LU-TP 22-01
January 2022

Competition between Phase Separation and Micellization in a Coarse-Grained Protein Model with Two Components

Fangxuan Lyu

Department of Astronomy and Theoretical Physics, Lund University

Master thesis (FYTM03, 30 ECTS) supervised by Anders Irbäck

LUND
UNIVERSITY

Abstract

Intrinsically disordered proteins (IDPs) are widely believed to play a key role in the formation of intracellular biomolecular condensates through liquid-liquid phase separation (LLPS). Furthermore, it has been shown *in vitro* that several IDPs are able to phase separate on their own. Computational modeling offers unique opportunities to investigate the driving forces and sequence-dependence of IDP LLPS. Prior computational studies have mainly focused on one-component systems. In this thesis, using a coarse-grained model, we investigate the aggregation behavior of systems composed of two hydrophobic/polar sequences, called A and B. On their own, sequence A forms droplets of a dense bulk phase, whereas sequence B forms micellar aggregates. Using Monte Carlo simulations, we investigate the competition between these two aggregation mechanisms in mixed systems containing both sequences. Keeping the total number of chains fixed, we monitor the transition from sequence A-dominated droplet formation to sequence B-dominated micellization as the fraction of sequence B chains is increased. The transition entails major changes in both internal organization and size of the observed aggregates, and we find that a variety of intermediate species exist.

Popular science description

The interior of cells, the basic structural and functional unit of living things, contains various subunits called organelles. Classical organelles, such as chloroplasts and mitochondria, are separated from their surroundings by a membrane. Recent experimental advances have shown that, in addition, there exist organelles that lack a surrounding membrane and rather behave as liquid droplets, as was first demonstrated about a decade ago. These membraneless bodies harbour high concentrations of proteins and nucleic acids, and serve as an important complement to classical organelles in the internal organization of cells. Their formation amounts to a phase separation, with the droplet and its background representing two distinct liquid phases, like a droplet of oil in water. Over the past decade, the character and function of membraneless organelles have become subjects of intense interest.

Currently, it is widely believed that intrinsically disordered proteins (IDPs) play an important role in biomolecular phase separation. Despite their lack of a well-defined 3D structure, IDPs are able to perform a variety of biological functions. Moreover, they frequently occur in membraneless organelles, and several such IDPs have been shown capable of generating droplets on their own. However, the ability of IDPs to phase separate has been shown to depend not only on their amino acid composition but also on the ordering of the amino acids along the protein sequence. Understanding this sequence-dependence is an important, largely unsolved problem. Some of the so far most extensively studied IDPs are linked to neurodegenerative diseases. The misfolding and aggregation of these IDPs may have a causative role in diseases such as Alzheimer's and Parkinson's. A potential phase separation could influence the aggregation properties of these disease-linked IDPs.

Exploring the mechanisms of biomolecular phase separation in realistic cellular environments is a challenge. To be able to investigate the basics of biomolecular phase separation, many experimental and computational studies have therefore focused on idealized systems with a single component, typically an IDP. Such studies have, in particular, provided insight into how the ability to phase separate depends on the amino acid sequence. However, droplets inside cells are more complicated, as they contain several different proteins as well as nucleic acids. In this thesis, we explore the phase behavior of a two-component system composed of two IDPs, using a simple coarse-grained model. By computer simulations for a fixed total number of molecules, we in particular investigate how the behavior of this system varies with the relative amount of the two components. We show that both the size and character of the clusters formed in these systems depend strongly on the relative amount of the two sequences.

Contents

List of acronyms

LLPS, liquid-liquid phase separation; IDP, intrinsically disordered protein; MC, Monte Carlo; MSW, million sweeps.

1 Introduction

In addition to classical membrane-bound organelles, eukaryotic cells contain a variety of membrane-less compartments or organelles, also known as biomolecular condensates [1]. Examples of such condensates include nucleoli [23] and Cajal bodies [25] in the cell nucleus, as well as cytoplasmic bodies such as P granules [2]. In 2009, a seminal study demonstrated that P granules form by a liquid-liquid phase separation (LLPS) process [2], which generates a high-density droplet of proteins and nucleic acids. Today, it is believed that LLPS is a frequent and significant regulating principle for biomolecular condensates, and therefore for the compartmentalization of cells. Furthermore, other membrane-less structures such as signalling complexes [12], centrosomes [32], membrane receptor clusters [26]share similar mechanisms. A functionally important consequence of the liquid-like nature of these structures is that they can exchange matter with the surrounding medium. LLPS has also been suggested to play a role in pathological protein aggregation related to neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) [11, 13], Alzheimer's disease and Parkinson's disease. Therefore, it is of great importance to understand the underlying biophysical mechanisms of LLPS.

There is mounting evidence that intrinsically disordered proteins (IDPs) play a vital role in driving phase separation *in vitro* [19, 16, 3] and *in vivo*. To elucidate the mechanisms of IDP LLPS, various theoretical and computational approaches have been employed. Two widely used analytical approaches are Flory-Huggins theory [9, 8], which describes the demixing of homopolymers in aqueous solution, and its extension by Voorn and Overbeek [20] to systems composed of two oppositely charged polymers. However, both these methods are based on mean-field theory, and insensitive to the ordering of amino acids along the chains. Another avenue is provided by the random phase approximation [30]. Lin *et al.* applied this method to explore how the distribution of charge along polyampholyte chains influences their phase behaviour [14]. A more general but computationally costlier approach to the study of IDP LLPS and its sequence-dependence is by biomolecular simulations based on coarse-grained models with explicit chains. Recent years have seen an increasing number of such studies, most of which have focused on one-component systems where all chains share the same sequence.

Nevertheless, a living cell is a multi-component entity, and our exploration of LLPS should not be limited to one-component systems. In this thesis, we used Monte Carlo methods to explore the aggregation behaviour of two-component systems in a coarse-grained hydrophobic/polar (H/P) model. In the systems considered, both components are 10-bead HP chains, and share the same composition. Specifically, we consider the alternating sequence $(HP)_5$ and the block sequence H_5P_5, called A and B, respectively. For a fixed total number of chains, we studied how the aggregation behaviour depends on the relative amount of A and B chains. One-component systems of either A or B chains have previously been studied using the same biophysical model [17, 18]. By finite-size scaling analysis of simulation data obtained for different system sizes, it was demonstrated that sequence A

phase separates, while sequence B does not do so [17, 18]. Qualitatively similar results were obtained in another study of one-component systems of HP chains, in which 37 different 20-bead HP sequences were considered [24]. A previous study of two-component systems was performed by Chan and coworkers [21], who focused on electrostatic forces rather than hydrophobicity.

The pure sequence A and pure sequence B systems are known from before to show very different aggregation behaviours. The pure sequence A system forms a single dominant droplet of a dense bulk phase, whereas the pure sequence B system forms several smaller aggregates with a micellar structure. In this thesis, we investigated the transition between these two types of behaviour, by numerical simulations for different A to B ratios, for a fixed total number of chains.

2 Methods

2.1 IDPs

Intrinsically disordered proteins (IDPs) are known to be a major driver of intracellular phase separation, and to be enriched in membraneless organelles [29]. They may be fully unstructured or contain large unstructured regions. In contrast to folded proteins with a well-defined 3D structure, IDPs are characterized by low-complexity, often repetitive, amino acid sequences.

Besides their role in phase separation, IDPs can perform several essential biological functions as a complement to folded proteins under physiological conditions [28]. It has been found that intrinsically disordered regions of eukaryotic proteomes are enriched in so-called short linear motifs (SLiMs) comprising approximately 3–10 amino acids, which act as binding sites and play important roles in cellular signaling [6, 5]. For example, the Axin scaffold protein recruits other proteins by their binding to a long intrinsically disordered region [31].

Statistically, IDPs tend to have a relatively high fraction of charged residues, which holds true for phase separating IDPs as well. Indeed, it has been shown that electrostatic interactions can be crucial in driving IDP LLPS [14]. However, there are also other interactions that can serve as the main driver of IDP LLPS, an example being interactions between aromatic amino acids. In this thesis, we consider a simple hydrophobic/polar protein model. Hydrophobicity is known to be a key factor in protein folding [15], and the set of hydrophobic amino acids include the ones usually classified as aromatic (phenylalanine, tryptophan and tyrosine).

2.2 Biophysical model

The lattice-based HP model has been widely used for elucidating basics of protein folding [7]. Here, the protein is represented by a self-avoiding chain of hydrophobic (H) or polar (P) beads on a lattice. In this thesis, we study the aggregation behavior of hydrophobic/polar chains in a simple off-lattice model. As in the lattice HP model, each amino acid is represented by a single bead, which reduces the size of the configuration space and makes it possible to study systems with a relatively large number of chains.

Using this model, we investigate mixed two-component systems, containing two different HP sequences. One is the alternating sequence $(\mathrm{HP})_5$, called A, and the other is the block sequence $\mathrm{H_5P_5}$, called B. The length of the bonds between adjacent beads in the chain, called b, is kept fixed. All beads are hard spheres with a diameter $r_\mathrm{d} = 0.75b$. If a bead pair is at a distance $r_{ij} < r_\mathrm{d}$, the potential becomes infinite. In addition to the hard-core repulsion, HH pairs have an attractive interaction. The range of this interaction is $r_h = 2b$. HP and PP pairs interact only through the hard-core repulsion. Therefore, the pair potential of the simulated system, E_{ij}, and total energy, E, can be formulated as

$$E_{ij} = \begin{cases} \infty, & \text{if } r_{ij} < r_\mathrm{d}, \\ u_{ij}, & \text{if } r_\mathrm{d} < r_{ij} < r_\mathrm{h}, \\ 0, & \text{if } r_{ij} > r_\mathrm{h}, \end{cases} \tag{2.1}$$

$$E = \sum_{i<j} E_{ij} \tag{2.2}$$

Here, $u_{ij} = -\epsilon$ for HH pairs ($\epsilon > 0$), while $u_{ij} = 0$ for HP and PP pairs. Throughout the paper, we express energies and lengths in units of ϵ and b, respectively.

2.3 Simulation techniques

We performed explicit-chain Monte Carlo (MC) simulations of this model in the canonical ensemble (fixed NVT), to explore the thermodynamic behavior of mixtures of sequence A and B chains. For each choice of N, V and T, we performed a simulation run comprising between 1–3 MSW, where MSW stands for million sweeps. In all cases, the analysis is based on averaging over the last 0.8 MSW, discarding the initial part of the runs for thermalization. Each sweep consists of a fixed number of attempted elementary moves, which is set equal to the total number of beads.

