Xinjiang Lü

Monte Carlo simulation and theoretical analysis of self-assembled semiflexible equilibrium polymers

By

Xinjiang Lü

Doctor of Philosophy

Department of Chemistry

Dr. James T. Kindt
Advisor

Dr. Keiji Morokuma
Committee Member

Dr. Michael C. Heaven
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Gradutate School

Date

Monte Carlo simulation and theoretical analysis of self-assembled semiflexible equilibrium polymers

By

Xinjiang Lü

B.S., University of Science and Technology of China, 2002

Advisor: Dr. James T. Kindt, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the Graduate School of

Emory University in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

2007

Abstract

This work is focused on Monte Carlo simulation and theoretical analysis of self-assembled semiflexible equilibrium polymers using F-actin as a model system.

Grand canonical Monte Carlo simulations of a simple model semiflexible equilibrium polymer system, consisting of hard sphere monomers reversibly self-assembling into chains of arbitrary length, have been performed using a novel sampling method to add or remove multiple monomers during a single MC move. Systems with two different persistence lengths and a range of bond association constants have been studied. We find first-order lyotropic phase transitions between isotropic and nematic phases near the concentrations predicted by a statistical thermodynamic theory, but with significantly narrower coexistence regions. A possible contribution to the discrepancy between theory and simulation is that the length distribution of chains in the nematic phase is bi-exponential, differing from the simple exponential distribution found in the isotropic phase and predicted from a mean-field treatment of the nematic. The additional short length-scale characterizing the distribution appears to arise from the lower orientational order of short chains. The dependence of this length-scale on chemical potential, bond association constant, and total monomer concentration has been examined.

Using grand canonical Monte Carlo simulations we study the equilibrium properties of actin self-assembly. The statistics of actin polymerization are described by a mechanism involving monomer activation and chain propagation with bond association constants derived from experimental free energy parameters. For efficiency in representing systems of extremely long, stiff chains we use a coarse-graining based on spherocylinders. We present results pertaining to the isotropic-nematic transition in this equilibrium biopolymer system.

We have also used Monte Carlo simulations to study the bundle formation in self-assembled semiflexible chain polymers with inter-chain attractions. Approximate phase diagrams are obtained for varied physical parameters, such as the chain flexibility, bonded and non-bonded interactions. The attraction induced microphase separation results in an equilibrium between a bundle and isotropic short chains. The chain length distributions of the phase separated system, as well as the bundle's shape and aspect ratio, etc. are presented and discussed. Our simulation results are analyzed and compared with related experimental and theoretical work. We also observed toroids and branched bundles during the course of our simulations.

Monte Carlo simulation and theoretical analysis of self-assembled semiflexible equilibrium polymers

By

Xinjiang Lü

B.S., University of Science and Technology of China, 2002

Advisor: Dr. James T. Kindt, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of
Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Chemistry

2007

Acknowledgments

First and formost, I would like to thank my advisor, Professor James T. Kindt, for his patient and insightful guidance during the course of my graduate studies. His support, training, and trust is more than essential to this dissertation, and will be forever appreciated.

I am grateful to Professor Keiji Morokuma and Professor Michael C. Heaven. They kindly serve as my committee members and mentors, and went out of their way to review my reports and give my constructive comments.

I acknowledge all members of Dr. Kindt's group. Especially, I would like to thank Dr. Jason de Joannis and Dr. Yong (Frank) Jiang for their invaluable help since I joined the research group. I thank Mr. Fuchang Yin and Dr. Hao Wang for their advice on bundle analysis in Chapter 4. My thanks also go to other past and present members for their constructive advice and interesting discussions in the past five years.

I would like to express my gratitude to my colleagues at Emerson Center and friends across the campus for their help with my research during the last five years.

Lastly, I can not thank my grandparents enough for their selfless sacrifice and unconditional support. I am also extremely grateful to the rest of my family, especially my father and my little aunt. I was very lucky to meet Krystal Lin, whose love and encouragement has supported me through my years at Emory. This dissertation is dedicated to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computer simulation plays an important role in modern scientific research. With a carefully constructed model, we can carry out rigorous experiments on computers, thus greatly enhance our understanding in various research areas such as materials and biological molecules. There are two main classes of classical computer simulation techniques: molecular dynamics (MD) and Monte Carlo (MC) [1].

In molecular dynamics simulations, we study how molecules move around with real time. The assumption here is that molecules obey Newtonian laws with quantum effects neglected. Starting from an initial configuration, the simulation system gives rich dynamics information of our interest. For an equilibrium system, the average over a long period of time is equal to the ensemble average.

The Monte Carlo method has a different approach, where we directly sample the configuration space and obtain the integration using pseudo random numbers. Depending on the form of the integrand $f(x)$ in the integral

$$\int_{x_1}^{x_2} \mathrm{d}x f(x), \tag{1.1}$$

various techniques have been invented to speed up the sampling efficiency. In the context of physics and chemistry, the integrand contains an exponential term because the system satisfies the Boltzmann distribution. The Metropolis algorithm was devised to efficiently sample such a distribution, which has been incorporated in most modern Monte Carlo algorithms [1, 2].

In this dissertation, we use Monte Carlo methods to study the self-assembly of semiflexible equilibrium polymers, i.e. chains formed from equilibrium association of monomers. Properties of our interest include, but not limited to, the isotropic-nematic phase transition, the polymer chain length distribution, the cross-linked network, and inter-chain attractions induced bundling.

1.1 Monte Carlo methods and ensembles

While we practice Monte Carlo simulations, we have one essential assumption that our pseudo random number generator (RNG) gives enough random numbers. Practically, if the period of the RNG is far greater than the number of Monte Carlo steps, we are confident about the randomness of our simulation. Therefore, in the following we focus on the Monte Carlo algorithms and related simulation techniques, which determines the efficiency of the simulation given our limited computational capacity.

To numerically integrate a definite integral [1, 3], we can evaluate the integrand $f(x)$ with x randomly sampled from the integration interval $[x_1, x_2]$. The integral then can be approximated by

$$I = (x_2 - x_2)\frac{\sum f(x_i)}{N}, \tag{1.2}$$

where the second term is the arithmetic average with N is sample size. If N is large enough, this formula is a very good approximation of the integral. The drawback of this brute sampling is its inefficiency for a complex $f(x)$ because this method may waste a lot of time in sampling insignificant points, that is, relatively small $f(x_i)'s$. It will be much more advantageous if the sampling algorithm favors big $f(x)$ values rather than small $f(x)'s$. One method is to utilize the inverted cumulative distribution

$$x = g^{-1}(y), \tag{1.3}$$

if we know $g(x) = \int_{x_1}^{x} \mathrm{d}x f(x)$. Through this formula, we map a uniform probability distribution on $[0, 1]$ onto a nonuniform probability distribution of $x's$ based on $f(x)$. The resulting integral is

$$I = \frac{1}{N}\sum f(x'). \tag{1.4}$$

Because it is usually difficult to obtain the inverted function, a generalized form of this method is to devise a weight function $w(x)$ to bias the sampling process,

$$I = \int_{x_1}^{x_2} \mathrm{d}x \; w(x) \frac{f(x)}{w(x)}. \tag{1.5}$$

The conventional rejection method chooses $f_{\max}(x)$ as $w(x)$ and is not efficient when $f(x)$ is very “spiky”. To work around the disadvantages, the more advanced hybrid method factors $f(x)$ into two parts: one invertible ”spiky” function and the other relatively smooth one [3].

