Elbaum-Garfinkle, Y. Kim, K. Szczepaniak, C. C.-H. Chen, C. R. Eckmann, S. Myong, and C. P. Brangwynne, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics, Proceedings of the National Academy of Sciences **112**(23), 7189–7194 (2015).

[27] M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles, Nature Chemistry **9**(11), 1118–1125 (2017).

[28] E. Zumbro and A. Alexander-Katz, Polymer stiffness regulates multivalent binding and liquid-liquid phase separation, Biophysical Journal **119**(9), 1849–1864 (2020).

[29] A. Khong, C. H. Kerr, C. H. Yeung, K. Keatings, A. Nayak, D. W. Allan, and E. Jan, Disruption of stress granule formation by the multifunctional cricket paralysis virus 1A protein, Journal of Virology **91**, e01779–16 (2017).

[30] V. Belzil, P. Valdmanis, P. Dion, H. Daoud, E. Kabashi, A. Noreau, J. Gauthier, P. Hince, A. Desjarlais, J.-P. Bouchard, et al., Mutations in FUS cause FALS and SALS in French and French Canadian populations, Neurology **73**(15), 1176–1179 (2009).

[31] A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, et al., A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation, Cell **162**(5), 1066–1077 (2015).

[32] S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen, Compositional control of phase-separated cellular bodies, Cell **166**(3), 651–663 (2016).

[33] W. T. Snead and A. S. Gladfelter, The control centers of biomolecular phase separation: How membrane surfaces, PTMs, and active processes regulate condensation, Molecular Cell **76**(2), 295–305 (2019).

[34] J. A. Ditlev, L. B. Case, and M. K. Rosen, Who's in and who's out—Compositional control of biomolecular condensates, Journal of Molecular Biology **430**(23), 4666–4684 (2018).

[35] P. St George-Hyslop, J. Q. Lin, A. Miyashita, E. C. Phillips, S. Qamar, S. J. Randle, and G. 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).

[36] M. Saito, D. Hess, J. Eglinger, A. W. Fritsch, M. Kreysing, B. T. Weinert, C. Choudhary, and P. Matthias, Acetylation of intrinsically disordered regions regulates phase separation, Nature Chemical Biology **15**(1), 51–61 (2019).

[37] J. Söding, D. Zwicker, S. Sohrabi-Jahromi, M. Boehning, and J. Kirschbaum, Mechanisms for active regulation of biomolecular condensates, Trends in Cell Biology **30**(1), 4–14 (2020).

[38] S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V. Pappu, and A. D. Gitler, Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties, Proceedings of the National Academy of Sciences **116**(16), 7889–7898 (2019).

[39] M. Dundr, M. D. Hebert, T. S. Karpova, D. Stanek, H. Xu, K. B. Shpargel, U. T. Meier, K. M. Neugebauer, A. G. Matera, and T. Misteli, In vivo kinetics of Cajal body components, Journal of Cell Biology **164**(6), 831–842 (2004).

[40] S. A. Fromm, J. Kamenz, E. R. Nöldeke, A. Neu, G. Zocher, and R. Sprangers, In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery, Angewandte Chemie International Edition **53**(28), 7354–7359 (2014).

[41] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry, Nature Reviews: Molecular Cell Biology **18**(5), 285–298 (2017).

[42] K. K. Nakashima, M. A. Vibhute, and E. Spruijt, Biomolecular chemistry in liquid phase separated compartments, Frontiers in Molecular Biosciences **6**, 21 (2019).

[43] A. A. Hyman and K. Simons, Beyond oil and water—Phase transitions in cells, Science **337**(6098), 1047–1049 (2012).

[44] R. W. Style, T. Sai, N. Fanelli, M. Ijavi, K. Smith-Mannschott, Q. Xu, L. A. Wilen, and E. R. Dufresne, Liquid-liquid phase separation in an elastic network, Physical Review X **8**, 011028 (2018).

[45] K. A. Rosowski, T. Sai, E. Vidal-Henriquez, D. Zwicker, R. W. Style, and E. R. Dufresne, Elastic ripening and inhibition of liquid–liquid phase separation, Nature Physics **16**(4), 422–425 (2020).

[46] Y. Shin, Y.-C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, and C. P. Brangwynne, Liquid nuclear condensates mechanically sense and restructure the genome, Cell **175**(6), 1481–1491 (2018).

[47] J. D. Wurtz and C. F. Lee, Chemical-reaction-controlled phase separated drops: formation, size selection, and coarsening, Physical Review Letters **120**, 078102 (2018).

[48] N. Kedersha, G. Stoecklin, M. Ayodele, P. Yacono, J. Lykke-Andersen, M. J. Fritzler, D. Scheuner, R. J. Kaufman, D. E. Golan, and P. Anderson, Stress granules and processing bodies are dynamically linked sites of mRNP remodeling, Journal of Cell Biology **169**(6), 871–884 (2005).

[49] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Germline P granules are liquid droplets that localize by controlled dissolution/condensation, Science **324**(5935), 1729–1732 (2009).