The move set consists of seven elementary updates. Three of these, called one-bead, two-bead and pivot-type moves, alter the internal structure of individual chains. The one-bead move updates the position of a randomly selected bead, excluding the end beads of a chain. This move is around the axis formed by its nearest neighbours. Likewise, the two-bead move updates the location of two consecutive beads along the chain. Finally, the pivot update rotates one part of the chain relative to the rest. The remaining four updates

are rigid-body translations or rotations of whole chains or a cluster of chains. In the cluster updates, the cluster construction is stochastic, following a Swendsen-Wang-type procedure [10, 27]. The resulting cluster is subject to a rigid-body rotation or translation, which is accepted whenever this attempted move does not cause any steric clashes. Note that the choice of weights in the probabilistic cluster construction is such that detailed balance is automatically fulfilled, without having to use any Metropolis accept/reject test. All our other elementary moves use a Metropolis test to ensure that detailed balance is fulfilled.

2.4 Data analysis

To analyse the aggregation properties, the simulated multi-chain configurations were divided into clusters of interacting chains. All analyses were carried out using MATLAB (v. R2020a).

2.4.1 Periodic boundary conditions

All simulations were carried out assuming periodic boundary conditions in all three dimensions. The boundary conditions must be taken into account in computing the distance between a pair of beads. Consider a cubic box, $0 < x, y, z < L$, and let (x_1, y_1, z_1) and (x_2, y_2, z_2) be two points in that box. The distance between the two points is then computed as $d = \sqrt{d_x^2 + d_y^2 + d_z^2}$, with $d_x = \min(|x_1 - x_2|, L - |x_1 - x_2|)$ and analogous expressions for d_y and d_z.

Keeping track of the periodic boundary conditions is important when analyzing the distribution of beads within clusters (see below). For this analysis, for convenience, we translated each observed cluster of interest so that its centre of mass is at the centre of the simulation box, by a two-step procedure. The purpose of the first step is to put the cluster sufficiently close to the box centre that periodic copies of the systems can be ignored in the subsequent calculations. To this end, we constructed 1D histograms of the beads belonging to the given cluster in all three dimensions. In each dimension (x, y or z), the beads are then shifted, taking periodicity into account, so as to have the bin with most beads at the centre of the box. After this approximate centring, in the second step, the centre of mass can be computed without keeping track of periodicity. A final rigid-body translation of the cluster brings its centre of mass to the centre of the box.

2.4.2 Cluster mass distributions

We divided the chains into clusters based on chain-chain contacts. Any pair of chains deemed in contact must end up in the same cluster. To decide whether two chains are in contact, we counted the number of HH bead contacts between the two chains. Two

H beads are said to be in contact if they have a non-zero pair energy E_{ij}, which holds if $r_{ij} < r_{\mathrm{h}}$ (see Eq. 2.1). Finally, we say that the two chains are in contact if there are n_{HH} or more HH bead contacts between them.

The division of a given configuration into clusters is done iteratively. When a cluster has been identified, some chain among those not yet assigned to a cluster is picked as a seed for the next cluster. All chains in contact with that chain are added to the new cluster. This step is repeated until there is no chain-chain contact between the new cluster and other chains.

In this thesis, we set the threshold for chain-chain contacts to $n_{\mathrm{HH}} = 1$ bead contacts. We also tested using other values such as 10 and 20. The use of a larger n_{HH} may prevent clusters in "accidental" contact from being identified as a single cluster, but we found that it may also lead to a failure to identify proper clusters, especially for sequence B-abundant systems. For simplicity, we decided to stick to the choice $n_{\mathrm{HH}} = 1$.

Having identified the clusters, we determined the number of chains, or mass, of each cluster, and constructed a cluster mass distribution, $P(m)$. Here, $P(m)$ is proportional to m times the number of observed clusters with mass m. Therefore, $P(m)$ is the probability of finding a random chain in a cluster with size m. In presenting computed $P(m)$ distributions below, for clarity, we used bins of size $\Delta m = 5$. Spikes in $P(m)$ indicate typical cluster sizes.

2.4.3 Cluster structure — bead density distributions

It turns out that the cluster mass distributions are dominated by 2–5 distinct peaks, depending on the A/B composition (see Sec. 3.2.3). Each peak defines a class of clusters. Visual inspection showed that clusters in the same class typically shared not only a similar mass but also a similar overall shape. To quantitatively elucidate the structural properties of a given cluster class, we analyzed the distribution of beads within the clusters. For a given class, we first computed a bead density distribution for each cluster in the class, and then averaged over all clusters in the class.

For a cluster class consisting of approximately spherical clusters, it would be natural to compute the radial distribution of beads around the center of mass, $P(r)$. However, for many of our cluster classes, the typical shape is ellipsoidal rather than spherical. Therefore, we computed a two-dimensional bead density distribution, $P(r, z)$, where r and z are cylindrical coordinates. This description is relevant for most of our cluster classes, but exceptions exist. In particular, cylindrical coordinates are of limited use when dealing with the large and curved clusters present in the $R_{240/60}$ system (see Sec. 3.2.2). In our bead density analysis, we therefore omitted the $R_{240/60}$ system.

To compute the distribution $P(r, z)$ for a cluster class, we conducted the following calculations for each cluster in the class:

1. Perform a rigid-body translation of the cluster to bring its center of mass to the

center of the simulation box. This step effectively eliminates the risk that the cluster crosses the periodic boundaries.

2. Let N_b be the number of beads in the cluster, and let $\mathbf{r}_\alpha = (x_\alpha, y_\alpha, z_\alpha)$ be the bead positions ($\alpha = 1, \ldots, N_b$). With $x_1 = x$, $x_2 = y$ and $x_3 = z$, compute the covariance matrix C_{ij}, defined by

$$C_{ij} = \frac{1}{N_b - 1} \sum_{\alpha=1}^{N_b} (x_{i,\alpha} - \overline{x}_i)(x_{j,\alpha} - \overline{x}_j), \qquad \overline{x}_i = \frac{1}{N_b} \sum_{\alpha=1}^{N_b} x_{i,\alpha} \, .$$

   Find the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the 3×3-matrix C_{ij}, and the corresponding eigenvectors $\mathbf{v}_i$. The eigenvector with the largest eigenvalue, $\mathbf{v}_1$, is the direction in which the spread of the beads is largest.

3. Perform a rigid-body rotation of the cluster such that $\mathbf{v}_1$ points in the z direction after the rotation. With the cluster oriented in this way, find the cylindrical coordinates z and r, where r is the distance to the z axis, for each bead in the cluster. Count the number of beads α with $r < r_\alpha < r + dr$ and $z < z_\alpha < z + dz$. Divide this number by the volume of this (ring-like) region, $V = (2\pi r dr + \pi dr^2)dz$, to obtain the bead density at (r, z).

Finally, the bead density of the cluster class was computed by averaging over all individual clusters.

Figure 1 illustrates the distributions of covariance matrix eigenvalues for large clusters in the $R_{300/0}$ and $R_{270/30}$ systems. In the $R_{300/0}$ system, the clusters are approximately spherical, and the eigenvalues are not clearly separated. By contrast, in the $R_{270/30}$ system, the largest eigenvalue is markedly larger than the other two, reflecting that the clusters are ellipsoidal rather than spherical in this system.

For the pure A and pure B systems, we also performed simulations in an elongated geometry. For these systems, in which large clusters are slab-like, we use one-dimensional bead density profiles, calculated as functions of the position in the elongated direction. These one-dimensional bead density distributions are easier to compute, as steps 2 and 3 in the above procedure are not needed.

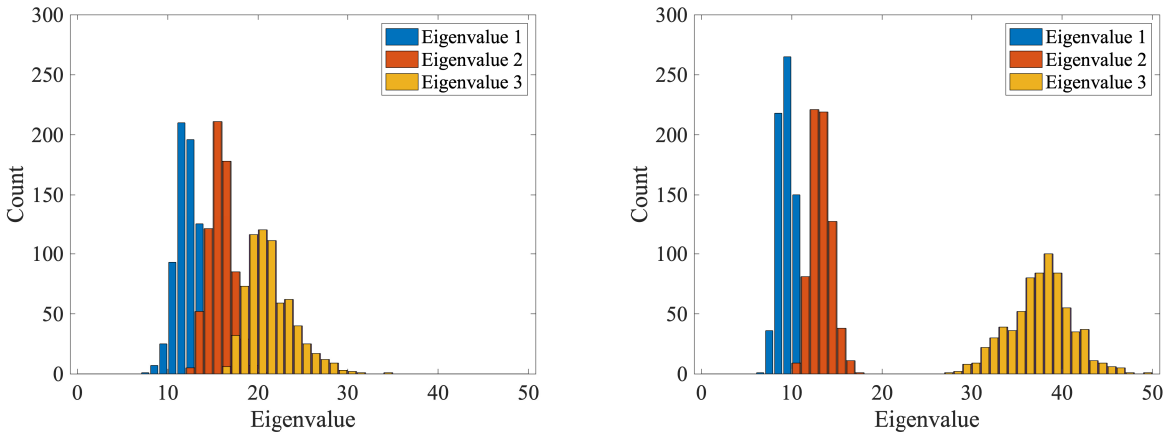


Figure 1: Covariance matrix eigenvalues for large clusters in (left) the $R_{300/0}$ system and (right) the $R_{270/30}$ system.

2.5 Simulation details

The MC simulations were conducted using the C code of Ref. [17], modified to enable the study of two-component systems. Before tackling the two-component systems, we studied the pure A and pure B systems and compared the results with those of Ref. [17]. For all systems, we performed simulations with 300 chains in a cubic box, at a density $\rho = 0.025b^3$. The pure A system, which undergoes phase separation, was also studied using an elongated simulation box, to facilitate estimation of the phase diagram. The number of chains and the density were the same as in the cubic simulations. Basic parameters of the simulated systems can be found in Table 1.

Table 1: Physical parameters of the simulated systems. The elongated geometry was only used for the pure sequence A system.

Geometry	Number of chains	Size (b^3)	ρ (b^{-3})	T (ϵ/k_B)
Cubic	300	49.32^3	0.025	2.6
Elongated	300	$14^2 \times 204$	0.025	2.6–3.1

When simulating the two-component systems, we found that the time required for thermalisation increased when increasing the fraction of sequence B chains (see Fig. 3 below). Therefore, we increased the total simulation time for systems with many sequence B chains. For all systems, the analysis was based on the final 0.8 MSW of the runs. The lengths of the different simulation runs can be found in Table 2.

Table 2: The total MC time, the MC time discarded for thermalisation, and the analyzed part of the runs for all the simulated systems, in units of MSW (10^6 sweeps).