1.1.1 Metropolis algorithm

In classical statistical mechanics, we are mostly interested in the ensemble average of an observable A [1, 4]:

$$\langle A \rangle = \frac{\int \mathrm{d}\mathbf{r}^N \; A(\mathbf{r}^N) \exp[-\beta U(\mathbf{r}^N)]}{\int \mathrm{d}\mathbf{r}^N \; \exp[-\beta U(\mathbf{r}^N)]}, \tag{1.6}$$

where $\beta = 1/(K_{\mathrm{B}}T)$, the momentum part of the phase space has been integrated out, and the observable is integrated over the configuration space of $3N$ dimensions. If the denominator, i.e., the configuration integral in Eq. (1.6) is denoted by Z_N, then the ensemble average can be rewritten as

$$\langle A \rangle = \int \mathrm{d}\mathbf{r}^N \; P(\mathbf{r}^N) A(\mathbf{r}^N), \tag{1.7}$$

where the probability density function $P(\mathbf{r}^N) = \exp[-\beta U(\mathbf{r}^N)]/Z_N$. Since the configuration integral of a many body system is not solvable analytically, we usually use computational methods, such as Monte Carlo sampling, to sample $P(\mathbf{r}^N)$ and evaluate variable properties instead of Z_N itself.

The general approach is that we sample the phase space by constructing a sequence of random configurations from the distribution, each of which is only dependent on the previous one. Such a sequence is known as a Markov chain. A proposed Monte Carlo scheme should not change the equilibrium distribution: in our case, the Boltzmann distribution. The equilibrium (or "balance") condition requires that the probability of leaving a configuration be equal to the probability of entering a configuration (or equal flux). We can apply a stricter detailed balance condition: equal flow between two configurations as follows

$$P(i) \times \alpha(i \to j) \times acc(i \to j) = P(j) \times \alpha(j \to i) \times acc(j \to i), \tag{1.8}$$

where i and j are two adjacent configurations in a Markov chain, α is the transition attempt probability between the two configurations, and *acc* is the probability of accepting such a transition.

1.1.2 Biased sampling methods

In the classical Metropolis algorithm, we have $\alpha(i \to j) = \alpha(j \to i)$, that is, we do not bias either transition direction. Then we have

$$\frac{acc(i \to j)}{acc(j \to i)} = \frac{P(j)}{P(i)} = \exp[-\beta(U_j(\mathbf{r}^N) - U_i(\mathbf{r}^N))]. \tag{1.9}$$

To improve its efficiency, later proposed Monte Carlo algorithms usually have asymmetric α. These biased sampling techniques are especially useful in simulations of polymers. For example, in the orientational bias sampling, a new bond angle is chosen not randomly but to favor the one lowering the total energy, then the acceptance probability is modified accordingly to satisfy the detailed balance. In the more advanced configurational bias method [5], the Rosenbluth factors [6] are computed for both the old and new configurations in a Monte Carlo step as follows:

$$W(n) = \prod_{i=1}^{l} w_i(n), \tag{1.10}$$

where l is the polymer chain length and $w_i(n)$ is

$$w_i(n) = \sum_{j=1}^{k} \exp[-\beta u_i(j)]. \tag{1.11}$$

Note that $w_1(n)$ for the first monomer is different from the rest ones of the polymer chain of length l. The number k is the total number of allowed growth sites for each step i and depends on the choice of lattice in simulation as well as the author's choice. $u_i(j)$ is the energy of each allowed growth site j in each step i. The weight for the old configuration, $W(o)$, is calculated in a similar way except that the configuration is already known, while for the new configuration the choice of each new monomer is determined by $\exp[-\beta u_i(j)]/w_i(n)$. Based on the condition of detailed balance (1.8), the acceptance probability is

$$acc(o \to n) = \min[1, W(n)/W(o)], \tag{1.12}$$

because the old and new energy terms: $\exp[-\beta U(n)]$ and $\exp[-\beta U(o)]$ are cancelled through both sides, respectively, where U is the energy of the configuration. This method can also be generalized to off-lattice models, where the discrete number of trial moves for each monomer becomes a continuum space.

1.1.3 Ensembles

The design of practical Monte Carlo moves in large part depends on the choice of simulation ensemble. According to the equation of state, we need three state variables to determine the state of a system. Some common ensembles in simulation are isobaric-isothermal ensemble (constant-NPT), canonical ensemble (constant-NVT) and grand canonical ensemble (constant-μVT) [1]. From different ensembles we should be able to obtain the same macroscopic properties provided that the Monte Carlo simulation is ergodic and reaches the equilibrium*. The choice then largely depends on whether it is convenient to study properties of our interest. For example, the NPT ensemble is an ideal choice since many experiments are carried out this way. Note that in this ensemble the volume V can change in accordance with the system change, which makes possible the complete phase transition in the simulation. Similarly, the μVT allows for the density (or number of particles N) change in the simulation while the chemical potential μ remains constant. The grand canonical ensemble is usually a natural choice

---

*This is the case in the limit of long simulations. The differences can be as small as possible with increasing computational capacity and efficient algorithms for the adsorption/desorption simulation. However, in some cases we are not interested in the complete phase transition, then we can choose other more convenient ensembles. In the conventional canonical ensemble, we have fixed the density (constant N and V), therefore during a phase transition the system will phase separate into two physically different parts, though this is not the true phase separation in the experimental sense. Using the periodic boundary condition, we replicate the system around the simulation box to reduce the finite size effect. This method works fine with a homogeneous system. With a fixed density, the finite simulation size prevents the system from undergoing a true phase separation. In this work, we focus on the application of NVT and μVT ensembles, which serve our purposes depending on our research interests.

In the past decades, numerous Monte Carlo moves were designed specifically for simulation of polymers. Mostly, we are interested in the phase transition phenomena, such as the glass transition or the melting process. In these cases, the density is so high that simple translational or rotational motions of a whole polymer are rarely successful. To improve the efficiency, we can design canonical ensemble MC moves to target the relaxation of local polymer structures such as microrelaxation moves: L-flips, U-flips, and end-group rotations [7]. These moves were original devised for lattice models, and later were extended into off-lattice models. For example, L-flips can be modified to be crankshaft moves at arbitrary angles. To move a whole polymer chain, the reptation is a snake-slithering-like motion, which can still be very effective for highly entangled polymers [8]. For self-assembling polymers, single monomer move is customarily carried out, though this move is not very efficient. To achieve equilibrium polymers, we can also have chain cleavage or fusion moves, which are very useful especially in lattice model simulation [9].

1.2 Actin protein: a model system for polymers

Actin protein is the key component of cytoskeleton and plays important roles in various cellular functions, such as cell shape and cell motility [10, 11]. The monomer form of actin protein, called G-actin (Globular-actin), is essentially a polypeptide chain consisting of approximately 275 amino acid residues. The sequence is highly conserved across species [12].

Under certain conditions, such as salt, ATP/ADP and pH, G-actin monomers reversibly assemble into long polymer chains, F-actin (filamentous-actin), which results in a polydisperse chain polymer system [11]. The actin filament, a double stranded helical polymer chain with a pitch of $\sim$ 72 nm, has a diameter of 7 $\sim$ 8 nm and can be as long as several microns [10]. With the regulation of cross-linking proteins, polycations and inert polymers, actin filaments can further form bundles (including rings) or cross-linked networks [13–15]. The mechanism of the polymerization itself remains unclear, though varied models have been proposed to explain the kinetics and thermodynamics. Generally actin polymerization is believed to involve three basic steps: monomer activation, nucleation (a possible trimer nucleus) and propagation. For exmaple, Niranjan et al. [16, 17] measured the equilibrium extent of polymerization and fit the data to the formula obtained from a proposed polymerization mechanism, where the authors extracted free energy parameters for the three polymerization steps. G-actin monomer-monomer association involves structural changes [12], and the bonding is noncovalent and a complex combining result of hydrophobic/hydrophilic interactions, electrostatic interactions, and hydrogen bonding [18, 19]. Computer simulations combined with *in vitro* experiments have been applied to study the polymerization, specifically the nucleation pathway problem [20, 21].