[50] N. Kedersha, M. R. Cho, W. Li, P. W. Yacono, S. Chen, N. Gilks, D. E. Golan, and P. Anderson, Dynamic shuttling of Tia-1 accompanies the recruitment of mRNA to mammalian stress granules, Journal of Cell Biology **151**(6), 1257–1268 (2000).

[51] K. E. Handwerger, J. A. Cordero, and J. G. Gall, Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure, Molecular Biology of the Cell **16**(1), 202–211 (2005).

[52] C. P. Brangwynne, T. J. Mitchison, and A. A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes, Proceedings of the National Academy of Sciences **108**(11), 4334–4339 (2011).

[53] Y. Lin, D. S. Protter, M. K. Rosen, and R. Parker, Formation and maturation of phase-separated liquid droplets by RNA-binding proteins, Molecular Cell **60**(2), 208–219 (2015).

[54] D. Bracha, M. T. Walls, and C. P. Brangwynne, Probing and engineering liquid-phase organelles, Nature Biotechnology **37**(12), 1435–1445 (2019).

[55] S. Alberti, A. Gladfelter, and T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates, Cell **176**(3), 419–434 (2019).

[56] N. Asherie, J. Pande, A. Lomakin, O. Ogun, S. R. Hanson, J. B. Smith, and G. B. Benedek, Oligomerization and phase separation in globular protein solutions, Biophysical Chemistry **75**(3), 213–227 (1998).

[57] T. Pawson and P. Nash, Assembly of cell regulatory systems through protein interaction domains, Science **300**(5618), 445–452 (2003).

[58] C. Sisu, A. J. Baron, H. M. Branderhorst, S. D. Connell, C. A. G. M. Weijers, R. de Vries, E. D. Hayes, A. V. Pukin, M. Gilbert, R. J. Pieters, H. Zuilhof, G. M. Visser, and W. B. Turnbull, The influence of ligand valency on aggregation mechanisms for inhibiting bacterial toxins, ChemBioChem **10**(2), 329–337 (2009).

[59] B. J. Tombling, C. Lammi, C. Bollati, A. Anoldi, D. J. Craik, and C. K. Wang, Increased valency improves inhibitory activity of peptides targeting Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9), ChemBioChem **22**(12), 2154–2160 (2021).

[60] H. Zuilhof, Fighting cholera one-on-one: The development and efficacy of multivalent cholera-toxin-binding molecules, Accounts of Chemical Research **49**, 274–285 (2016).

[61] T. R. Branson and W. B. Turnbull, Bacterial toxin inhibitors based on multivalent scaffolds, Chemical Society Reviews **42**(11), 4613–4622 (2013).

[62] J. Jin, X. Xie, C. Chen, J. Park, C. Stark, D. James, M. Olhovsky, R. Linding, Y. Mao, and T. Pawson, Eukaryotic protein domains as functional units of cellular evolution, Science Signaling **2**, ra76 (2009).

[63] B. S. Schuster, E. H. Reed, R. Parthasarathy, C. N. Jahnke, R. M. Caldwell, J. G. Bermudez, H. Ramage, M. C. Good, and D. A. Hammer, Controllable protein phase separation and modular recruitment to form responsive membraneless organelles, Nature Communications **9**(1), 1–12 (2018).

[64] D. M. Mitrea and R. W. Kriwacki, Phase separation in biology; functional organization of a higher order, Cell Communication and Signaling **14**, 1 (2016).

[65] P. Yang, C. Mathieu, R.-M. Kolaitis, P. Zhang, J. Messing, U. Yurtsever, Z. Yang, J. Wu, Y. Li, Q. Pan, J. Yu, E. W. Martin, T. Mittag, H. J. Kim, and J. P. Taylor, G3BP1 is a tunable switch that triggers phase separation to assemble stress granules, Cell **181**(2), 325–345 (2020).

[66] J. Guillén-Boixet, A. Kopach, A. S. Holehouse, S. Wittmann, M. Jahnel, R. Schlüssler, K. Kim, I. R. Trussina, J. Wang, D. Mateju, et al., RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation, Cell **181**(2), 346–361 (2020).

[67] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, Phase separation by low complexity domains promotes stress granule assembly and drives Pathological Fibrillization, Cell **163**(1), 123–133 (2015).

[68] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins, Elife **6**, e30294 (2017).

[69] S. Banjade, Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu, and M. K. Rosen, Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck, Proceedings of the National Academy of Sciences **112**(47), 6426–6435 (2015).

[70] C. V. Falkenberg, M. L. Blinov, and L. M. Loew, Pleomorphic ensembles: formation of large clusters composed of weakly interacting multivalent molecules, Biophysical Journal **105**(11), 2451–2460 (2013).

[71] C. V. Falkenberg, J. H. Carson, and M. L. Blinov, Multivalent molecules as modulators of RNA granule size and composition, Biophysical Journal **113**(2), 235–245 (2017).

[72] V. Nguemaha and H. Zhou, Liquid-liquid phase separation of patchy particles illuminates diverse effects of regulatory components on protein droplet formation, Scientific Reports **8**(1), 1–11 (2018).