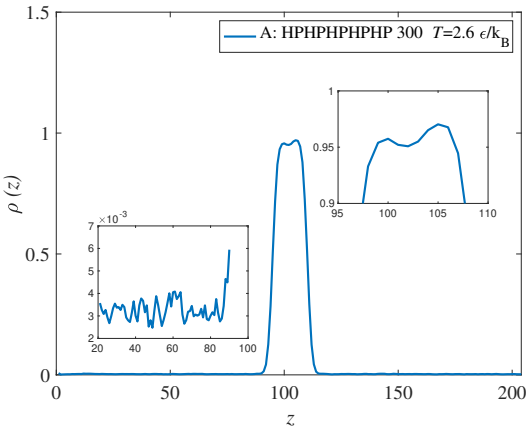
Ratio A/B ($R_{A/B}$)	MC time	Thermalisation	Analysis range
300/0	1	0.2	0.2–1
270/30	1	0.2	0.2–1
240/60	1	0.2	0.2–1
180/120	1.5	0.7	0.7–1.5
150/150	3	2.2	2.2–3
120/180	2	1.2	1.2–2
60/240	3	2.2	2.2–3
40/260	2	1.2	1.2–2
30/270	3	2.2	2.2–3
20/280	3	2.2	2.2–3
10/290	2	1.2	1.2–2
0/300	2	1.2	1.2–2

3 Results

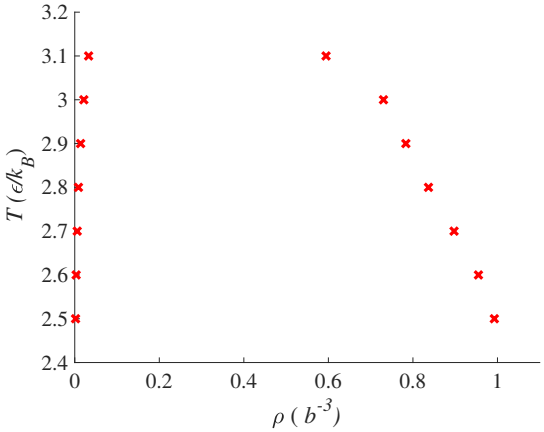
Using the model and methods described in Sec. 2, we performed thermodynamic simulations of various systems composed of 300 HP chains. This section presents the results of our simulations. We first discuss the phase diagram of a one-component system, where all the 300 chains share the same sequence, namely the alternating sequence A, $(HP)_5$. We then turn to the main aim of this thesis, which is to explore the condensation behavior in systems where sequence A is mixed with the block sequence B, (H_5P_5).

3.1 Warm-up exercise: phase behaviour of one-component systems

Previous work has shown that sequence A undergoes phase separation, through a finite-size scaling analysis of simulation results for different system sizes [17]. In the mixed two-phase regime, a dense and a dilute phase were observed to coexist, with densities ρ_h and ρ_ℓ, respectively. For this system, phase separation is observed if T is below a critical temperature, $T < T_c$, and the total density ρ lies between the two pure-phase densities, $\rho_\ell(T) < \rho < \rho_h(T)$. If ρ is larger but not much larger than ρ_ℓ and the system is cubic, an approximately spherical droplet of the dense phase is formed, in a dilute background. The phase behavior of this sequence was further studied in Ref. [18], where, in particular, it was demonstrated that using an elongated rather than a cubic system leads to markedly reduced finite-size effects. This has to do with the fact that the droplets are slab-like rather than spherical in an elongated geometry, and therefore can grow without increasing their



(a) Density profile



(b) Phase diagram

Figure 2: (a) Bead density profile of a one-component system with 300 chains of sequence A ($\rho = 0.025\, b^{-3}$) simulated in an elongated geometry, at $T = 2.6\, \epsilon/k_B$. (b) Temperature-density phase diagram of the system, based on data for seven temperatures. The left and right branches correspond to $\rho_\ell(T)$ and $\rho_h(T)$, respectively, as estimated from bead density profiles like the one in (a).

surface area [18].

Following Ref. [18], we simulated 300 chains of sequence A using an elongated box. The box dimensions were $14^2 \times 204\, b^3$, yielding a total bead density $\rho = 0.025\, b^{-3}$. Simulations were performed for seven different temperatures $T < T_c$. For each T, we computed a bead density profile $\rho(z)$, where z denotes the coordinate in the elongated direction. Here, before averaging, each analyzed snapshot was shifted so as to have the droplet at the center of the box. Figure 2a shows the computed average profile at $T = 2.6\, \epsilon/k_B$. From this figure, one can estimate ρ_h ($0.9550\, b^{-3}$) from the plateau at the center of the droplet and ρ_ℓ ($0.0033\, b^{-3}$) from the much lower background density. These two values, ρ_h and ρ_ℓ, yield two points on the coexistence curve that defines the mixed two-phase region (Fig. 2b). A similar analysis was presented in Ref. [18], but for a slightly different total density ρ. Qualitatively, our results are in good agreement with these prior findings.

We also simulated a cubic system with 300 chains of sequence B, using $\rho = 0.025\, b^{-3}$ and $T = 2.6\, \epsilon/k_B$. Unlike the alternating sequence A, the block sequence B is known not to undergo phase separation [17]. In line with prior work, we found the sequence B system to form several similar-size assemblies, rather than one dominant droplet(see Fig. 4 below). The assemblies had a micellar structure, with the hydrophobic part of the chains in the core and the polar part pointing outwards. A similar behavior was observed at $T = 5.2\, \epsilon/k_B$, but with fewer micelles.

3.2 Two-component systems

Having obtained results in line with expectations for the pure sequence A and pure sequence B systems, we proceeded to investigate mixed two-component systems with various A and B ratios. All systems consisted of 300 chains in a cubic box and were studied using $T = 2.6\,\epsilon/k_B$ and a total bead density $\rho = 0.025\,b^{-3}$. In what follows, we will often refer to the systems studied as $R_{A/B}$, where A and B are the numbers of chains with the sequences A and B. In total, we conducted Monte Carlo simulations for 12 different systems $R_{A/B}$, including the two one-component systems $R_{300/0}$ and $R_{0/300}$ (see Table 2). Below, we discuss the results from these simulations. We first examine the run-time evolution of the energy in the simulations, mainly to check for convergence. We then discuss representative snapshots from the different simulations. Finally, we present and discuss cluster mass distributions and bead density distributions inside large clusters for the different systems.

Prior studies have shown that the phase behavior of one-component systems of HP chains depends on the distribution of hydrophobicity along the sequence. In this thesis, we study two-component A/B systems and vary the A to B ratio, rather than the hydrophobicity distribution along a single chain.

3.2.1 Monte Carlo evolution of the energy

All our simulations were started from random initial configurations. The temperature T and density ρ are such that all the systems studied spontaneously form large clusters (with $\gtrsim 20$ chains), whose properties we wish to investigate. How long it will take for representative clusters to form is system dependent. To decide how large a part of the generated trajectories to discard for thermalization, we examined the Monte Carlo evolution of the energy E.

Figure 3 shows the run-time evolution of E in 10 of the 12 systems (two systems were omitted to make the figure readable). In the sequence A-dominated systems, the thermalization was fast, with no notable drift in E after 0.2 MSW. In systems with 60 or more chains of sequence B, the thermalization proceeded considerably slower. For example, in the $R_{60/240}$ system, a steady drift in E occurred out to about 2.0 MSW. Because of the system dependence of the time required for thermalization, the total simulation time was chosen to vary between 1.0 MSW and 3.0 MSW (Table 2). Despite this adjustment, we cannot rule out that the results obtained for the most slowly converging systems are slightly affected by incomplete thermalization, but we believe that these effects are small. All the analyses discussed below are based on data from the final 0.8 MSW of the simulations.

In Fig. 3, it is also interesting to compare the equilibrium values of the potential energy E among the different systems. Despite that the sequences A and B share the same composition (5 H + 5 P), it can be seen that the potential energy depends strongly on the A to B ratio. The energy is highest for the pure sequence A system, and decreases steadily as the number of sequence B chains is increased, until a minimum is reached around 270–

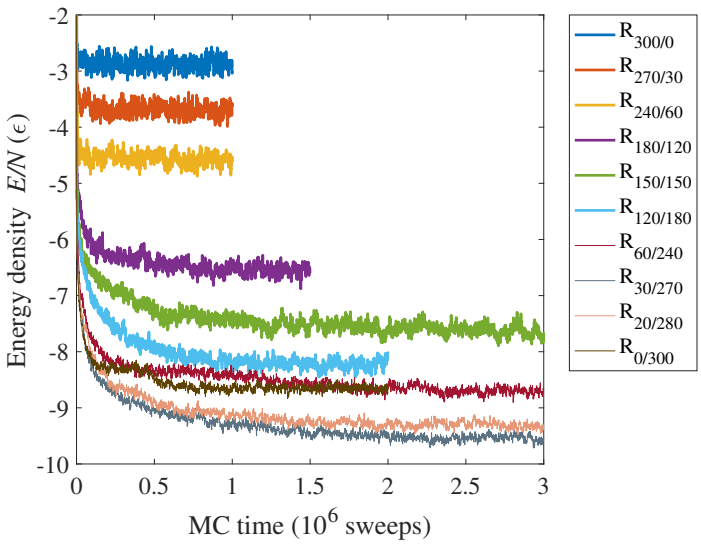


Figure 3: MC evolution of the energy density E/N for some of the 300-chain systems with different A to B ratios, at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. The total simulation time was chosen to be system dependent, because the speed of the thermalisation process varies.

280 sequence B chains. The most sequence B-dominated systems, $R_{10/290}$ and $R_{0/300}$, have slightly higher energies than the $R_{30/300}$ and $R_{20/300}$ systems. Overall, the energy difference between the sequence A- and sequence B-dominated systems indicate that sequence B is more interaction-prone and potentially forms denser clusters.

The cluster properties at equilibrium depend on the balance between energy and entropy, which is such that the free energy is minimized. When a free chain joins an existing cluster, there is a cost in entropy that must be balanced by reduced energy. One may compare it to the folding process that brings a globular protein to its native state [22]. Upon folding, entropy is lost, which is compensated for by a decrease in energy. Note that clusters are not static at equilibrium, but exchange chains with their dilute surroundings. Next, we take a more detailed look at the size and character of the equilibrium clusters in the different systems.

3.2.2 Snapshots from the simulations

Figure 4 shows representative snapshots from some of our simulations of two-component A/B systems, to give an idea of how the size and character of the aggregates vary with the A to B ratio. The two extreme cases of pure A and pure B systems have been studied before using the same model [17]. The pure A system forms a large, approximately spherical cluster in a dilute background, while the pure B system forms several similar-size micellar structures (Fig. 4). As indicated by Fig. 4, one key difference between the two systems is that the large clusters formed by the pure A system are homogeneous, with H and P beads

fully mixed, whereas H beads are concentrated in the interior of the aggregates formed by the pure B system, with P beads facing outwards.