In this work we focus on the equilibrium problems of actin polymerization. Both experiments and theories show that the length distribution obeys the exponential law. Here we assume the association constants for all steps are indentical, i.e., the so-called isodesmic model. Some discrepancies regarding the distribution in *in vitro* actin experiments have been attributed to facts such as nonequilibrium and low measurement resolution. With a concentration above ~ 2 mg/ml, the actin filaments transition disordered to ordered, namely isotropic-nematic phase transition [22, 23]. This phenomenon has been successfully explained in the seminal work by Onsager [24] and other further work [25, 26] on rigid monodisperse rods. Later efforts were focused on incorporating the unique semiflexibility feature of F-actin-like biopolymers [27, 28], which is different from organic polymers such as polyethylene (which is very flexible). The polydispersity of the length distribution also has a strong influence on the phase behavior. Early simulations of semiflexible equilibrium polymers were carried out on a cubic lattice model [29], later more realistic continuum (or off-lattice) models were used to simulate the self-assembly and phase behavior of such a system [30, 31]. The length distribution was generally found to follow an exponential probability distribution. However, there has been no further examination of the distribution's dependence on different macroscopic phases, where most studies are focused on the relationship between the mean chain length (or aggregate size) and the adjustable physical parameters.

However, under real biological conditions, a myriad of third party proteins regulate the growth of actin filaments [32], therefore the length distribution can be significantly different from equilibrium polymers. For example, the phalloidin can be added into the actin solution to bind acitn and prevent its depolymerization. Gelsolin has a function of severing actin filaments and is widely used to regulate the filament length in actin experiments [22]. Also the cell is an confined environment rather than an infinite macroscopic space, the effect of which has been studied *in vitro* in a cell-sized water droplet coated with a phospholipid membrane as a simple model of a living cell membrane. [33]

1.3 Thesis outline

The rest of this thesis is organized as follows.

In Chapter Two, we apply a novel sampling method to grand canonical Monte

Carlo simulations of semiflexible equilibrium polymers. We show that the first-order isotropic-nematic phase transitions have significantly narrower coexistence regions compared with theoretical predictions based on a mean field statistical thermodynamic theory, which is partially due to the unique "bi-exponential" distribution in the nematic phase arising from the lower orientational order of short chains. We further show that a theory representing the nematic phase as an equilibrium mixture of randomly oriented chains below a critical length L_c and chains obeying a Gaussian orientational distribution above L_c reproduces this distribution qualitatively, whereby the agreement between predicted and simulated phase boundaries is improved, with a narrower isotropic-nematic coexistence region, compared with a monoexponential chain distribution.

In Chapter Three, we study the phase transition of actin filaments with association constants indirectly obtained from experiments based on a model where the actin polymerization involves three steps: monomer activation, nucleation, and propagation. The limit case of large association constants in the isodesmic model is also examined. In this work, a long chain is modeled as a chain of jointed spherocylinders to reduce the computational expense while maintaining the chain length dependence on the association constants as in the hard sphere chain model.

In Chapter Four, we use Monte Carlo simulations to study the phase behavior and structures of self-assembled semiflexible chain polymers with inter-chain attractions. Approximate phase diagrams are obtained for varied physical parameters, such as the chain flexibility, bonded and non-bonded interactions. The attraction induced microphase separation results in an equilibrium between a bundle and isotropic short chains. The chain length distribution of the phase separated system, as well as the bundle's shape and aspect ratio, etc. are presented and discussed. Our simulation results are analyzed and compared with related experimental and theoretical work. Other structures formed during the course of our simulations, such as toroids and branched bundles, are also presented.

Chapter 2

Monte Carlo simulation of the self-assembly and phase behavior of semiflexible equilibrium polymers

Grand canonical Monte Carlo (MC) simulations of a simple model semiflexible equilibrium polymer system, consisting of hard sphere monomers reversibly self-assembling into chains of arbitrary length, have been performed using a novel sampling method to add or remove multiple monomers during a single MC move. Systems with two different persistence lengths and a range of bond association constants have been studied. We find first-order lyotropic phase transitions between isotropic and nematic phases near the concentrations predicted by a statistical thermodynamic theory, but

with significantly narrower coexistence regions. A possible contribution to the discrepancy between theory and experiment is that the length distribution of chains in the nematic phase is bi-exponential, differing from the simple exponential distribution found in the isotropic phase and predicted from a mean-field treatment of the nematic. The additional short length-scale characterizing the distribution appears to arise from the lower orientational order of short chains. The dependence of this length-scale on chemical potential, bond association constant, and total monomer concentration has been examined. To better elucidate this phenomenon, we show that a theory representing the nematic phase as an equilibrium mixture of randomly oriented chains below a critical length L_c and chains obeying a Gaussian orientational distribution above L_c reproduces this distribution qualitatively. The agreement between predicted and simulated phase boundaries is improved, with a narrower isotropic-nematic coexistence region, compared with a monoexponential chain distribution. We find, furthermore, that the critical length L_c scales as the inverse of monomer concentration, irrespective of the bond strength of the chains, but that at the phase boundary, the fraction of monomers found in disordered chains peaks at a certain bond strength.

2.1 Introduction

A recently developed Monte Carlo simulation algorithm for treating equilibrium self-assembled structures within the grand canonical ensemble has been applied to the reversible self-assembly of particles into chains, i.e., equilibrium polymerization. Using this "polydisperse insertion, removal, and resizing" (PDIRR) algorithm, the isotropic-nematic transition of semiflexible equilibrium polymers has been studied over a range of parameter space and compared to predictions of analytic theory. Experimental examples of equilibrium polymer systems that exhibit an isotropic-nematic transition include wormlike micelles [34] and self-assembled protein fibers like F-actin [22, 35]. The goal of the present study is not to model a particular system, but to test the predictions of analytical theory and explore the general nature of the phenomenon.

The spontaneous phase transition of anisotropic hard-wall particles to an orientationally ordered nematic liquid crystalline phase due to excluded-volume interactions was first given a theoretical underpinning by Onsager citeonsager49. Flory [36] used a lattice model to predict the effects of internal degrees of freedom on this transition for semiflexible polymers; later work by Grosberg, Khokhlov, and Semenov [37, 38] and by Odijk [25] developed approaches based on a second virial approximation. Simulation studies of ordering in semiflexible polymer models have been performed to test these theories [39, 40]. Following early work by McMullen et al. [41], the coupling of equilibrium polymerization to nematic ordering has been studied analytically by a number of authors [27, 42–45]. Many of the complicating details– internal flexibility, end effects, treatment of excluded volume beyond the second virial approximation, functional optimization of orientational distributions have been treated separately, but simulation is still the only practical way to address the overall problem of a polydisperse, equilibrium system of chains with internal flexibility at moderate packing fractions. Lattice simulations of the I-N transition have shown strong artifacts relating to the discreteness of positions and orientations [29, 46]. Recent off-lattice simulation studies by Fodi and Hentschke [30] and by Chatterji and Pandit [31] have confirmed the qualitative features predicted by mean-field theory, including the first-order nature of the transition in 3-dimensional systems, as well as the significant increase in average chain length that accompanies nematic ordering. The present study uses biased, multistep grand canonical MC moves with the goal of rapidly equilibrating large systems ($5000 - 25000$ monomers) and determining phase coexistences as precisely as possible over a range of physical parameters. The model for monomer association was chosen to allow independent selection of chain flexibility and the association strength of the chaining interaction, facilitating comparison with theoretical predictions.

2.2 Methods

2.2.1 Model description

We begin with an association constant for the bonding of a hard-sphere monomer to another monomer or to the end of a chain of bound monomers, $K_{\text{assoc}} = \exp(\beta E_{\text{bond}})$, with $\beta = (k_B T)^{-1}$ and E_{bond} the bond free energy. The association constant is defined with respect to a reference system of ideal monomers and chains lacking hard-sphere excluded volume interactions, with the standard reference state of an n-mer defined at a concentration of one n-mer per volume σ^3.[*] In the present work, an "isodesmic" model is used in which K_{assoc} is independent of the length of the chain. At equilibrium, the chemical potential of an n-mer in the reference system is:

$$\mu_n = n\mu - (n-1)A_{\text{bond}} = n\mu - (n-1)k_B T \ln K_{\text{assoc}}, \tag{2.1}$$

where $\mu = \mu_1$, the chemical potential of a free monomer.[†] In the present model, ring structures are not permitted.