[73] D. W. Sanders, N. Kedersha, D. S. Lee, A. R. Strom, V. Drake, J. A. Riback, D. Bracha, J. M. Eeftens, A. Iwanicki, A. Wang, et al., Competing protein-RNA interaction networks control multiphase intracellular organization, Cell **181**(2), 306–324 (2020).

[74] J. Huskens, L. J. Prins, R. Haag, and B. J. Ravoo, *Multivalency: Concepts, research and applications*, John Wiley & Sons, 2018.

[75] M. Ludden, A. Mulder, R. Tampé, D. Reinhoudt, and J. Huskens, Molecular printboards as a general platform for protein immobilization: A supramolecular solution to nonspecific adsorption, Angewandte Chemie International Edition **46**(22), 4104–4107 (2007).

[76] L. G. Suárez, W. Verboom, and J. Huskens, Cyclodextrin-based supramolecular nanoparticles stabilized by balancing attractive host–guest and repulsive electrostatic interactions, Chemical Communications **50**(55), 7280–7282 (2014).

[77] R. Mejia-Ariza, L. Graña-Suárez, W. Verboom, and J. Huskens, Cyclodextrin-based supramolecular nanoparticles for biomedical applications, Journal of Materials Chemistry B **5**(1), 36–52 (2017).

[78] L. L. Kiessling, J. E. Gestwicki, and L. E. Strong, Synthetic multivalent ligands as probes of signal transduction, Angewandte Chemie International Edition **45**(15), 2348–2368 (2006).

[79] C. T. Varner, T. Rosen, J. T. Martin, and R. S. Kane, Recent advances in engineering polyvalent biological interactions, Biomacromolecules **16**(1), 43–55 (2015).

[80] N. R. Bennett, D. B. Zwick, A. H. Courtney, and L. L. Kiessling, Multivalent antigens for promoting B and T Cell activation, ACS Chemical Biology **10**(8), 1817–1824 (2015).

[81] T. Kawasaki and T. Kawai, Toll-like receptor signaling pathways, Frontiers in Immunology **5**, 461 (2014).

[82] M. K. Slifka and I. J. Amanna, Role of multivalency and antigenic threshold in generating protective antibody responses, Frontiers in Immunology **10**, 956 (2019).

[83] E. B. Puffer, J. K. Pontrello, J. J. Hollenbeck, J. A. Kink, and L. L. Kiessling, Activating B Cell signaling with defined multivalent ligands, ACS Chemical Biology **2**(4), 252–262 (2007).

[84] Y. Zhang, M. Cheng, J. Cao, Y. Zhang, Z. Yuan, Q. Wu, and W. Wang, Multivalent nanoparticles for personalized theranostics based on tumor receptor distribution behavior, Nanoscale **11**(11), 5005–5013 (2019).

[85] M. E. Davis, Design and development of IT-101, a cyclodextrin-containing polymer conjugate of camptothecin, Advanced Drug Delivery Reviews **61**(13), 1189–1192 (2009).

[86] M. A. van Dongen, C. A. Dougherty, and M. M. Banaszak Holl, Multivalent polymers for drug delivery and imaging: The challenges of conjugation, Biomacromolecules **15**(9), 3215–3234 (2014).

[87] C. B. Carlson, P. Mowery, R. M. Owen, E. C. Dykhuizen, and L. L. Kiessling, Selective tumor cell targeting using low-affinity, multivalent interactions, ACS Chemical Biology **2**(2), 119–127 (2007).

[88] F. J. Martinez-Veracoechea and D. Frenkel, Designing super selectivity in multivalent nano-particle binding, Proceedings of the National Academy of Sciences **108**(27), 10963–10968 (2011).

[89] I. Langmuir, The adsorption of gases on plane surfaces of glass, mica and platinum., Journal of the American Chemical Society **40**(9), 1361–1403 (1918).

[90] G. V. Dubacheva, T. Curk, B. M. Mognetti, R. Auzély-Velty, D. Frenkel, and R. P. Richter, Superselective targeting using multivalent polymers, Journal of the American Chemical Society **136**(5), 1722–1725 (2014).

[91] G. V. Dubacheva, T. Curk, D. Frenkel, and R. P. Richter, Multivalent recognition at fluid surfaces: The interplay of receptor clustering and superselectivity, Journal of the American Chemical Society **141**(6), 2577–2588 (2019).

[92] T. Curk, J. Dobnikar, and D. Frenkel, Optimal multivalent targeting of membranes with many distinct receptors, Proceedings of the National Academy of Sciences **114**(28), 7210–7215 (2017).

[93] T. Curk, C. A. Brackley, J. D. Farrell, Z. Xing, D. Joshi, S. Direito, U. Bren, S. Angioletti-Uberti, J. Dobnikar, E. Eiser, D. Frenkel, and R. J. Allen, Computational design of probes to detect bacterial genomes by multivalent binding, Proceedings of the National Academy of Sciences **117**(16), 8719–8726 (2020).

[94] Z. Zhang, M. M. Ali, M. A. Eckert, D.-K. Kang, Y. Y. Chen, L. S. Sender, D. A. Fruman, and W. Zhao, A polyvalent aptamer system for targeted drug delivery, Biomaterials **34**(37), 9728–9735 (2013).