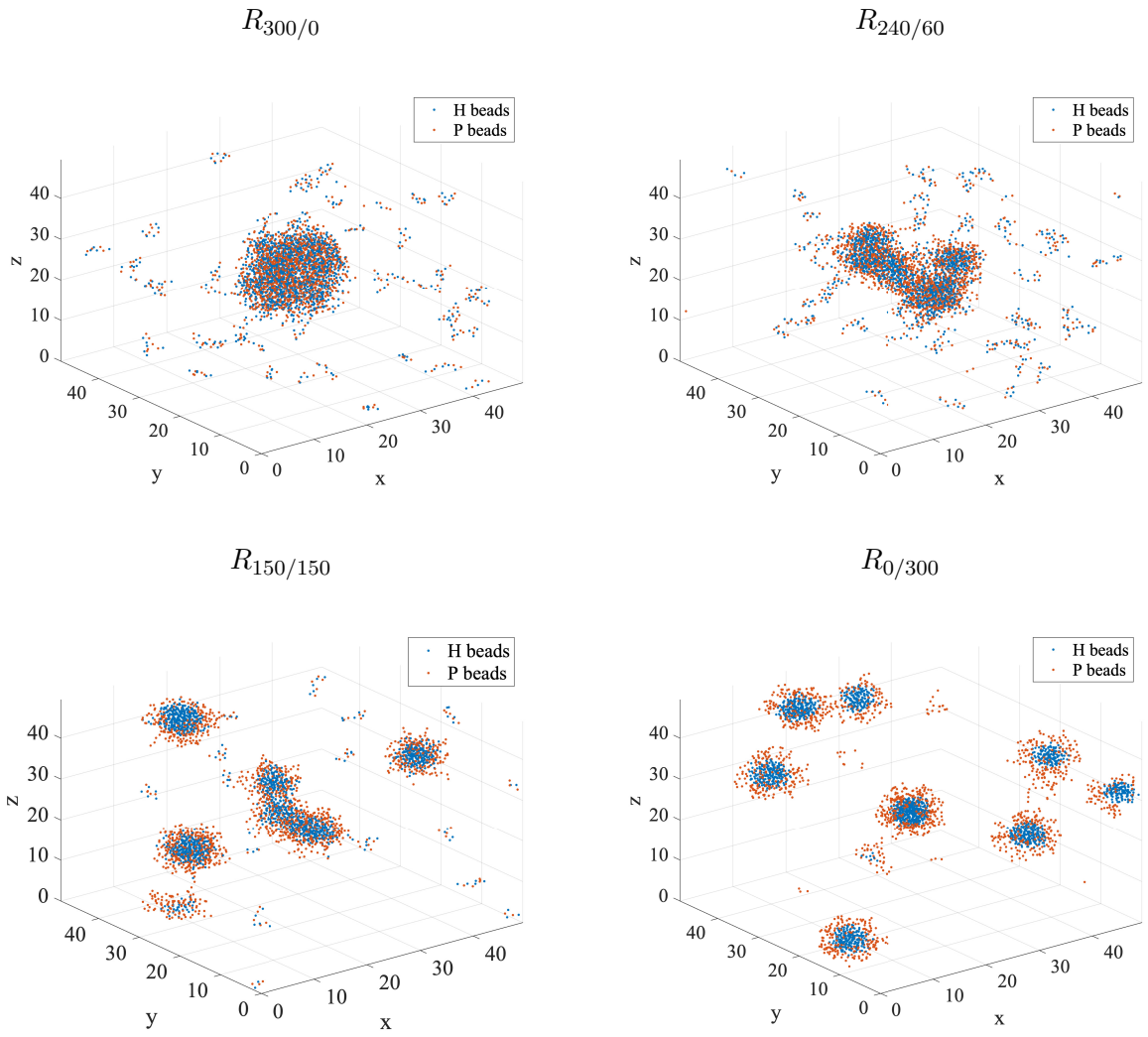


Figure 4: Representative snapshots from simulations with 300 HP chains for four different A to B ratios ($R_{300/0}$, $R_{240/60}$, $R_{150/150}$, and $R_{0/300}$), at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. The blue dots represent hydrophobic beads and the orange dots indicate polar beads. Snapshots were produced using MATLAB.

In a recent study, Brangwynne and coworkers investigated the aggregation of HP chains using a different coarse-grained model [24]. They focused on one-component systems. In particular, they demonstrated that the degree of "blockiness" of the sequence strongly influences the shape of the aggregates. A simple estimate of blockiness is provided by the average length of hydrophobic segments in a given sequence. A slightly more advanced measure of blockiness is the pattern parameter κ introduced by Das and Pappu [4], which

has been widely used in protein studies. For sequences with low blockiness, large spherical liquid-like droplets were observed in a dilute background, in agreement with the results of Ref. [24] and the snapshot shown in Fig. 4 for our alternating sequence A. As the degree of blockiness was increased, a variety of aggregates, differing in both size and character, were observed, which were termed structured liquid, strings, membranes, worm-like micelles and finally micelles [24]. The observation of micelles for high degrees of blockiness matches well with the results of Ref. 18 and the snapshot in Fig. 4 for the B system. The worm-like, rather than spherical, micelles were characterized by a thin string-like hydrophobic core surrounded by a polar surface.

We investigated two-component A/B systems by simulations for several different A to B ratios. Changing the A to B ratio of the system effectively leads to an altered overall degree of blockiness. We found that the A/B system exhibits a single dominant droplet as long as the amount of B sequence chains stays sufficiently low. However, when adding B sequence chains, the spherical shape seen in the pure A system gets distorted. In the $R_{240/60}$ system, we still observed a single large droplet, but the droplet had an irregular, far from spherical shape (Fig. 4). We also simulated a 400-chain system with the same A to B ratio, $R_{320/80}$, and observed similar but longer aggregates. The aggregates observed for this A to B ratio are reminiscent of the fibre-like structures observed by Brangwynne and coworkers in 400-chain one-component simulations for the 20-bead sequence $(H_3P_2)_4$.

When further decreasing the A to B ratio, we found that the large aggregate breaks up into several smaller ones. The smaller aggregates often had an approximately spherical, micellar structure, as in the pure B system. However, other shapes were also observed, as is illustrated by the snapshot from the $R_{150/150}$ simulation shown in Fig. 4. In this simulation, a structurally diverse set of aggregates occurred, including both spherical and elongated micellar structures, with one to three hydrophobic cores.

In systems with a small fraction of sequence B chains, as in the pure A system, we observe one large droplet (Fig. A1). The pure A system is known to undergo phase separation [17, 18]. However, in the systems with a non-zero fraction of sequence B chains, the large droplet does not necessarily represent a dense bulk phase.

3.2.3 Cluster mass distributions

We now turn to a quantitative analysis of the clusters that occurred in our simulations of the different two-component A/B systems. To this end, we first decomposed each snapshot from the simulations into clusters, following the procedure outlined in Sec. 2.4.2. In this section, we study the size of the simulated clusters. By counting cluster sizes in individual snapshots and then averaging over snapshots, we obtained cluster mass distributions, which give the probability of finding a randomly selected chain in a cluster of a given size. The resulting cluster mass distributions all exhibit two or more spikes (Figs. 5 and A2). Each spike signifies a statistically preferred cluster size. The spikes have a non-zero width, since the clusters do not have a well-defined 3D structure and may exchange chains with the

surroundings.

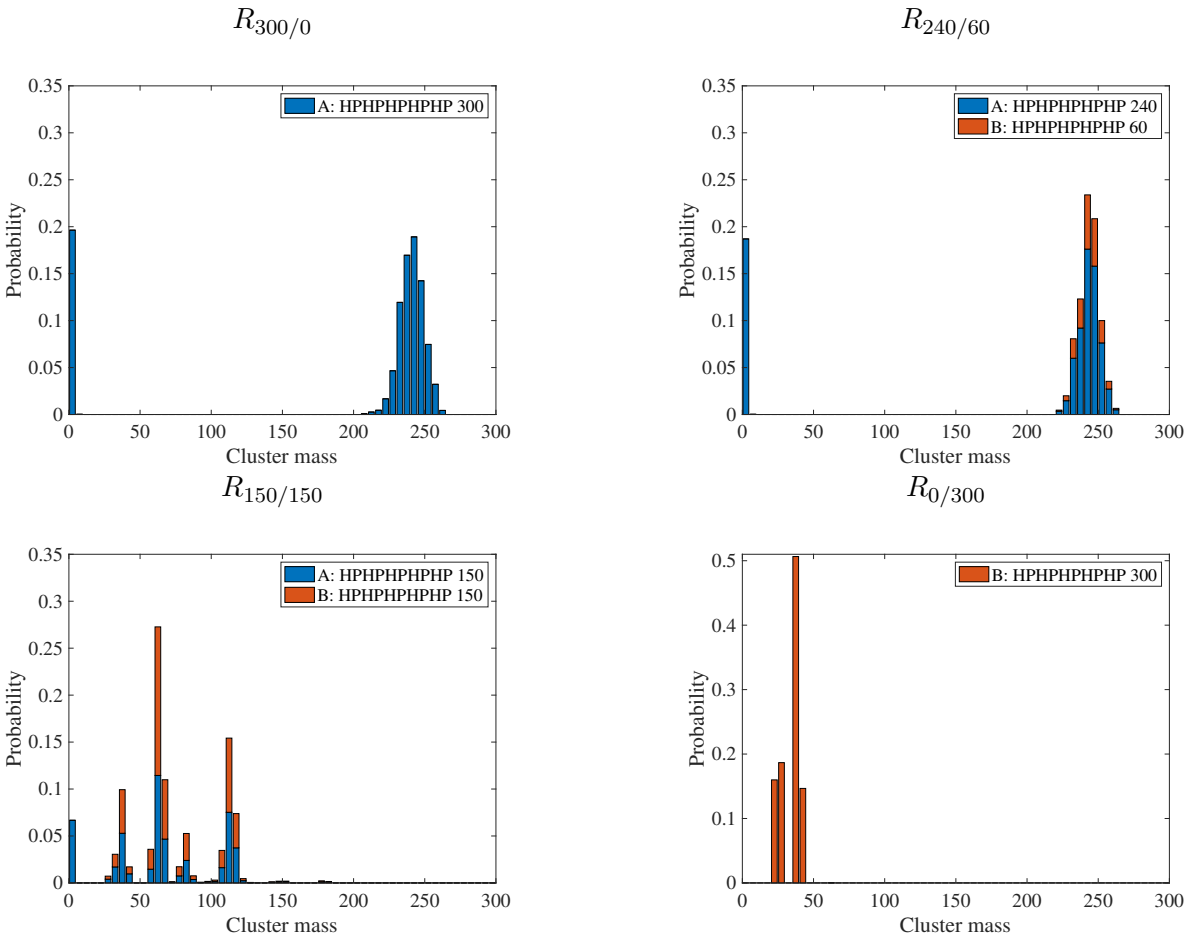


Figure 5: Cluster mass distribution in 300-chain systems with different A to B ratios ($R_{300/0}$, $R_{240/60}$, $R_{150/150}$, $R_{0/300}$). The mass of a cluster is defined as the number of chains it contains. Each system was simulated at $T = 2.6\,\epsilon/k_B$ and fixed total bead-density $\rho = 0.025\,b^{-3}$. The total probability of a given mass was partitioned into contributions from sequence A (blue) and sequence B (orange), respectively. In each system, there are two or more typical cluster sizes, corresponding to distinct spikes in the distribution.