Bonds between neighboring monomers in a chain are constrained to a length equal to the hard sphere diameter σ. The bending potential for the angle between any three consecutive bound monomers is an infinite square well: $U(\theta) = \infty$ for $\theta < \theta_{\text{min}}$, $U(\theta) = 0$ for $\theta \geq \theta_{\text{min}}$. The same potential was used in a previous study of equilibrium network formation [49]; the present case differs only in that no junctions or branchpoints are permitted.

---

[*] Depending on context, one may consider K_{assoc} as a dimensionless value or as an inverse concentration with units σ^3.

[†] In systems with fixed polydispersity, relationships between chemical potentials of different sized chains are not known a priori; an iterative or adaptive extension [47, 48], of the present method would be necessary to treat such cases to result in a predetermined composition.

2.2.2 Polydisperse insertion, removal, and resizing (PDIRR) moves

Because part of the motivation for this study was to demonstrate the consistency and utility of the PDIRR algorithm, Monte Carlo sampling was performed primarily through biased chain addition, removal, and resizing moves. A general description of the PDIRR algorithm has been reported recently [50]. Here, specifics of implementation and justification of the method for a semiflexible equilibrium polymer model will be described.

The basis of the PDIRR method is that insertion of an aggregate containing an arbitrary number of monomers, or the change in an aggregate's size by an arbitrary number of monomers, can be attempted within a single move. Chain addition or chain removal/resizing moves are selected with equal probability. The first step within a trial addition move is the insertion of a hard-sphere monomer at a random position within the simulation box. If the distance between the trial monomer and any existing particle in the system is less than the hard-sphere diameter σ, the move fails. Otherwise, in subsequent steps a chain is grown unidirectionally from this monomer until a hard-sphere overlap is obtained or the maximum number of particles allowed in the system is reached. For the next step within the chain addition move, a trial particle (index $n = 2$) is added to a randomly selected position on the sphere of radius σ centered at the first particle (index $n = 1$). For steps to add particles $n = 3$ and above, the next trial particle is added to a randomly chosen position on the truncated section of the sphere of radius σ centered on trial particle n-1 that satisfies a minimum bond angle condition, $(\mathbf{r}_n - \mathbf{r}_{n-1}) \cdot (\mathbf{r}_{n-1} - \mathbf{r}_{n-2}) > -\cos\theta_{\min}$. If the jth step results in a hard-sphere overlap, the polydisperse insertion algorithm dictates that the addition of one of the intermediates in the growth of the chain, from a monomer to an $(j-1)$-mer, be attempted.[‡] The acceptance probability for this chain addition is given by $\min[1, W]$ where:

$$W = \sum_{n=1}^{j-1} \omega_n \tag{2.2}$$

and ω_n represents a grand canonical weight for the n-mer, having the form used in standard GCMC:

$$\omega_n = \frac{1}{N_{\text{ch}} + 1} V \exp(\beta\mu_n) = \frac{V}{N_{\text{ch}} + 1} \frac{1}{K_{\text{assoc}}} [K_{\text{assoc}} \exp(\beta\mu)]^n \tag{2.3}$$

where N_{ch} represents the number of chains of all sizes already present in the system, including "chains" of length 1, and μ_n is defined in Eq. (2.1). If the move is to be accepted, then one of the chain growth stages is selected for insertion; the aggregation number n of the chain to be inserted is selected at random with a probability ω_n/W.

To satisfy detailed balance, the factor W must be calculated in a similar manner for the removal/resizing step. First, a chain of arbitrary length is selected at random with equal probability from among the aggregates in the system, whose number we will call $N_{\text{ch}} + 1$ for consistency with the addition step. To calculate W for this n-

---

[‡] A neighbor list of particles within a distance 9 σ of the trial particle is generated at the first monomer's insertion, to be updated every eight steps thereafter.

mer, a series of "dummy" extension steps to a randomly selected end of the chain are performed, until a hard-sphere overlap is encountered when the attempt is made to extend the chain to an $(n'+1)$-mer. The weight W is the sum of weights from ω_1 to $\omega_{n'}$ as given by Eq. (2.3), and the acceptance probability for removal of the n-mer is given by $\min[1, W-1]$.

In case of an unsuccessful removal, a resizing attempt is made. A new size $n'' \neq n$ for the chain between 1 and n' (inclusive) is selected with a probability $\omega_{n''}/(W-\omega_n)$, and the resizing move is accepted with a probability $\mathrm{acc}_{\mathrm{resize}} = \min[1, (W-\omega_n)/(W-\omega_{n''})]$. If the move is successful, monomers are removed from the previously selected end if $n'' < n$, and monomers are added to the chain end at the "dummy" positions already generated when $n'' > n$.

2.2.3 Justification of PDIRR algorithm

To demonstrate the validity of this approach, we need only to show that it should give an equivalent probability distribution to a simple GCMC approach for a mixture of n-mers whose chemical potentials are given by Eq. (2.1). The ratio of probabilities of inserting and removing any chain of a given size in a conventional GCMC simulation would in such a case be:

$$\begin{aligned}\frac{P(N_n \rightarrow N_n+1)}{P(N_n+1 \rightarrow N_n)} &= \frac{\min[1, V(N_n+1)^{-1}\exp(\beta\mu_n - \beta U_{\mathrm{ext}})]}{\min[1, V^{-1}(N_n+1)\exp(-\beta\mu_n + \beta U_{\mathrm{ext}})]} \\ &= \frac{V}{N_n+1}\exp(\beta\mu_n - \beta U_{\mathrm{ext}}). \qquad (2.4)\end{aligned}$$

Under the current PDIRR method, the chance that an n-mer is chosen for insertion in an addition move is ω_n/W, while the probability that any n-mer will be chosen for removal is equal to the number of n-mers (N_n+1) divided by the number of chains in the system $(N_{\mathrm{ch}}+1)$:

$$\begin{aligned}\frac{P(N_{\mathrm{ch}} \rightarrow N_{\mathrm{ch}}+1, N_n \rightarrow N_n+1)}{P(N_{\mathrm{ch}}+1 \rightarrow N_{\mathrm{ch}}, N_n+1 \rightarrow N_n)} &= \frac{\omega_n W^{-1}\min[1,W]}{(N_n+1)(N_{\mathrm{ch}}+1)^{-1}\min[1,W^{-1}]} \\ &= \omega_n(N_{\mathrm{ch}}+1)(N_n+1)^{-1}. \qquad (2.5)\end{aligned}$$

Substitution of Eq. (2.3) (the definition of ω_n) into Eq. (2.5) yields the same result as Eq. (2.4), given that $U_{\mathrm{ext}} = 0$ for this hard-sphere system for any allowed conformation. One detail that remains is that W in general may depend on the random process of chain growth beyond n monomers, and so may have a distribution of values. As in the off-lattice configuration bias algorithm that provided the idea for the present method, the required detailed balance condition is saved by the condition of "super-detailed balance" [5] : for any given n-mer in a given system configuration, the use of dummy re-growth moves ensures that the probability distribution of values for W will be identical for the insertion and removal processes.