[95] L. Albertazzi, F. J. Martinez-Veracoechea, C. M. A. Leenders, I. K. Voets, D. Frenkel, and E. W. Meijer, Spatiotemporal control and superselectivity in supramolecular polymers using multivalency, Proceedings of the National Academy of Sciences **110**(30), 12203–12208 (2013).

[96] P. I. Kitov and D. R. Bundle, On the nature of the multivalency effect: A thermodynamic model, Journal of the American Chemical Society **125**(52), 16271–16284 (2003).

[97] D. van Effenterre and D. Roux, Adhesion of colloids on a cell surface in competition for mobile receptors, Europhysics Letters (EPL) **64**(4), 543–549 (2003).

[98] S. Maslanka Figueroa, D. Fleischmann, S. Beck, and A. Goepferich, The effect of ligand mobility on the cellular interaction of multivalent nanoparticles, Macromolecular Bioscience **20**(4), 1900427 (2020).

[99] F. J. Martinez-Veracoechea and M. E. Leunissen, The entropic impact of tethering, multivalency and dynamic recruitment in systems with specific binding groups, Soft Matter **9**(12), 3213–3219 (2013).

[100] V. M. Krishnamurthy, V. Semetey, P. J. Bracher, N. Shen, and G. M. Whitesides, Dependence of effective molarity on linker length for an intramolecular protein-ligand system, Journal of the American Chemical Society **129**(5), 1312–1320 (2007).

[101] R. S. Kane, Thermodynamics of multivalent interactions: Influence of the linker, Langmuir **26**(11), 8636–8640 (2010).

[102] S. Wang and E. E. Dormidontova, Nanoparticle targeting using multivalent ligands: computer modeling, Soft Matter **7**, 4435–4445 (2011).

[103] S. Liese and R. R. Netz, Influence of length and flexibility of spacers on the binding affinity of divalent ligands, Beilstein Journal of Organic Chemistry **11**(1), 804–816 (2015).

[104] C. D. Walkey, J. B. Olsen, H. Guo, A. Emili, and W. C. W. Chan, Nanoparticle size and surface chemistry determine serum protein adsorption and macrophage uptake, Journal of the American Chemical Society **134**(4), 2139–2147 (2012).

[105] M. R. Caplan and E. V. Rosca, Targeting drugs to combinations of receptors: a modeling analysis of potential specificity, Annals of Biomedical Engineering **33**(8), 1113–1124 (2005).

[106] X. Tian, S. Angioletti-Uberti, and G. Battaglia, On the design of precision nanomedicines, Science Advances **6**, eaat0919 (2020).

[107] S. Angioletti-Uberti, Exploiting receptor competition to enhance nanoparticle binding selectivity, Physical Review Letters **118**(6), 068001 (2017).

[108] M. Lundqvist, J. Stigler, G. Elia, I. Lynch, T. Cedervall, and K. A. Dawson, Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts, Proceedings of the National Academy of Sciences **105**(38), 14265–14270 (2008).

[109] M. P. Monopoli, C. Åberg, A. Salvati, and K. A. Dawson, Biomolecular coronas provide the biological identity of nanosized materials, Nature Nanotechnology **7**(12), 779–786 (2012).

[110] P. M. Biesheuvel and A. Wittemann, A modified box model including charge regulation for protein adsorption in a spherical polyelectrolyte brush, The Journal of Physical Chemistry B **109**(9), 4209–4214 (2005).

[111] J. Satulovsky, M. A. Carignano, and I. Szleifer, Kinetic and thermodynamic control of protein adsorption, Proceedings of the National Academy of Sciences **97**(16), 9037–9041 (2000).

[112] F. Fang and I. Szleifer, Kinetics and thermodynamics of protein adsorption: a generalized molecular theoretical approach, Biophysical Journal **80**(6), 2568–2589 (2001).

[113] B. M. Rubenstein, I. Coluzza, and M. A. Miller, Controlling the folding and substrate-binding of proteins using polymer brushes, Physical Review Letters **108**, 208104 (2012).

[114] K. Yu and J. N. Kizhakkedathu, Synthesis of functional polymer brushes containing carbohydrate residues in the pyranose form and their specific and nonspecific interactions with proteins, Biomacromolecules **11**(11), 3073–3085 (2010).

[115] J. C.-H. Kuo, J. G. Gandhi, R. N. Zia, and M. J. Paszek, Physical biology of the cancer cell glycocalyx, Nature Physics **14**(7), 658–669 (2018).

[116] G. V. Dubacheva, T. Curk, R. Auzély-Velty, D. Frenkel, and R. P. Richter, Designing multivalent probes for tunable superselective targeting, Proceedings of the National Academy of Sciences **112**(18), 5579–5584 (2015).

[117] W. G. Noid, Perspective: Coarse-grained models for biomolecular systems, The Journal of Chemical Physics **139**, 090901 (2013).

[118] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Coarse-grained protein models and their applications, Chemical Reviews **116**(14), 7898–7936 (2016).

[119] M. G. Saunders and G. A. Voth, Coarse-graining methods for computational biology, Annual Review of Biophysics **42**(1), 73–93 (2013).