Figure 5 shows cluster mass distributions for the same four systems studied in Fig. 4 (cluster mass distributions for all the simulated systems can be found in Fig. A2). Consistent with Fig. 4, the calculated cluster mass distributions for the $R_{300/0}$ and $R_{240/60}$ systems show two pronounced peaks, one corresponding to a large droplet and the other to free molecules and tiny aggregates. The high-mass peak is centered slightly below 250 chains. Worth noting is that the low-mass peak is dominated by sequence A. Essentially all sequence B chains end up in the large cluster, reflecting their high interaction propensity.

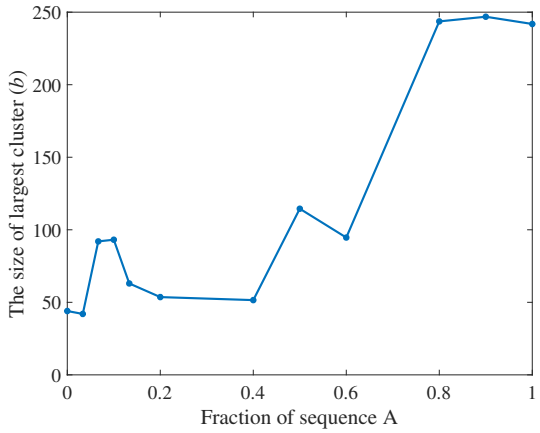
By contrast, in the $R_{0/300}$ system, only mid-size clusters were observed, with 20–50 chains.

The absence of low-mass clusters is not unexpected given the strong dominance of sequence A chains in the low-mass clusters of the $R_{240/60}$ system. While the clusters in the $R_{0/300}$ have a relatively uniform size, the cluster mass distribution still has two distinct peaks, separated by a narrow but deep gap. This gap indicates that there may exist two main cluster types with some fundamental structural difference. However, beyond size, we have not been able to identify any structural property that discriminates the two cluster types.

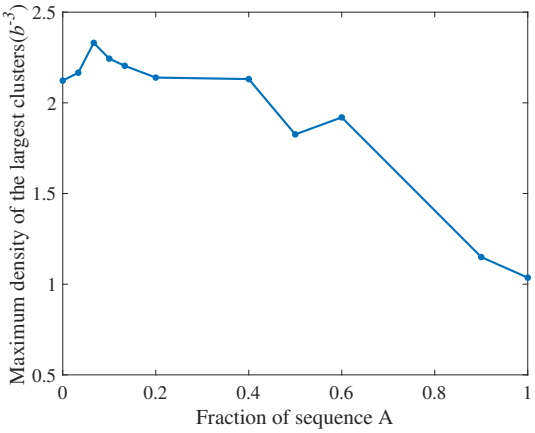
Finally, for the 50-50 system $R_{150/150}$, we find as many as five distinct peaks in the cluster mass distribution. A peak corresponding to low-mass clusters is present in this system, and these clusters are, again, dominated by sequence A chains. The next two peaks, centered at $\sim$40 and $\sim$60, both correspond to approximately spherical clusters. The fraction of sequence B chains is somewhat higher for the clusters with $\sim$60 chains than for those with $\sim$40 chains, which suggests that the $\sim$60 clusters may be denser. The remaining two peaks, centered at $\sim$80 and $\sim$110, correspond to worm-like rather than spherical micelles, with two and three hydrophobic cores, respectively.

Figure 6a shows how the size of the largest cluster varies with the A/B composition of the system. The largest cluster size was computed by averaging over the center of the highest-mass peak in the cluster mass distribution. The averaging was restricted to bins with probability $>$0.05. Figure 6a clearly shows that the size of the largest cluster increases substantially as the fraction of sequence A chains is increased from 0.6 to 0.8. In this region, the system changes in character. The system contains a single dominant cluster if the fraction of sequence A chains is 0.8 or higher, whereas several smaller clusters are present if this fraction is 0.6 or lower. Overall, there is a tendency for the size of the largest cluster to increase as the fraction of sequence A chains is increased. However, for systems with 5—10% sequence A chains, a small peak can be seen, which we believe is statistically significant. Figures 6c and 6d show two examples of what the largest cluster may look like in these systems. In both cases, the cluster is somewhat elongated rather than spherical, with sequence A chains forming a string-like structure on the surface, which may support the formation of a slightly larger cluster.

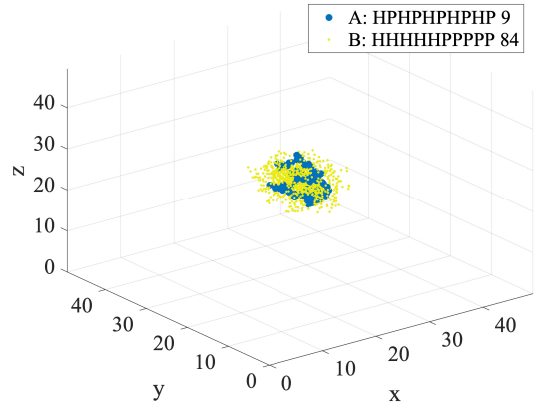
We also studied the bead density of the largest clusters. As discussed in Sec. 3.2.4 below, we computed the average bead density as a function of the cylindrical coordinates r and z. Figure 6b shows the maximum average bead density, as a function of the fraction of sequence A chains. There is a tendency for the bead density to decrease when increasing fraction of sequence A chains. The sequence B-dominated systems have maximum bead densities roughly a factor 2 higher than that of sequence A-dominated systems.



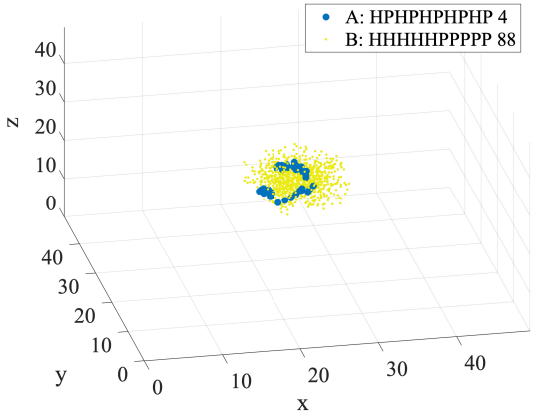
(a) The largest cluster size



(b) The highest density of large cluster



(c) $R_{30/270}$



(d) $R_{20/280}$

Figure 6: Cluster properties in 300-chain two-component A/B systems at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. (a) Average mass of the the largest cluster against the fraction of sequence A chains. (b) The average maximum bead density, as obtained using a grid with $dr = dz = b$, against the fraction of sequence A chains. (c, d) Snapshots of the largest type of cluster in the $R_{30/270}$ and $R_{20/280}$ systems, with sequence A chains in blue and sequence B chains in yellow.

3.2.4 Bead density distributions

In addition to cluster mass, we also studied the distribution of beads within different types of clusters. Following the procedure outlined in Sec. 2.4.3, we computed bead density distributions around the center of mass as functions of cylindrical coordinates, $\rho(r, z)$. This two-dimensional distribution was found more informative than a simple one-dimensional radial distribution, $\rho(r)$, because many of the observed cluster types have an ellipsoidal rather than a spherical shape. We display the computed density distributions as heat maps.

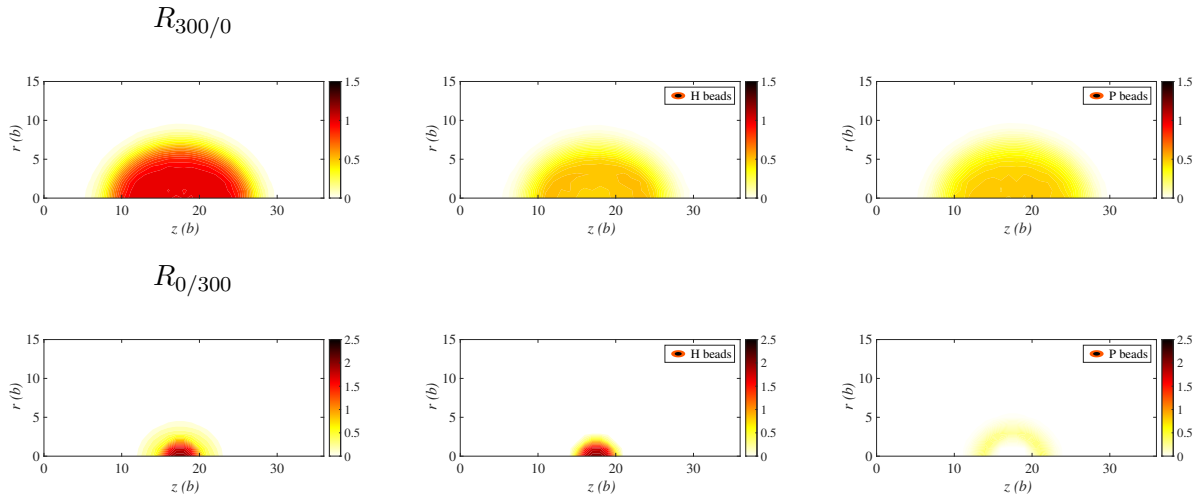


Figure 7: Heat maps showing bead density distributions $\rho(r, z)$, calculated as functions of cylindrical coordinates r and z, for clusters formed by the pure sequence A and pure sequence B systems in 300-chain simulations at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. Left panel: total bead density, mid panel: H bead density, right panel: P bead density.

In addition to the total bead density, we also show the respective contributions from H and P beads in separate figures. Figures 7 and 8 show the results obtained for a few selected systems. Additional bead density distributions can be found in Fig. A3. Figure 7 shows bead density distributions for the large clusters present in the one-component systems $R_{300/0}$ and $R_{0/300}$. In both cases, the clusters are approximately spherical. However, the clusters extend over a larger volume in the pure A system than in the pure B system. This is not surprising given that the pure A clusters have a markedly higher mass. In addition, it turns out that the beads are less densely packed in pure A than in pure B clusters. In fact, the bead density in the core is approximately a factor two higher in pure B clusters than in pure A clusters ($2.12b^{-3}$ versus $1.04b^{-3}$). Another important difference between the two systems is in the distributions of H and P beads. In the pure A system, H and P beads are well mixed, with similar density distributions. By contrast, the pure B clusters have a core dominated by H beads, and a corona dominated by P beads. The shapes of the H and P distributions illustrate the micellar character of the pure B clusters.