The ratio of the probability of resizing a chain from n to n'' to the probability of the reverse move is:

$$
\begin{aligned}
&\frac{P(N_n+1, N_{n''} \to N_n, N_{n''}+1)}{P(N_n, N_{n''}+1 \to N_n+1, N_{n''})} \\
&= \frac{(N_n+1)N_{\text{ch}}^{-1}\,(1-\min[1,W^{-1}])\,\omega_{n''}(W-\omega_n)^{-1}\min[1,(W-\omega_n)/(W-\omega_{n''})]}{(N_{n''}+1)N_{\text{ch}}^{-1}\,(1-\min[1,W^{-1}])\,\omega_n(W-\omega_{n''})^{-1}\min[1,(W-\omega_{n''})/(W-\omega_n)]} \\
&= \frac{(N_n+1)\omega_{n''}}{(N_{n''}+1)\omega_n}.
\end{aligned}
\tag{2.6}
$$

In both numerator and denominator, the probability is a product of the chance of choosing a given size aggregate from the N_{ch} chains in the system; the probability that the removal move will fail; the probability that a given size will be selected; and the acceptance probability for the move. To justify the resizing algorithm, we show that this ratio is consistent with the addition/removal ratios. The resizing move is equivalent to removing the n-mer and adding an n''-mer in a single compound move, so the ratio of the forward and reverse resizing probabilities should equal the ratio of the product of these addition and removal probabilities to the product of the reverse move probabilities:

$$
\begin{aligned}
&\frac{P(N_n+1, N_{n''} \to N_n, N_{n''}+1)}{P(N_n, N_{n''}+1 \to N_n+1, N_{n''})} \\
&= \frac{P(N_{\text{ch}}+1 \to N_{\text{ch}}, N_n+1 \to N_n)P(N_{\text{ch}} \to N_{\text{ch}}+1, N_{n''} \to N_{n''}+1)}{P(N_{\text{ch}}+1 \to N_{\text{ch}}, N_{n''}+1 \to N_{n''})P(N_{\text{ch}} \to N_{\text{ch}}+1, N_n \to N_n+1)} \\
&= [\omega_n(N_{\text{ch}}+1)(N_n+1)^{-1}]^{-1}\omega_{n''}(N_{\text{ch}}+1)(N_{n''}+1)^{-1} = \frac{(N_n+1)\omega_{n''}}{(N_{n''}+1)\omega_n},
\end{aligned}
\tag{2.7}
$$

where the substitution in the third line comes from Eq. (2.5), and the final result matches Eq. (2.6) to demonstrate consistency. Again, the reliance on super-detailed balance is implicit: resizing from a given n-mer to a given n''-mer will yield the same distribution of W as the reverse move, as both cases will sample the same subset of "dummy" aggregates containing the larger chain.

2.2.4 Data analysis

The pressure of the self-assembled system at equilibrium with a chainlength distribution $f(L)$ can be formally written in terms of the grand partition function Ξ as:

$$p = \left[\frac{\mathrm{d}\ln\Xi(\mu, V, T; f(L))}{\mathrm{d}V}\right]_{\mu,T} = \left[\frac{\partial \ln\Xi}{\partial V}\right]_{\mu,T,f(L)} + \left[\frac{\partial \ln\Xi}{\partial f(L)}\right]_{\mu,T}\left[\frac{\partial f(L)}{\partial V}\right]_{\mu,T}. \tag{2.8}$$

At equilibrium, however, the distribution $f(L)$ is optimized, so the derivative of $\ln\Xi$ with respect to the distribution is zero; the pressure of the system is equal to the pressure of a fixed-composition system with the same length distribution. The virial contribution to the pressure from a system of polymers interacting through hard-sphere sites can be calculated from a weighted site-site distribution function [51]:

$$p = \rho k_B T\left[\frac{1}{M} + \frac{2\pi\sigma^3\rho}{3}\lim_{r\to\sigma^+}\left(\tau^*(r)g(r)\right)\right], \tag{2.9}$$

where ρ is the total concentration of monomers, M is the mean chain length (i.e., the average total number of bound and unbound monomers in the system divided by the number of aggregates in the system – including free monomers), $g(r)$ is the radial distribution function for the monomers, and $\tau^*(r)$ is defined as:

$$\tau^*(r) = \sigma^{-1}\left\langle(\mathbf{r}_{ab}\cdot\mathbf{r}_{ij})/r_{ab}\right\rangle_{r_{ab}=r}, \tag{2.10}$$

in which $\mathbf{r}_{ab}$ represents the displacement vector between two monomers and $\mathbf{r}_{ij}$ represents the center-of-mass displacement vector between two chains, where the center-of-mass of each chain is evaluated without respect to periodic boundary conditions; the magnitude of $\mathbf{r}_{ij}$ can be much greater than the box size for chains that wrap around the periodic boundaries multiple times. It is important to note that contributions to $g(r)$ from pairs of particles on the same chain, bonded or non-bonded, do not enter the pressure calculation of Eqs. (2.9) and (2.10) (as $r_{ij} = 0$ for $i = j$) except for non-bonded pairs of monomers on different periodic images of the same chain. To obtain the second term in Eq. (2.9), a histogram of the weighted site-site distribution function was recorded (with bin size 0.005 σ) every 1000 MC moves during the simulation, and a non-linear least squares fit to a cubic polynomial was used to estimate the value of the function at $r = \sigma$. Every 1000 MC moves, the bond vector order parameter S was also determined, using the method of Eppenga and Frenkel [52]. A length-dependent order parameter S_i was also determined, where $S_i = (3 \langle \cos^2 \theta \rangle - 1)/2$ where θ is the angle between a bond vector and the director of the system as a whole, while the average runs over all bonds in chains of length i. Chain length dependent order parameters were similarly recorded for the bond vectors at positions 1 and $i - 1$ (chain ends), and $i/2$ (chain midpoints).

2.2.5 Other simulation details

For local conformational relaxation, each PDIRR move of either type (successful or not) is followed by a series of "crankshaft" moves in which an interior monomer is rotated about the line connecting its two bonded neighbors by a randomly selected angle up to 2π. The move is accepted with 100% probability if the new angle satisfies both hard-sphere excluded volume requirements and the bond angle constraints described above. The first move in the series is performed on a randomly selected interior monomer. A local neighbor list generated for the first move is then employed for crankshaft moves on the five nearest bonded neighbors (or until a chain end was reached) in both directions along the chain. All (up to 11) moves are attempted whether or not the previous moves are successful.

The radial distribution function and length distribution function were calculated every 1000 move attempts. In general, between 3×106 and 2×107 PDIRR steps were performed for each data point after an initial equilibration period of between 105 and 107 PDIRR steps. A typical production run of 2×107 PDIRR moves required between 12 and 48 hours of AMD Athlon 1600 MP processor time. For certain sets of conditions, in order to improve the precision of pressure calculations, up to 108 PDIRR moves were used.

A $(40\ \sigma)^3$ box with periodic boundary conditions was used for most simulations, except at the highest value of K_{assoc} and l_p, for which a $(60\ \sigma)^3$ box was used to ensure a sufficient number of monomers for good statistics near the I-N transition. The time required for an isotropic phase to make the transition to the nematic phase depended on the conditions and choice of parameters, and was naturally very long near coexistence conditions; in most cases, therefore, nematic phase calculations were initiated using a nematic configuration obtained from a prior simulation using the same persistence length but different chemical potential or K_{assoc}. As noted a recent study [53], chain length distributions in equilibrium polymer systems relaxing towards a new equilibrium are qualitatively different from equilibrium distributions; the achievement of a stable chain length distribution, similar to those shown in Fig. 2.2, was used as the primary criterion to establish that equilibration had occurred.