[120] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, The MARTINI Force Field: Coarse grained model for biomolecular simulations, The Journal of Physical Chemistry B **111**(27), 7812–7824 (2007).

[121] I. M. Lifshitz, A. Y. Grosberg, and A. R. Khokhlov, Some problems of the statistical physics of polymer chains with volume interaction, Review of Modern Physics **50**(3), 683–713 (1978).

[122] M. Reches, P. W. Snyder, and G. M. Whitesides, Folding of electrostatically charged beads-on-a-string as an experimental realization of a theoretical model in polymer science, Proceedings of the National Academy of Sciences **106**(42), 17644–17649 (2009).

[123] K. A. Dill, S. Bromberg, K. Yue, H. S. Chan, K. M. Fiebig, D. P. Yee, and P. D. Thomas, Principles of protein folding — A perspective from simple exact models, Protein Science **4**(4), 561–602 (1995).

[124] H. Li, R. Helling, C. Tang, and N. Wingreen, Emergence of preferred structures in a simple model of protein folding, Science **273**(5275), 666–669 (1996).

[125] T. Veitshans, D. Klimov, and D. Thirumalai, Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties, Folding and Design **2**(1), 1–22 (1997).

[126] M.-H. Hao and H. A. Scheraga, Monte Carlo simulation of a first-order transition for protein folding, The Journal of Physical Chemistry **98**(18), 4940–4948 (1994).

[127] A. Šali, E. Shakhnovich, and M. Karplus, How does a protein fold?, Nature **369**(6477), 248–251 (1994).

[128] A. Šali, E. Shakhnovich, and M. Karplus, Kinetics of protein folding: A lattice model study of the requirements for folding to the native state, Journal of Molecular Biology **235**(5), 1614–1636 (1994).

[129] J. Skolnick and A. Kolinski, Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics, Journal of Molecular Biology **221**(2), 499–531 (1991).

[130] L. A. Mirny, V. Abkevich, and E. I. Shakhnovich, Universality and diversity of the protein folding scenarios:a comprehensive analysis with the aid of a lattice model, Folding and Design **1**(2), 103–116 (1996).

[131] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano, Phase equilibria of lattice polymers from Histogram Reweighting Monte Carlo Simulations, Macromolecules **31**(3), 912–918 (1998).

[132] P. de Gennes, Exponents for the excluded volume problem as derived by the Wilson method, Physics Letters A **38**(5), 339–340 (1972).

[133] M. Peto, A. Kloczkowski, and R. L. Jernigan, Shape-dependent designability studies of lattice proteins, Journal of Physics: Condensed Matter **19**(28), 285220 (2007).

[134] J. L. England, B. E. Shakhnovich, and E. I. Shakhnovich, Natural selection of more designable folds: A mechanism for thermophilic adaptation, Proceedings of the National Academy of Sciences **100**(15), 8727–8731 (2003).

[135] C. J. Camacho and D. Thirumalai, Kinetics and thermodynamics of folding in model proteins, Proceedings of the National Academy of Sciences **90**(13), 6369–6372 (1993).

[136] D. K. Klimov and D. Thirumalai, Cooperativity in protein folding: from lattice models with sidechains to real proteins, Folding and Design **3**(2), 127–139 (1998).

[137] Levinthal, C, Are there pathways for protein folding?, Journal de Chimie Physique **65**, 44–45 (1968).

[138] A. Kolinski, W. Galazka, and J. Skolnick, On the origin of the cooperativity of protein folding: implications from model simulations, Proteins: Structure, Function, and Bioinformatics **26**(3), 271–287 (1996).

[139] A. Sariban, K. Binder, and D. Heermann, Critical phenomena in polymer mixtures: Monte Carlo simulation of a lattice model, Colloid and Polymer Science **265**(5), 424–431 (1987).

[140] P. J. Flory, *Principles of polymer chemistry*, Cornell University Press, 1953.

[141] A. Sariban and K. Binder, Monte Carlo simulation of a lattice model for ternary polymer mixtures, Colloid and Polymer Science **266**(5), 389–397 (1988).

[142] S. Bromberg and K. A. Dill, Side-chain entropy and packing in proteins, Protein Science **3**(7), 997–1009 (1994).

[143] C. Perego and R. Potestio, Computational methods in the study of self-entangled proteins: a critical appraisal, Journal of Physics: Condensed Matter **31**(44), 443001 (2019).

[144] M. A. Soler and P. F. N. Faísca, Effects of Knots on Protein Folding Properties, PLoS One **8**(9), 1–10 (2013).

[145] P. F. N. Faísca, R. D. M. Travasso, T. Charters, A. Nunes, and M. Cieplak, The folding of knotted proteins: insights from lattice simulations, Physical Biology **7**(1), 016009 (2010).

[146] M. A. Soler, A. Nunes, and P. F. N. Faísca, Effects of knot type in the folding of topologically complex lattice proteins, The Journal of Chemical Physics **141**(2), 025101 (2014).

[147] D. S. Protter, B. S. Rao, B. Van Treeck, Y. Lin, L. Mizoue, M. K. Rosen, and R. Parker, Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly, Cell Reports **22**(6), 1401–1412 (2018).