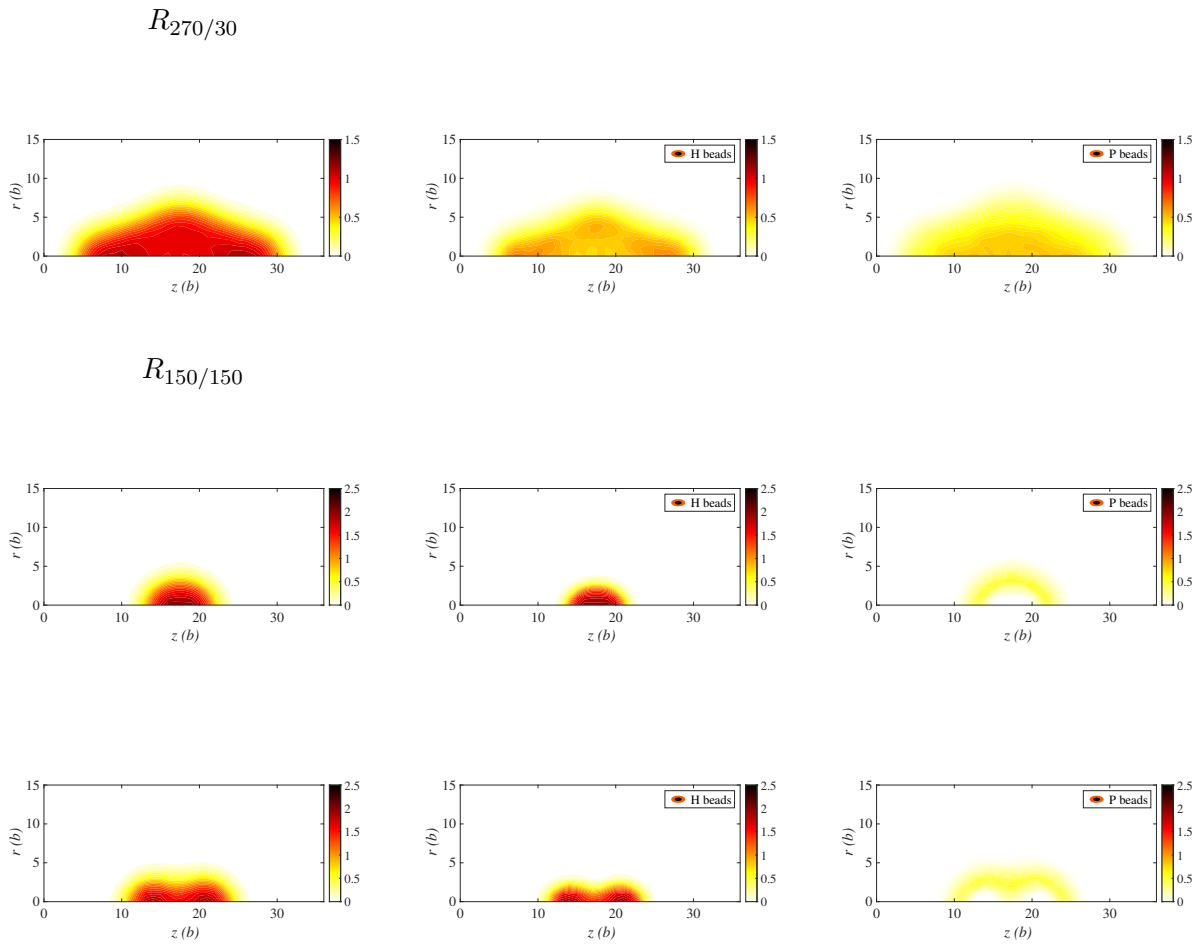


Figure 8: Heat maps showing bead density distributions $\rho(r, z)$, calculated as functions of cylindrical coordinates r and z, for clusters formed by the $R_{270/30}$ and $R_{150/150}$ systems in 300-chain simulations at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. The $R_{150/150}$ data are for two cluster types, corresponding to the third (top row) and fourth (bottom row) peaks in the cluster mass distribution. Left panel: total bead density, mid panel: H bead density, right panel: P bead density.

Figure 8 shows bead density distributions for clusters in the two-component systems $R_{270/30}$ and $R_{150/150}$. The $R_{270/30}$ clusters have a mass similar to that of pure A clusters. The bead density distributions reveal, however, some differences between the $R_{270/30}$ and pure A clusters. One difference is in the overall shape, which is ellipsoidal rather than spherical for the $R_{270/30}$ clusters. Another difference is that H and P are slightly less well mixed in the $R_{270/30}$ clusters. Additionally, the highest H bead density is not found in the core of the $R_{270/30}$ clusters.This behaviour probably reflects that the B chains tend to reside on the surface of the otherwise sequence A-dominated clusters, with their hydrophobic half pointing inwards, as found by visual inspection. The cluster mass distribution indicates

that $R_{150/150}$ system contains four cluster types. The bead density distributions suggest that all four cluster types have a demixed micellar structure, with an H core and a P corona (Figs. 8 and A3). However, the shape of the hydrophobic core varies. The two cluster types with lowest masses both have a single, approximately spherical H core. The third peak in mass corresponds to clusters with an elongated hydrophobic core, which exhibits of two distinct, approximately spherical H-rich regions (Fig. 8). The bead density profiles of the fourth and final cluster type have a similar structure (Fig. A3). However, visual inspection of snapshots from the simulations suggests have the core of these clusters have three spherical H-rich regions, two of which sometimes are in contact. This structure is lost when averaging over all snapshots.

4 Discussion & Summary

In this thesis, we have extended prior work on the phase behavior of two HP sequences [17, 18], called A and B, by investigating the aggregation behavior of mixed two-component systems containing both these sequences. To this end, we conducted MC simulations for 12 different A to B ratios, at a fixed total number of chains, set to 300. Depending on the relative amount of sequence A and B chains, we found that a variety of aggregates can be observed, which differ in size, shape and internal organization.

Generally, sequence A-dominated systems are characterized by the presence of a single dominant cluster with relatively homogeneous distributions of H and P beads, whereas sequence B-dominated systems form several similar-size micellar clusters with a hydrophobic core and a polar surface. The latter clusters are denser than the sequence A-droplets, as reflected in lower potential energy. The behavior of general mixed A/B systems can in part be understood in terms of a competition between these two aggregation mechanisms, sequence A-driven phase separation and sequence B-driven micellization.

The outcome of this competition depends, of course, on the A to B ratio, and we identified three major regimes. First, if the fraction of sequence B chains is small, then a large sequence A-dominated droplet is present, comprising a total of almost 250 chains (Fig. 6a). A small minority of these are sequence B chains, typically located near the surface of the droplet. While the droplet size changes only marginally when adding small amounts of sequence B chains, the shape of the droplet soon turns ellipsoidal rather than spherical. A cluster characteristic of the second regime can be found in the snapshot for the $R_{240/60}$ system in Fig. 4. This cluster is also large, but its structure is different. It contains several sequence B-dominated micellar structures, which are connected into a single large cluster by a surrounding network of mainly sequence A-chains. Moreover, the micellar structures do not fall on a straight line, leading to a curved rather than ellipsoidal overall shape. In the third and final regime, the propensity of sequence B to form micellar structures is the dominant driving force, and the large cluster is split into several micellar structures, with sequence A-chains on their surfaces. Larger clusters with two or more micellar cores

occur as well, if a sufficient amount of "glue" is available, in the form of sequence A chains. Examples of such multi-micelle clusters can be found, for instance, in the $R_{180/120}$ and $R_{150/150}$ simulations (Figs. 4 and A1). Another type of cluster larger than a typical micelle was observed in the $R_{20/280}$ and $R_{30/270}$ simulations (Fig. A1). These clusters have a single but extended (non-spherical) core. We speculate that they appear when the amount of sequence A chains is sufficiently high to permit the formation energetically favorable sequence A-layers around micelles, but insufficient to cover full set of typical spherical micelles.

We note that in pure sequence A systems, at the temperature and volume studied in this paper, droplet formation sets in when the number of chains reaches a value of $\approx$ 40 (corresponding to $\rho_\ell(T)$, see Fig. 2b). Therefore, the formation of pure sequence A droplets is possible down to approximately this number of sequence A chains. However, the results presented suggest that sequence A-dominated clusters are observed only if the number of sequence A chains is $\gtrsim$270. This observation implies that sequence A chains interact strongly with sequence B chains. Sequence A chains tend to form layers surrounding sequence B-dominated micelles rather than pure sequence A droplets. The strong interaction between the two sequences leads to new types of aggregates, not present in the one-component systems.

Simulating the aggregation behavior of systems with many chains is a challenge. In this thesis, for computational reasons, we limited ourselves to systems with 300 chains. We believe that this number of chains is sufficient to get a good overall picture of the character of the aggregates, and how this depends on the A to B ratio. However, for precise statements about asymptotic properties in the large-system limit, the size dependence of the results would have to be assessed. For given values of the temperature, the total density ρ, and the A to B ratio, one would then investigate how the results vary with, say, the volume, while keeping both ρ and the A to B ratio fixed. This type of analysis would in particular be needed in order to conclude whether or not separation into two bulk phases takes place in systems with a small but non-zero B to A ratio. It is also worth noting that the systems studied here are computationally challenging. While we feel confident that our simulations are sufficiently long to provide a reliable overall description of the systems studied, it is possible that, for example, details of the cluster mass distributions, still suffer from some statistical uncertainties.

When decreasing the A to B ratio, we effectively increase the overall degree of blockiness in our system. The finding that a rich variety of aggregates can be observed on the path between the two pure systems that exhibit phase separation and micellization, respectively, matches well with previous results on one-component HP systems with a varying degree of blockiness [24].

Acknowledgments

Foremost, I would like to give my sincere appreciation to my supervisor, Anders Irbäck, for his enlightening guidance in the completion of this dissertation. I have benefited greatly from his weekly instruction, during which I have gained a better understanding of LLPS and developed a strong interest in it. Also, I am grateful to all the teachers in the department of astronomy and theoretical physics, for their insightful lectures which are of substantial help to my research.

In addition, my deep gratitude also goes to my parents, who gave me endless love and financial support for this two-year study and life overseas.

Last but not least, I would like to express my heartfelt thanks to the friends I met in Sweden, for their assistance when I encountered difficulty, for the nights that we were studying together in the library or at home, and for the wonderful time we have had in the last two years.

References

[1] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, 18:285–298, 2017.

[2] 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*, 324:1729–1732, 2009.

[3] K. A. Burke, A. M. Janke, C. L. Rhine, and N. L. Fawzi. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell*, 60:231–241, 2015.

[4] R. K. Das and R. V. Pappu. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.*, 110:13392–13397, 2013.