2.3 Performance of the PDIIR method

The advantage of the PDIRR method in the present context over single-particle biased addition/removal algorithms is its faster sampling of chain length distributions, achieved through an increased mean step size during the random walk of chain growth and shrinkage. This advantage can be approximately quantified. At equilibrium, the mean step size of shortening a chain must equal the mean step size for lengthening a chain; the former is easier to estimate, as removal moves are not subject to excluded volume constraints. The probability of shortening a chain of length n_0 by n monomers in a resizing move is proportional to $(\omega_{n_0-n}/\omega_{n_0}) = [K_{\mathrm{assoc}} \exp(\beta\mu)]^{-n}$. If the value of $G \equiv K_{\mathrm{assoc}} \exp(\beta\mu) \leq 1$, then the degree of resizing will be comparable to the average chainlength. Otherwise, the average number of monomers to be resized is determined by which evaluates to $(G-1)^{-1}$ when n_0 is large. There is no advantage to the PDIRR method when $G \gg 1$, as under such circumstances it is unlikely that more than one monomer would be added or removed in a single PDIRR move under equilibrium conditions. The optimal performance regime for the method, where $(G-1)$ is positive but small, corresponds to semidilute conditions where considerable volume is available in the solution but chains do begin to crowd each other. Fortunately for the present problem, this is the regime in which the I-N transition is observed, as the density of the system first becomes great enough that excluded volume effects drive orientational ordering. Average resizing step sizes $\langle n \rangle$ between 3 and 11 are observed in our simulations, in accordance with these predictions, for systems near the transition. The computational cost per step will be a factor of $\sim \langle n \rangle$ for the PDIRR method compared to the single-monomer addition (or less, as the PDIRR method can employ neighbor lists more effectively, given that the moves are all made in a localized area), while the change in rate of growing or shrinking a long chain will scale as $\langle n \rangle^2$, so the expected benefit will be $\sim \langle n \rangle$. Mean step sizes on the approach to equilibrium, either during the initial polymerization or during an I-N transition, may be yet larger.

No unique criterion is convenient to test the relative efficiencies of MC algorithms. For one test, we used as a starting point a system configuration generated with $l_p = 10\ \sigma$, $K_{\mathrm{assoc}} = 5000$, $\exp(\beta\mu) = 3 \times 10^{-4}$ and observed the rate of relaxation of average chain length M from 36 to 24 when the equilibrium constant was lowered to 2000 σ^3 and $\exp(\beta\mu)$ increased to 7.5×10^{-4} (which does not significantly change the average total monomer concentration of the system). The real-time rate of change of the PDIRR code was 3.5 times faster than the monomer addition/removal code. With $(G-1)^{-1} = 2$, this is not a particularly favorable case for the PDIRR method in equilibrium sampling. For the growth of a new isotropic phase ($K_{\text{assoc}} = 5000$, $\exp(\beta\mu) = 2.2 \times 10^{-4}$, $l_p = 100\ \sigma$) from an initially empty box of edge length 60 σ, the results are more dramatic; the known equilibrium density of 0.079 σ^{-3} (17000 particles) and mean chain length of 21.5 σ were reached 15 times more quickly by the PDIRR code than by the single monomer addition code.

2.4 Results on semiflexible equilibrium polymers

2.4.1 Simulation results for $K_{\text{assoc}} = 5000$

We first present detailed results at a single value of the monomer association constant, $K_{\text{assoc}} = 5000$ to illustrate general structural and thermodynamic features of the isotropic and nematic phase of equilibrium polymers. Fig. 2.1(a) shows isotherms of chemical potential versus total monomer concentration, which show little dependence on persistence length at low densities. Effects of bond entropy that will lead to a stiffness-dependent association behavior when bond energy is fixed [54] are not seen here because the model allows the direct selection of association free energy, through Kassoc, for chains of any stiffness.

Growth law, isotropic phase

An ideal monomer association model predicts that the mean chain length M will scale with the square root of the total monomer concentration ρ, or more precisely $M = 0.5 + 0.5(1 + 4K_{\mathrm{assoc}}\rho)^{1/2}$, in the isotropic phase. As shown in Fig. 2.1(b), the ideal model accurately describes the simulation results in the limit of low monomer concentration. Upward deviations from the ideal model predictions at higher concentrations, previously observed in simulations of flexible [55–57] and semiflexible [30] equilibrium polymer, appear to be nearly independent of persistence length for the isotropic phase in the regime depicted in Fig. 2.1. This deviation was predicted by Gelbart et al. [58] to arise from excluded volume considerations (i.e., a monomer at the end of the chain occupies a greater excluded volume than does an interior monomer) which increase in importance as the concentration increases. In the more rigid polymer systems ($l_p = 100\ \sigma$ and $l_p = 1000\ \sigma$), the isotropic phase is observed to become unstable above a certain chemical potential with respect to a nematic phase, with nematic order parameter greater than 0.5 [see Fig. 2.1(c)]. The discontinuous change in concentration that goes along with ordering indicates a first-order lyotropic transition between two phases of different concentrations, separated by a two-phase region.

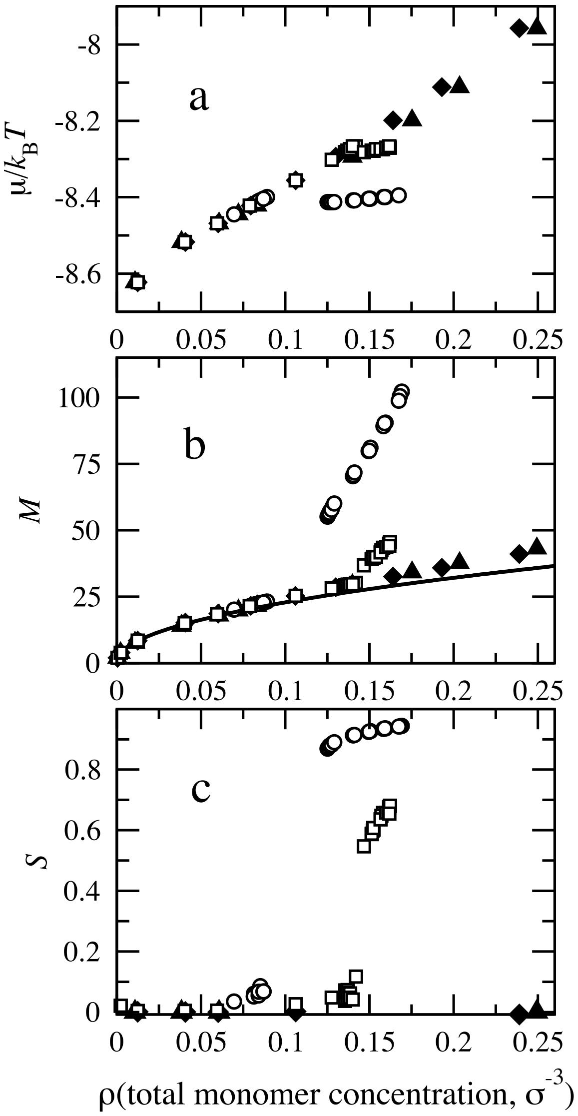


Figure 2.1: Dimensionless monomer chemical potential μ/k_BT [Panel (a)], mean number of monomers per chain, M [Panel (b)], and orientational order parameter S [Panel (c)] as a function of total monomer concentration ρ for equilibrium polymer systems of various persistence lengths, l_p. Open circles: $l_p = 1000\ \sigma$; open squares, $l_p = 100\ \sigma$; closed diamonds, $l_p = 10\ \sigma$; closed triangles, $l_p = 4\ \sigma$. The solid curve in Panel (b) gives the ideal chain growth law, $M = 0.5 + 0.5(1 + 4K_{\text{assoc}}\rho)^{1/2}$. Monomer association constant $K_{\text{assoc}} = 5000$ for all systems.

Chain length distributions and length-dependent order parameter

Distributions of chain lengths in isotropic and nematic phases near coexistence are shown in Fig. 2.2. The isotropic phases of the semiflexible polymers always yielded a simple exponential distribution, $\rho L \propto A \exp(-L/M)$, where M coincides with the mean chain length. The nematic phases, on the other hand, gave an approximate sum of two exponentials, $\rho_L = A \exp(-L/M_{\text{short}}) + B \exp(-L/M_{\text{long}})$ indicating that different considerations govern the growth of short and long chains. This difference is related to the difference in the degree of ordering; as previous authors have noted [30, 42, 48], short chains have lower orientational order than long chains in a polydisperse mixture. The inset to Fig. 2.2 shows that the range of short chainlengths where deviation from the single exponential distribution is observed matches the range over which the chainlength-dependent orientational order parameter S_i rises rapidly. Figure 2.3, a snapshot from the $l_p = 1000\ \sigma$ simulation with the short $(L \leq 10\ \sigma)$ chains highlighted in black, gives a visual impression of the orientational disorder of the short chains.