[148] K. Kremer and K. Binder, Monte Carlo simulation of lattice models for macromolecules, Computer Physics Reports **7**(6), 259–310 (1988).

[149] J. C. le Guillou and J. Zinn-Justin, Critical exponents for the n-vector model in three dimensions from Field Theory, in *Large-Order Behaviour of Perturbation Theory*, volume 7 of *Current Physics–Sources and Comments*, pages 527–530, Elsevier, 1990.

[150] M. Fixman, Radius of Gyration of Polymer Chains, The Journal of Chemical Physics **36**(2), 306–310 (1962).

[151] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, volume 1, Academic Press, 2001.

[152] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of State Calculations by Fast Computing Machines, The Journal of Chemical Physics **21**(6), 1087–1092 (1953).

[153] J. Reiter, T. Edling, and T. Pakula, Monte Carlo simulation of lattice models for macromolecules at high densities, The Journal of Chemical Physics **93**(1), 837–844 (1990).

[154] I. Coluzza, H. Muller, and D. Frenkel, Designing refoldable model molecules, Physical Review E **68**(4), 046703 (2003).

[155] M. Doi, S. F. Edwards, and S. F. Edwards, *The theory of polymer dynamics*, volume 73, Oxford University Press, Oxford, 1988.

[156] N. Madras and A. D. Sokal, The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk, Journal of Statistical Physics **50**(1), 109–186 (1988).

[157] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo simulation of spin-glasses, Physical Review Letters **57**(21), 2607–2609 (1986).

[158] P. J. Flory and M. Volkenstein, Statistical mechanics of chain molecules, Biopolymers **8**(5), 699–700 (1969).

[159] A. N. Semenov and M. Rubinstein, Thermoreversible gelation in solutions of associative polymers. 1. Statics, Macromolecules **31**(4), 1373–1385 (1998).

[160] W. M. Jacobs and D. Frenkel, Phase transitions in biological systems with many components, Biophysical Journal **112**(4), 683–691 (2017).

[161] M. Chiappini, E. Eiser, and F. Sciortino, Phase behaviour in complementary DNA-coated gold nanoparticles and fd-viruses mixtures: a numerical study, The European Physical Journal E **40**, 7 (2017).

[162] H. Liu, S. K. Kumar, and F. Sciortino, Vapor-liquid coexistence of patchy models: Relevance to protein phase behavior, The Journal of Chemical Physics **127**, 084902–04906 (2007).

[163] D. Deviri and S. A. Safran, Equilibrium size distribution and phase separation of multivalent, molecular assemblies in dilute solution, Soft Matter **16**(23), 5458–5469 (2020).

[164] E. Zumbro and A. Alexander-Katz, Multivalent polymers can control phase boundary, dynamics, and organization of liquid-liquid phase separation, bioRxiv , 424934 (2020).

[165] A. N. Singh and A. Yethiraj, Liquid–liquid phase separation as the second step of complex coacervation, The Journal of Physical Chemistry B **125**(12), 3023–3031 (2021).

[166] A. Musacchio, M. Noble, R. Pauptit, R. Wierenga, and M. Saraste, Crystal structure of a Src-homology 3 (SH3) domain, Nature **359**(6398), 851–855 (1992).

[167] V. S. Bhatt, D. Zeng, I. Krieger, J. C. Sacchettini, and J.-H. Cho, Binding mechanism of the N-terminal SH3 domain of CrkII and proline-rich motifs in cAbl, Biophysical Journal **110**(12), 2630–2641 (2016).

[168] T. S. Harmon, A. S. Holehouse, and R. V. Pappu, Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins, New Journal of Physics **20**(4), 045002 (2018).

[169] S. Souquere, S. Mollet, M. Kress, F. Dautry, G. Pierron, and D. Weil, Unravelling the ultrastructure of stress granules and associated P-bodies in human cells, Journal of Cell Science **122**(20), 3619–3626 (2009).

[170] M. Doi, *Introduction to polymer physics*, Oxford university press, 1996.

[171] A. Mittal, N. Lyle, T. S. Harmon, and R. V. Pappu, Hamiltonian switch Metropolis Monte Carlo simulations for improved conformational sampling of intrinsically disordered regions tethered to ordered domains of proteins, Journal of Chemical Theory and Computation **10**(8), 3550–3562 (2014).

[172] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Germline P granules are liquid droplets that localize by controlled dissolution/condensation, Science **324**(5935), 1729–1732 (2009).

[173] A. K. L. Leung, J. M. Calabrese, and P. A. Sharp, Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules, Proceedings of the National Academy of Sciences **103**(48), 18125–18130 (2006).

[174] Y. Jo and Y. Jung, Interplay between intrinsically disordered proteins inside membraneless protein liquid droplets, Chemical Science **11**(5), 1269–1275 (2020).

[175] W. Xing, D. Muhlrad, R. Parker, and M. K. Rosen, A quantitative inventory of yeast P body proteins reveals principles of composition and specificity, Elife **9**, e56525 (2020).

[176] J. B. Woodruff, B. F. 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 (2017).

[177] I. Owen and F. Shewmaker, The role of post-translational modifications in the phase transitions of intrinsically disordered proteins, International Journal of Molecular Sciences **20**(21), 5501–5514 (2019).