[5] N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, and T. J. Gibson. Attributes of short linear motifs. *Mol. Biosyst.*, 8:268–281, 2012.

[6] F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Travé, T. J. Gibson, et al. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.*, 13:603, 2008.

[7] K. A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501–1509, 1985.

[8] P. J. Flory. Thermodynamics of high polymer solutions. *J. Chem. Phys.*, 10:51–61, 1942.

[9] M. L. Huggins. Solutions of long chain compounds. *J. Chem. Phys.*, 9:440–440, 1941.

[10] A. Irbäck, N. Linnemann, B. Linse, and S. Wallin. Aggregate geometry in amyloid fibril nucleation. *Phys. Rev. Lett.*, 110:058101, 2013.

[11] H. J. 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*, 495:467–473, 2013.

[12] J. Li, T. McQuade, A. B. Siemer, J. Napetschnig, K. Moriwaki, Y.-S. Hsiao, E. Damko, D. Moquin, T. Walz, A. McDermott, et al. The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell*, 150:339–350, 2012.

[13] Y. R. Li, O. D. King, J. Shorter, and A. D. Gitler. Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.*, 201:361–372, 2013.

[14] Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan. Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.*, 117:178101, 2016.

[15] A. E. Mirsky and L. Pauling. On the structure of native, denatured, and coagulated proteins. *Proc. Natl. Acad. Sci. USA*, 22:439, 1936.

[16] 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*, 163:123–133, 2015.

[17] D. Nilsson and A. Irbäck. Finite-size scaling analysis of protein droplet formation. *Phys. Rev. E*, 101:022413, 2020.

[18] D. Nilsson and A. Irbäck. Finite-size shifts in simulated protein droplet phase diagrams. *J. Chem. Phys*, 154:235101, 2021.

[19] 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, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell*, 57:936–947, 2015.

[20] J. T. G. Overbeek and M. Voorn. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.*, 49:7–26, 1957.

[21] T. Pal, J. Wessén, S. Das, and H. S. Chan. Subcompartmentalization of polyampholyte species in organelle-like condensates is promoted by charge-pattern mismatch and strong excluded-volume interaction. *Phys. Rev. E*, 103:042406, 2021.

[22] R. Phillips, J. Kondev, J. Theriot, H. G. Garcia, and N. Orme. *Physical biology of the cell*. Garland Science, 2012.

[23] Y. Shav-Tal, J. Blechman, X. Darzacq, C. Montagna, B. T. Dye, J. G. Patton, R. H. Singer, and D. Zipori. Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. *Mol. Biol. Cell*, 16:2395–2413, 2005.

[24] A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *J. Chem. Phys.*, 152:075101, 2020.

[25] M. Strzelecka, S. Trowitzsch, G. Weber, R. Lührmann, A. C. Oates, and K. M. Neugebauer. Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nat. Struct. Mol. Biol.*, 17:403–409, 2010.

[26] X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, and R. D. Vale. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*, 352:595–599, 2016.

[27] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86, 1987.

[28] V. N. Uversky. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.*, 114:6557–6560, 2014.

[29] V. N. Uversky, I. M. Kuznetsova, K. K. Turoverov, and B. Zaslavsky. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.*, 589:15–22, 2015.

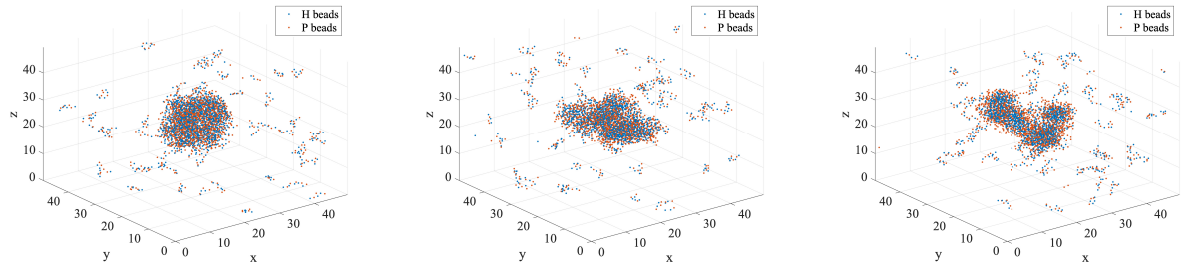
[30] J. Wittmer, A. Johner, and J. Joanny. Random and alternating polyampholytes. *EPL*, 24:263, 1993.

[31] B. Xue, P. R. Romero, M. Noutsou, M. M. Maurice, S. G. Rüdiger, A. M. William Jr, M. J. Mizianty, L. Kurgan, V. N. Uversky, and A. K. Dunker. Stochastic machines as a colocalization mechanism for scaffold protein function. *FEBS Lett.*, 587:1587–1591, 2013.

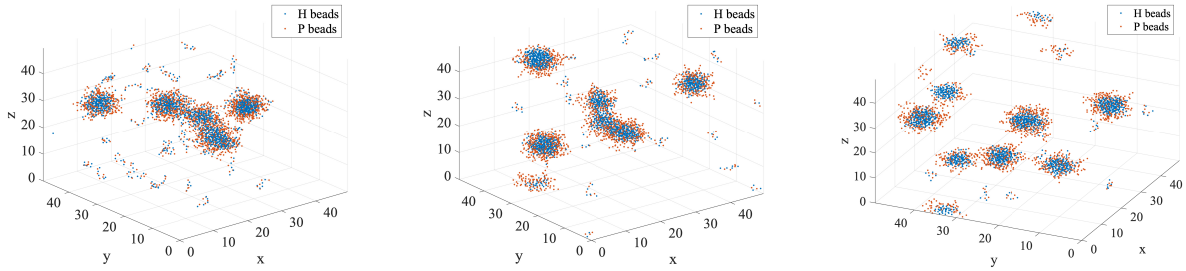
[32] D. Zwicker, M. Decker, S. Jaensch, A. A. Hyman, and F. Jülicher. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proc. Natl. Acad. Sci. U.S.A.*, 111:E2636–E2645, 2014.

A Appendix

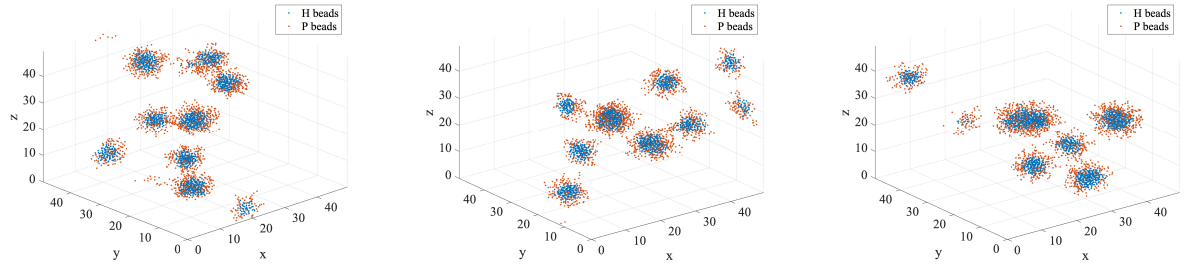
(a) $R_{300/0}$ (b) $R_{270/30}$ (c) $R_{240/60}$



(d) $R_{180/120}$ (e) $R_{150/150}$ (f) $R_{120/180}$



(g) $R_{60/240}$ (h) $R_{40/260}$ (i) $R_{30/270}$



(j) $R_{20/280}$ (k) $R_{10/290}$ (l) $R_{0/300}$

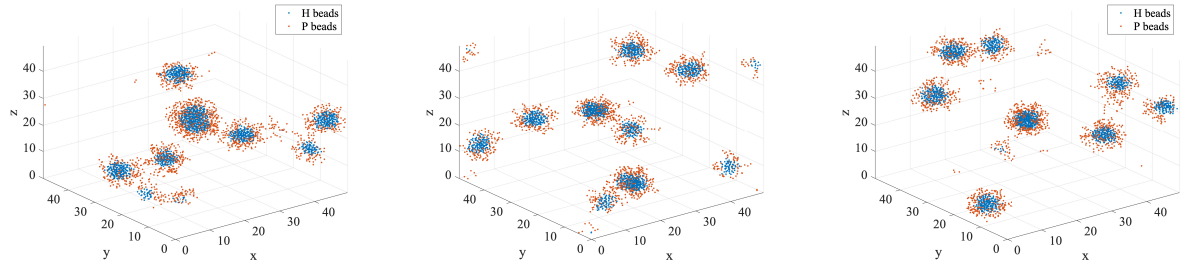


Figure A1: Representative snapshots from all simulations with 300 HP chains for different A to B ratios at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. The blue dots represent hydrophobic beads and the orange dots indicate polar beads. Snapshots were produced using MATLAB.