An exponential distribution is expected when the free energy of a chain changes linearly with each monomer, as is the case in long chain limit of the nematic phase, where the orientational distribution is unaffected by increasing chain length, or in the isotropic phase. In the short chain limit of the nematic phase, the orientational entropy changes rapidly with chain length, so the total orientational entropy does not change linearly with chain length. Compared to the long-chain limiting distribution,

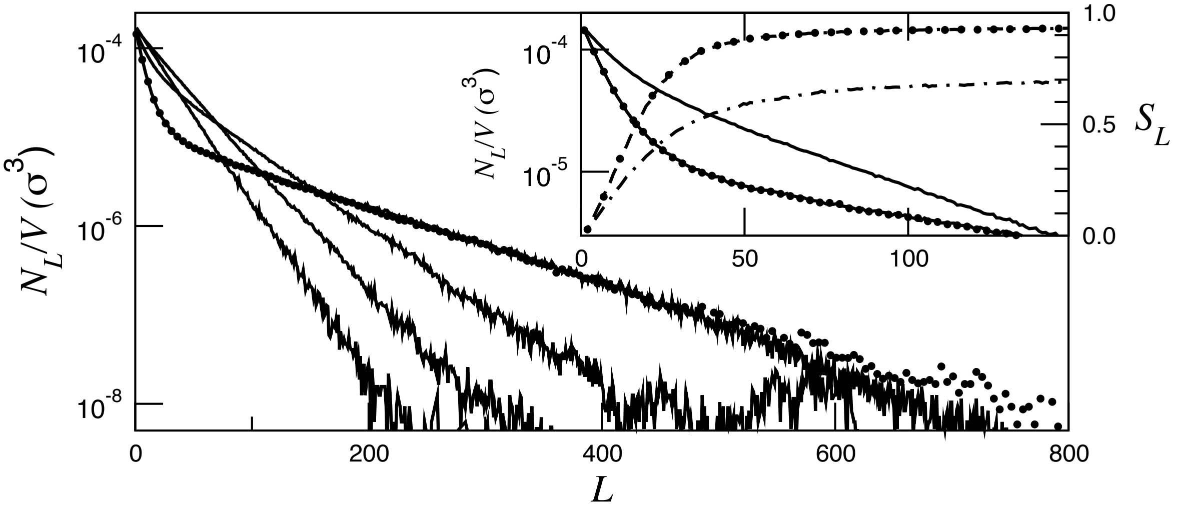


Figure 2.2: Chain length distribution. Average number density of chains of L monomers versus chain length L for $K_{\text{assoc}} = 5000$. Data sets I_{1000} and N_1000 correspond to isotropic and nematic phases at $l_p = 1000\ \sigma$, $\mu/k_BT = -8.414$; I_100 and N_100, isotropic and nematic phases at $l_p = 100\ \sigma$, $\mu/k_BT = -8.278$. Inset: expanded view of chain length distribution (solid curves) and chain length-dependent order parameter S_L (dot-dash curves) for nematic phases. Filled circles represent data obtained with a simulation box of volume $(60\ \sigma)^3$; for all other data, box size is $(40\ \sigma)^3$. $K_{\text{assoc}} = 5000$ for all data.

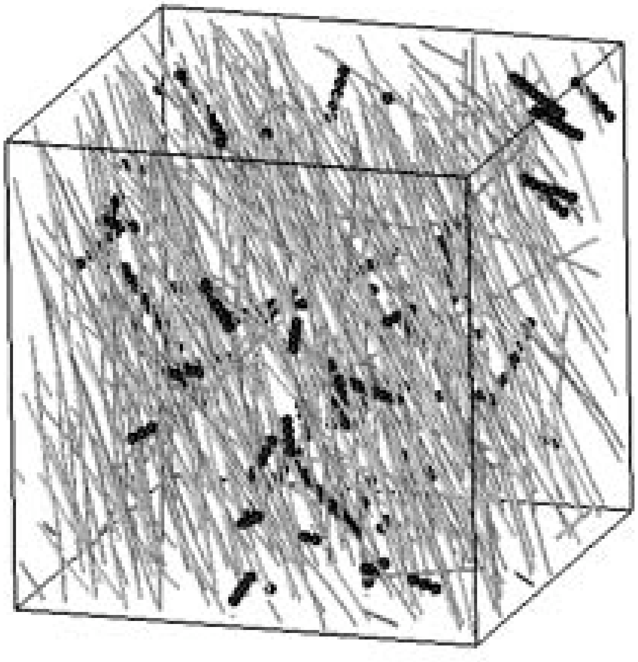


Figure 2.3: Snapshot from MC simulation of nematic phase. Short chains (10-mers and smaller) are highlighted as black spheres; longer chains shown as grey sticks. $K_{\text{assoc}} = 5000$, $l_p = 1000\ \sigma$, $G = 1.11$, box size $= (40\ \sigma)^3$.

an enhancement in the concentration of short chains is consistent with the higher orientational entropy of these chains.

Finite size effects and deflection length

Figure 2.2 also demonstrates that these simulation results are quantitatively unchanged when the box edge length is increased from 40 σ to 60 σ, in a test of box size effects under a single set of conditions [$l_p = 1000\ \sigma$, $K_{\text{assoc}} = 5000$, $\exp(\beta\mu) = 2.22 \times 10^{-4}$]. This apparent insensitivity to finite-size effects is rather unexpected, as several length-scales in the system (persistence length $= 1000\ \sigma$, mean chain length $= 56\ \sigma$, longest observed chain length $\approx 800\ \sigma$) are significantly greater than the original box size of 40 σ. A likely explanation is that a shorter lengthscale, the deflection length [25] $\lambda \approx \langle\theta^2\rangle\, l_p/2 \approx (1-S)l_p/3$, governs the structure of the nematic phase. This formula predicts $\lambda = 20\ \sigma$ or $34\ \sigma$ depending on whether the order parameter S averaged over all chains or for long chains only is used.

The deflection length represents the characteristic chain contour length over which the chain orientation is free to fluctuate without being deflected towards the director by collisions with other chains. We have independently determined a related quantity from the simulation by calculating the bond angle correlation function, $f(j) = \langle \mathbf{r}_i \cdot \mathbf{r}_{i+j} \rangle$, with $\mathbf{r}_i$ the bond vector of the ith bond in a chain. For a free ideal chain the result is a simple exponential with decay length equal to persistence length, as we indeed observe for the isotropic phase. In the nematic phase, as shown in Fig. 2.4, we find that the bond orientational correlation function $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+j} \rangle$ decays not to zero (as it does for an isotropic phase, given long enough chain lengths) but to approximately $\langle \cos^2 \theta \rangle = 2/3(S+1/2)$ as j increases. The apparent rise in this correlation function at high L is probably the result of poor statistics (as the number of long chains is small). The deflection lengths estimated from the roughly exponential decay is 41 σ for the system with $l_p = 1000\,\sigma$, and 23 σ for the system with $l_p = 100\,\sigma$. Like the chain length distribution and order parameter shown in Fig. 2.2, the bond correlation function is not affected by the change in box size. All of these results suggest that the present simulations are not strongly influenced by finite size effects, in spite of the common-sense notion that the aggregate size should never be larger than the box dimension. We cannot, however, rule out the possibility that fluctuations on much larger scales than our box size may influence to the large-scale order or thermodynamic stability of the nematic phase in macroscopic systems.