[178] N. A. Yewdall, A. A. André, T. Lu, and E. Spruijt, Coacervates as models of membraneless organelles, Current Opinion in Colloid & Interface Science **52**, 101416 (2021).

[179] A. Bah and J. D. Forman-Kay, Modulation of intrinsically disordered protein function by post-translational modifications, Journal of Biological Chemistry **291**(13), 6696–6705 (2016).

[180] M. Hofweber and D. Dormann, Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics, Journal of Biological Chemistry **294**(18), 7137–7150 (2019).

[181] K. Fong, Y. Li, W. Wang, W. Ma, K. Li, R. Z. Qi, D. Liu, Z. Songyang, and J. Chen, Whole-genome screening identifies proteins localized to distinct nuclear bodies, Journal of Cell Biology **203**(1), 149–164 (2013).

[182] J. R. Buchan and R. Parker, Eukaryotic stress granules: The ins and outs of translation, Molecular Cell **36**(6), 932–941 (2009).

[183] J. Li and D. J. Mooney, Designing hydrogels for controlled drug delivery, Nature Reviews Materials **1**(12), 1–17 (2016).

[184] A. S. Hoffman, Hydrogels for biomedical applications, Advanced Drug Delivery Reviews **64**, 18–23 (2012).

[185] I. Y. Jung, J. S. Kim, B. R. Choi, K. Lee, and H. Lee, Hydrogel based biosensors for in vitro diagnostics of biochemicals, proteins, and genes, Advanced Healthcare Materials **6**(12), 1601475 (2017).

[186] C. Stoffelen and J. Huskens, Zwitterionic supramolecular nanoparticles: self-assembly and responsive properties, Nanoscale **7**(17), 7915–7919 (2015).

[187] C. Stoffelen, J. Voskuhl, P. Jonkheijm, and J. Huskens, Dual stimuli-responsive self-assembled supramolecular nanoparticles, Angewandte Chemie International Edition **53**(13), 3400–3404 (2014).

[188] H. Wang and T. Wu, A general lithography-free method of microscale/nanoscale fabrication and patterning on Si and Ge surfaces, Nanoscale Research Letters **7**(1), 1–7 (2012).

[189] M. S. Lord, M. Foss, and F. Besenbacher, Influence of nanoscale surface topography on protein adsorption and cellular response, Nano Today **5**(1), 66–78 (2010).

[190] H. Jeon, C. G. Simon Jr, and G. Kim, A mini-review: Cell response to microscale, nanoscale, and hierarchical patterning of surface structure, Journal of Biomedical Materials Research Part B: Applied Biomaterials **102**(7), 1580–1594 (2014).

[191] A. Baumgärtner and M. Muthukumar, Effects of surface roughness on adsorbed polymers, The Journal of Chemical Physics **94**(5), 4062–4070 (1991).

[192] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles, The Journal of Chemical Physics **96**(3), 1776–1783 (1992).

[193] E. Marinari and G. Parisi, Simulated tempering: A new Monte Carlo scheme, Europhysics Letters (EPL) **19**(6), 451–458 (1992).

[194] P. J. Tummino and R. A. Copeland, Residence time of receptor-ligand complexes and its effect on biological function, Biochemistry **47**(20), 5481–5492 (2008).

[195] S. T. Hyde and G. E. Schröder-Turk, Geometry of interfaces: topological complexity in biology and materials, Interface Focus **2**(5), 529–538 (2012).

[196] C. Kumar, A. Palacios, V. A. Surapaneni, G. Bold, M. Thielen, E. Licht, T. E. Higham, T. Speck, and V. Le Houérou, Replicating the complexity of natural surfaces: technique validation and applications for biomimetics, ecology and evolution, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **377**, 2138 (2019).

[197] M. A. Meyers, P.-Y. Chen, A. Y.-M. Lin, and Y. Seki, Biological materials: Structure and mechanical properties, Progress in Materials Science **53**(1), 1–206 (2008).

[198] I. Dimitrovski, D. Kocev, S. Loskovska, and S. Džeroski, Hierarchical classification of diatom images using ensembles of predictive clustering trees, Ecological Informatics **7**(1), 19–29 (2012).

[199] D. Losic, G. Rosengarten, J. Mitchell, and N. Voelcker, Pore architecture of diatom frustules: Potential nanostructured membranes for molecular and particle separations, Journal of Nanoscience and Nanotechnology **6**, 982–989 (2006).

[200] D. Bonifazi, S. Mohnani, and A. Llanes-Pallas, Supramolecular chemistry at interfaces: Molecular recognition on nanopatterned porous surfaces, Chemistry – A European Journal **15**(29), 7004–7025 (2009).

[201] X. Liu, P. K. Chu, and C. Ding, Surface nano-functionalization of biomaterials, Materials Science and Engineering: R: Reports **70**(3), 275–302 (2010).

[202] H. Jeon, C. G. Simon Jr, and G. Kim, A mini-review: Cell response to microscale, nanoscale, and hierarchical patterning of surface structure, Journal of Biomedical Materials Research Part B: Applied Biomaterials **102**(7), 1580–1594 (2014).