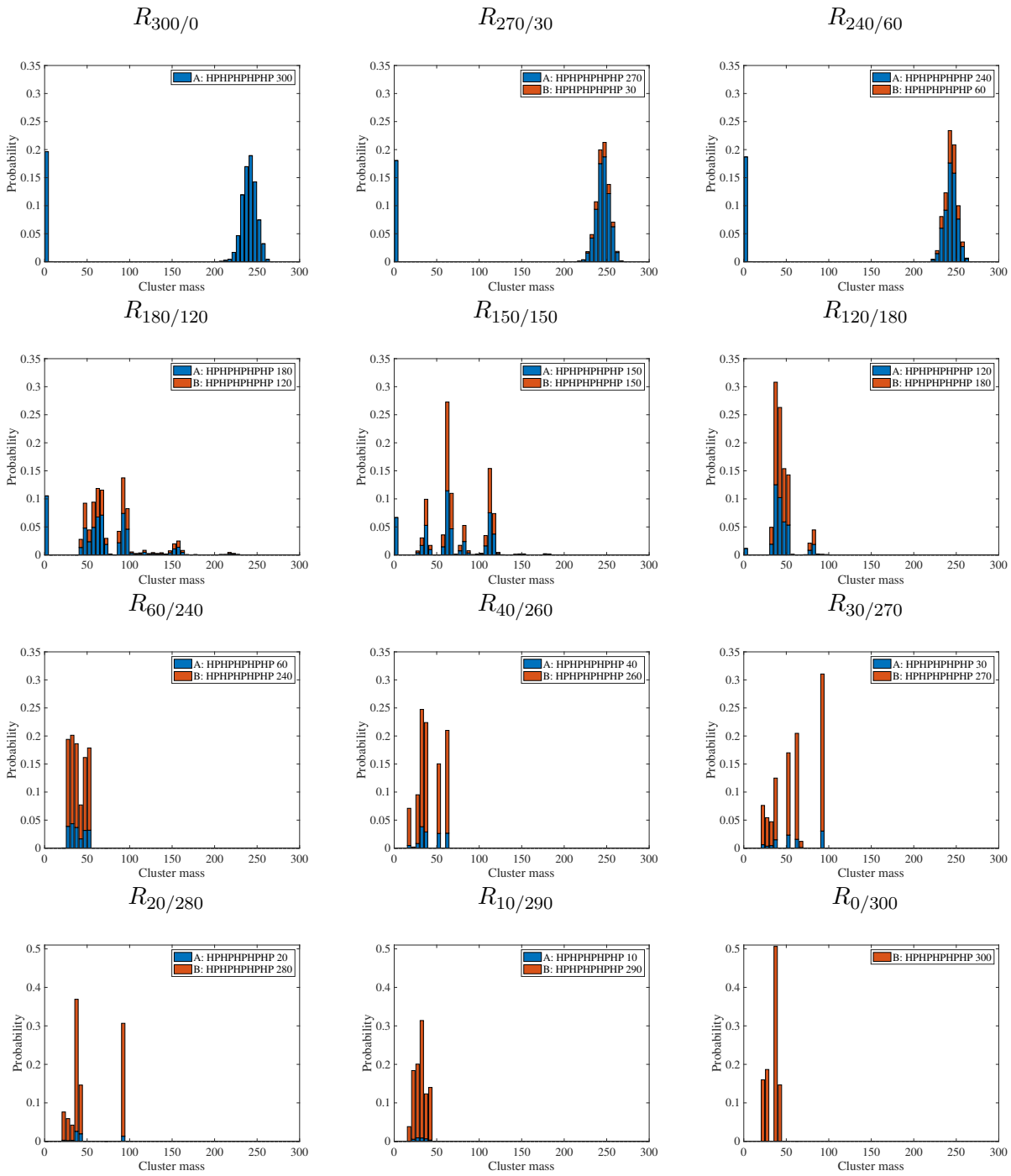
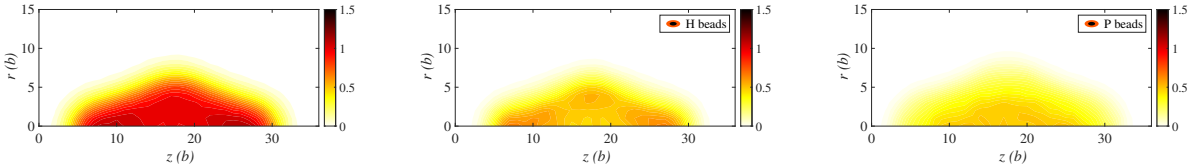
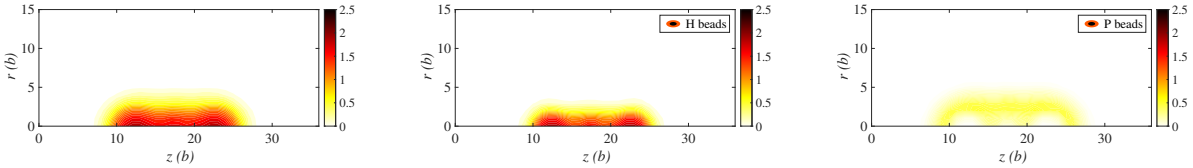


Figure A2: Cluster mass distribution in all simulated systems with different A to B ratios.The mass of a cluster is defined as the number of chain it contains. Each system was simulated at $T = 2.6\,\epsilon/k_B$ and fixed total bead-density $\rho = 0.025\,b^{-3}$. The total probability of a given mass was partitioned into contributions from sequence A (blue) and sequence B (orange), respectively. In each system, there are two or more typical cluster sizes, correspomding to distinct spikes in the distribution.

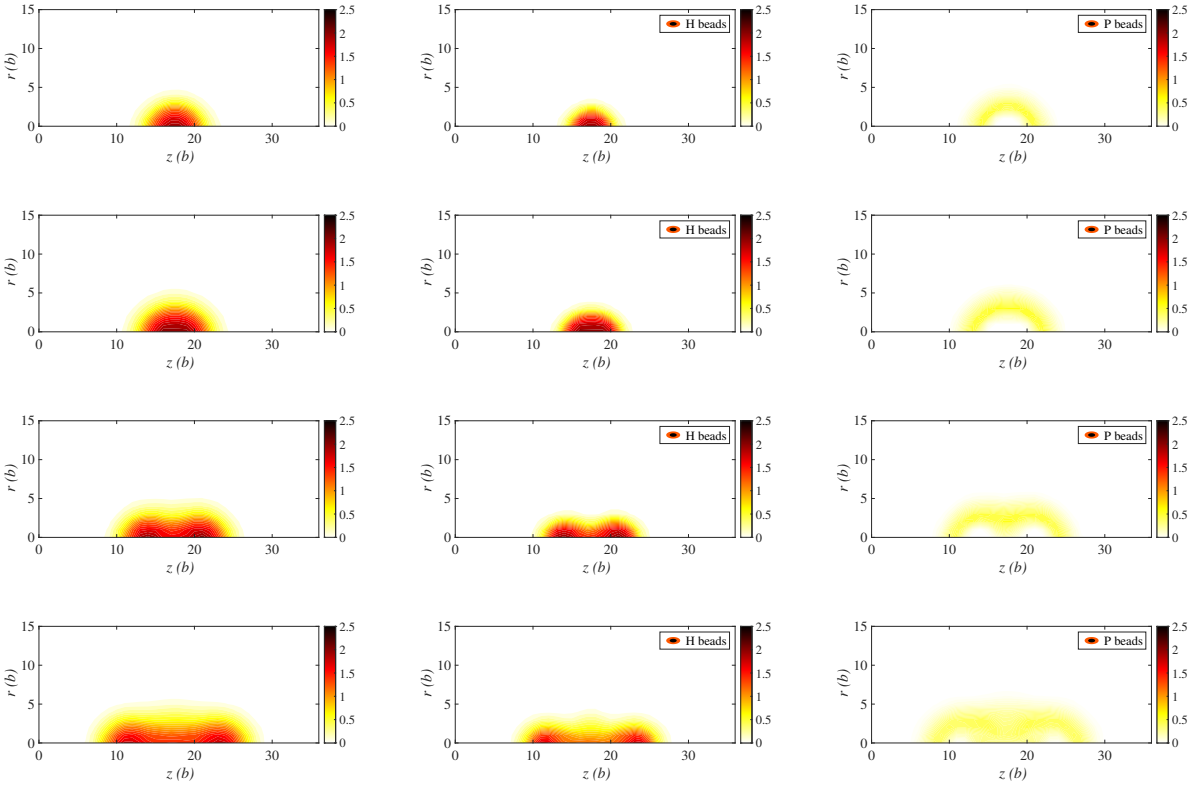
$R_{270/30}$



$R_{180/120}$



$R_{150/150}$



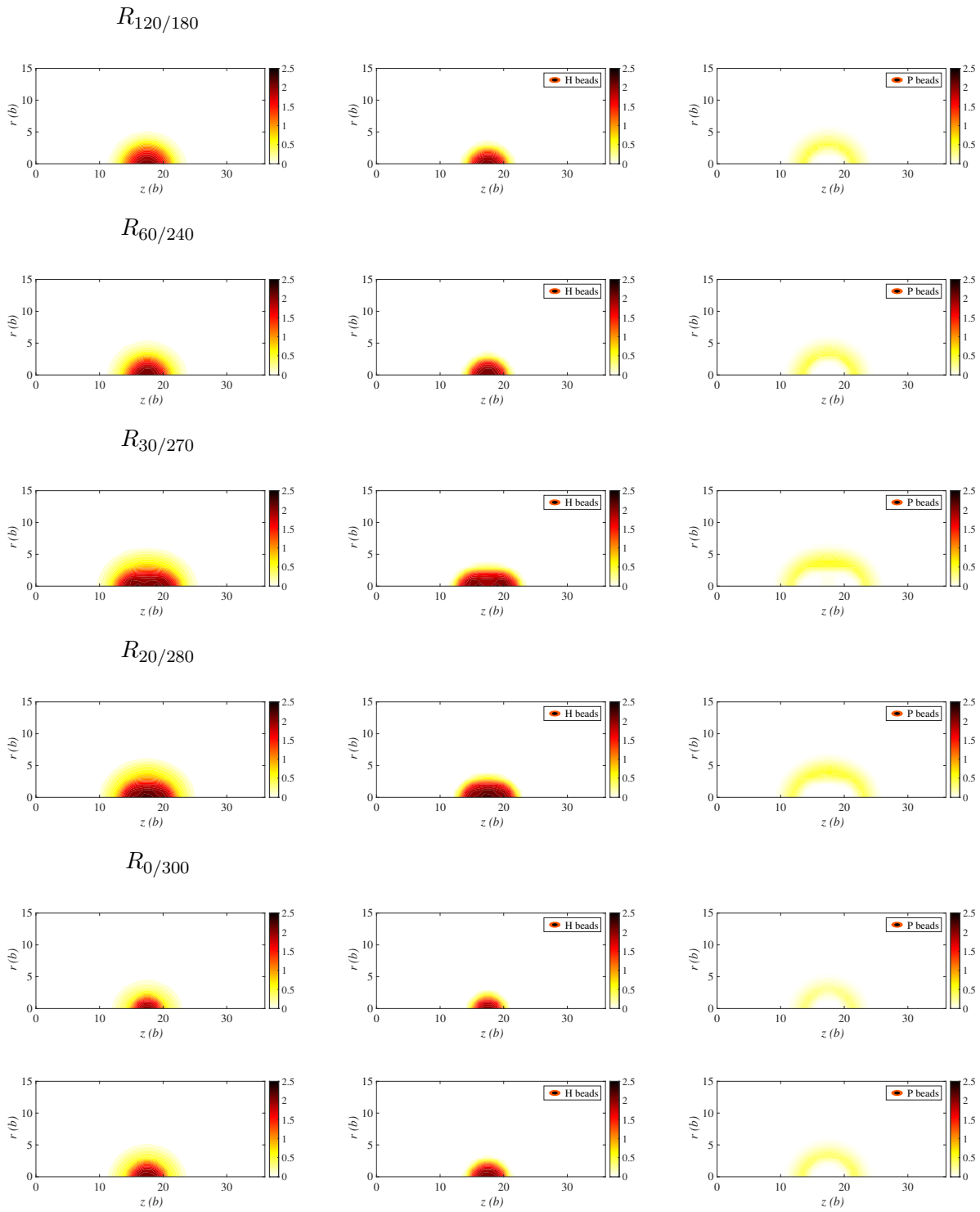


Figure A3: Heat maps showing bead density distributions $\rho(r, z)$, calculated as functions of cylindrical coordinates r and z, for clusters formed by all A/B systems in 300-chain simulations at $T = 2.6\,\epsilon/k_B$ and $\rho = 0.025\,b^{-3}$. Left panel: total bead density, mid panel: H bead density, right panel: P bead density.