[203] L. R. Snyder, J. J. Kirkland, and J. W. Dolan, *Introduction to modern liquid chromatography*, John Wiley & Sons, 2011.

[204] O. Coskun, Separation techniques: chromatography, Northern Clinics of Istanbul **3**(2), 156–160 (2016).

[205] C. Pettinari, F. Marchetti, N. Mosca, G. Tosi, and A. Drozdov, Application of metal – organic frameworks, Polymer International **66**(6), 731–744 (2017).

[206] S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit, and H. J. Kulik, Understanding the diversity of the metal-organic framework ecosystem, Nature Communications **11**(1), 1–10 (2020).

[207] R. J. Kuppler, D. J. Timmons, Q.-R. Fang, J.-R. Li, T. A. Makal, M. D. Young, D. Yuan, D. Zhao, W. Zhuang, and H. Zhou, Potential applications of metal-organic frameworks, Coordination Chemistry Reviews **253**(23), 3042–3066 (2009).

[208] G. Férey and C. Serre, Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences, Chemical Society Reviews **38**(5), 1380–1399 (2009).

[209] C. Janiak and J. K. Vieth, MOFs, MILs and more: concepts, properties and applications for porous coordination networks (PCNs), New Journal of Chemistry **34**(11), 2366–2388 (2010).

[210] R. Lin, S. Xiang, W. Zhou, and B. Chen, Microporous metal-organic framework materials for gas separation, Chem **6**(2), 337–363 (2020).

[211] L. Feng, K. Wang, X. Lv, T. Yan, and H. Zhou, Hierarchically porous metal–organic frameworks: synthetic strategies and applications, National Science Review **7**(11), 1743–1758 (2019).

[212] V. Mouarrawis, R. Plessius, J. I. van der Vlugt, and J. N. H. Reek, Confinement effects in catalysis using well-defined materials and cages, Frontiers in Chemistry **6**, 623 (2018).

[213] G. L. Dignon, W. Zheng, and J. Mittal, Simulation methods for liquid–liquid phase separation of disordered proteins, Current Opinion in Chemical Engineering **23**, 92–98 (2019).

[214] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model, PLOS Computational Biology **14**(1), 1–23 (2018).

[215] M. K. Hazra and Y. Levy, Biophysics of phase separation of disordered proteins is governed by balance between short- and long-range interactions, The Journal of Physical Chemistry B **125**(9), 2202–2211 (2021).

[216] G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M. A. Czekalska, et al., Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions, Nature Communications **12**(1), 1–14 (2021).

[217] M. P. Allen et al., Introduction to molecular dynamics simulation, Computational Soft Matter: From Synthetic Polymers to Proteins **23**(1), 1–28 (2004).

[218] K. Kempfer, J. Devemy, A. Dequidt, M. Couty, and P. Malfreyt, Development of coarse-grained models for polymers by trajectory matching, ACS Omega **4**(3), 5955–5967 (2019).

[219] E. Bianchi, R. Blaak, and C. N. Likos, Patchy colloids: state of the art and perspectives, Physical Chemistry Chemical Physics **13**(14), 6397–6410 (2011).

[220] A. Gooneie, S. Schuschnigg, and C. Holzer, A review of multiscale computational methods in polymeric materials, Polymers **9**(1), 16 (2017).

[221] M. Paloni, R. Bailly, L. Ciandrini, and A. Barducci, Unraveling molecular interactions in liquid–liquid phase separation of disordered proteins by atomistic simulations, The Journal of Physical Chemistry B **124**(41), 9009–9016 (2020).

[222] A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, A. M. D'Ordine, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal, and N. L. Fawzi, TDP-43 α-helical structure tunes liquid-liquid phase separation and function, Proceedings of the National Academy of Sciences **117**(11), 5883–5894 (2020).

[223] R. K. Das, K. M. Ruff, and R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins, Current Opinion in Structural Biology **32**, 102–112 (2015).

[224] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The protein data bank, Nucleic Acids Research **28**(1), 235–242 (2000).

[225] S. E. Reichheld, L. D. Muiznieks, F. W. Keeley, and S. Sharpe, Direct observation of structure and dynamics during phase separation of an elastomeric protein, Proceedings of the National Academy of Sciences **114**(22), 4408–4415 (2017).

[226] M. Praprotnik and L. Delle Site, Multiscale molecular modeling, Biomolecular Simulations **924**, 567–583 (2013).

[227] M. Praprotnik, L. D. Site, and K. Kremer, Multiscale simulation of soft matter: From scale bridging to adaptive resolution, Annual Review of Physical Chemistry **59**(1), 545–571 (2008).

[228] H. Kim, N. C. Kim, Y.-D. Wang, E. A. Scarborough, J. Moore, Z. Diaz, K. S. MacLea, B. Freibaum, S. Li, A. Molliex, et al., Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS, Nature **495**(7442), 467–473 (2013).

[229] A. Patel, L. Malinovska, S. Saha, J. Wang, S. Alberti, Y. Krishnan, and A. A. Hyman, ATP as a biological hydrotrope, Science **356**(6339), 753–756 (2017).