Figure 2.4 also shows the effects of bond placement within a chain on the ordering of bonds. For long chains, bonds at chain ends have less orientational order than chain interiors. As predicted by Khokhlov and Semenov [59], we find that the end segments are roughly twice as disordered as the interior segments (i.e., $\langle \cos^2 \theta \rangle_{\text{end}} \approx \langle \cos^2 \theta \rangle^2_{\text{middle}}$).

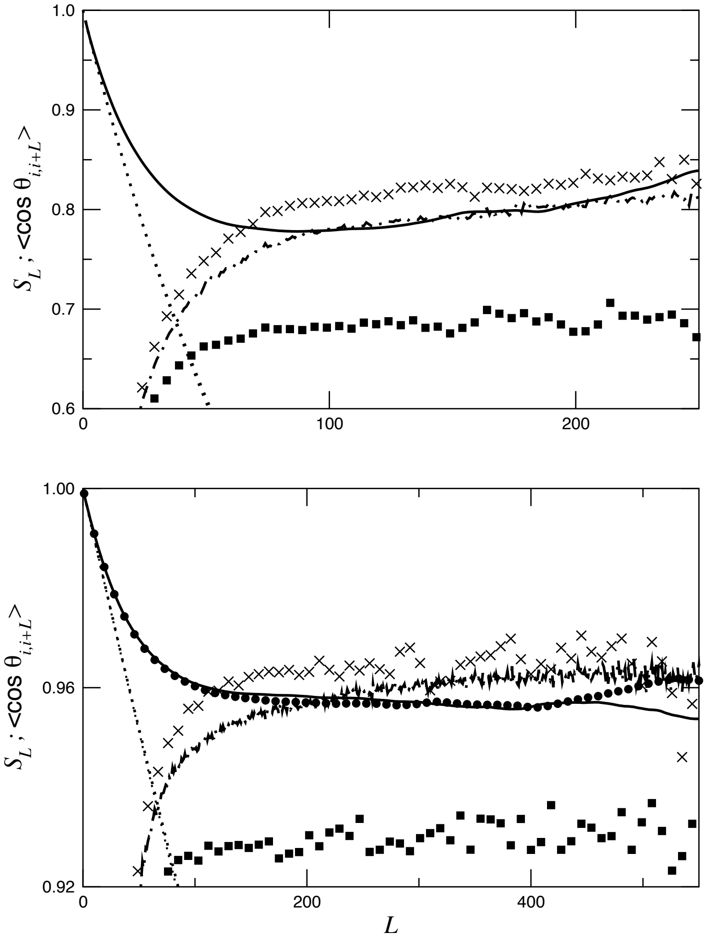


Figure 2.4: Bond angle correlation function $\langle \cos\theta_{i,i+L} \rangle$ (nematic phase: solid curve; isotropic phase: dotted curve), chain length dependent order parameter S_L (dot-dash curves), and order parameters of chain end (squares) and chain midpoint (×'s) bond vectors as a function of length L, for nematic phases. (See Fig. 2.2 for simulation parameters.) Filled circles ($l_p = 1000\ \sigma$ only) represent box size of $(60\ \sigma)^3$; other data obtained with $(40\ \sigma)^3$ box.

2.4.2 General comparison between simulation and theory

Simulations of isotropic and nematic systems in the vicinity of the I-N transition were performed for $l_p = 100\ \sigma$ and $l_p = 1000\ \sigma$ at five values of K_{assoc} between 320 and 12500. The same qualitative features (e.g., first-order isotropic-nematic phases transitions, functional forms of single- and bi-exponential chain length distributions in the isotropic and nematic phases, respectively) were observed in all of these systems. A benefit of the simplicity of the present model, which facilitates the precise and independent choice of K_{assoc} and l_p, is the opportunity for direct comparison with analytical theory. We compare our results with predictions based on the analytic free energy expressions of van der Schoot and Cates for isotropic and nematic phases of semiflexible spherocylinders [27].

Isotropic phase properties

The isotropic phase free energy per unit volume f_{iso} of Ref. [27] can be written (in units of k_BT) :

$$f_{\text{iso}} = -\rho E(1 - M^{-1}) + \left(B + \frac{\kappa_I}{M}\right)\rho^2 + \frac{\rho}{M}\left(\ln\frac{\rho}{M} - 1\right) + \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M. \tag{2.11}$$

The first term gives the cohesive energy in terms of the chain scission energy E, corresponding to $\ln K_{\text{assoc}}$ in the present model. The second term approximates steric interactions among chains within the second virial approximation, including end effects through the parameter κ_I. The third term is the translational free energy of a system in which the total number density of chains is ρ/M, and the remaining terms account for the mixing entropy of the polydisperse system for a single exponential distribution of chain concentration versus chain length. The mean chain length M that minimizes this free energy is

$$M_{\text{iso}}(\rho) = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\rho \exp(E + \kappa_I \rho)}. \tag{2.12}$$

The free energy of the nematic phase contains additional terms to account for the effects of orientational order on the rotational, conformational, and steric entropies of the chains.

$$\begin{aligned} f_{\text{nem}} &= -\rho E(1 - M^{-1}) + \left(\frac{4}{\sqrt{\pi\alpha}}B + \frac{\kappa_N}{M}\right)\rho^2 + \rho\left(\frac{1}{M}\ln\left(\frac{\alpha}{4}\right) + \frac{\alpha}{4l_p}\right) \\ &\quad + \frac{\rho}{M}\left(\ln\frac{\rho}{M} - 1\right) + \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M. \end{aligned} \tag{2.13}$$

Orientational order is included as $\alpha = 2\langle\theta^2\rangle - 1 \approx 3/(1-S)$; large α corresponds to a highly ordered nematic phase. The second virial term is now dependent on orientational order as $\alpha^{-1/2}$, while the rotational and configurational free energies penalties for nematic ordering scale as $\ln(\alpha)/M$ and α/l_p, respectively. The mean chainlength M becomes

$$M_{\text{nem}} = \frac{1}{2} + \frac{1}{2}\sqrt{1 + \rho\alpha \exp(E + \kappa_N \rho)}. \tag{2.14}$$

Like M, α is not an independent parameter but is determined at each choice of density, association constant, and persistence length l_p by minimization of the free energy;

unlike M, α is not available as a simple expression in terms of the other system quantities (unless some terms in Eq. (2.13) are assumed negligible), so it is determined by a numerical minimization of fnem.

In addition to the bond strength and persistence length, the analytical model uses three characteristic parameters with units of volume: B, κ_I , and κ_N. While van der Schoot and Cates [27] have used geometrical arguments to determine values of these parameters for spherocylinders, the simulation model employs tangent hard-sphere chains rather than spherocylinders. We have chosen values for the first two of these factors to reproduce the simulation data in the isotropic phase in the limit of low monomer density. As shown in Fig. 2.5, the growth law $M_{\text{iso}}(\rho)$ of Eq. (2.12) fits simulation results from five values of K_{assoc}, and two values of l_p, with one adjustable parameter, $\kappa_I = 1.45\ \sigma^3$. Also shown in Fig. 2.5, the chemical potential μ_{iso} obtained from the first derivative of f_{iso} with respect to ρ in Eq. 2.11:

$$\mu_{\text{iso}} = -E(1 - M^{-1}) + 2\left(B + \frac{\kappa_I}{M}\right)\rho + \frac{1}{M}\ln\frac{\rho}{M} + \frac{M-1}{M}\ln(M-1) - \ln M \quad (2.15)$$

is likewise in good agreement with the simulation results for monomer concentrations below $\rho = 0.1\ \sigma^{-3}$ with the second virial coefficient set to $B = 0.85\ \sigma^3$. As deviation from between theory and simulation in the measured concentration range depends only on ρ, this discrepancy can be attributed to a simple neglect of third and higher-order terms in the virial expansion of pressure, independent of persistence length and chain length distribution.

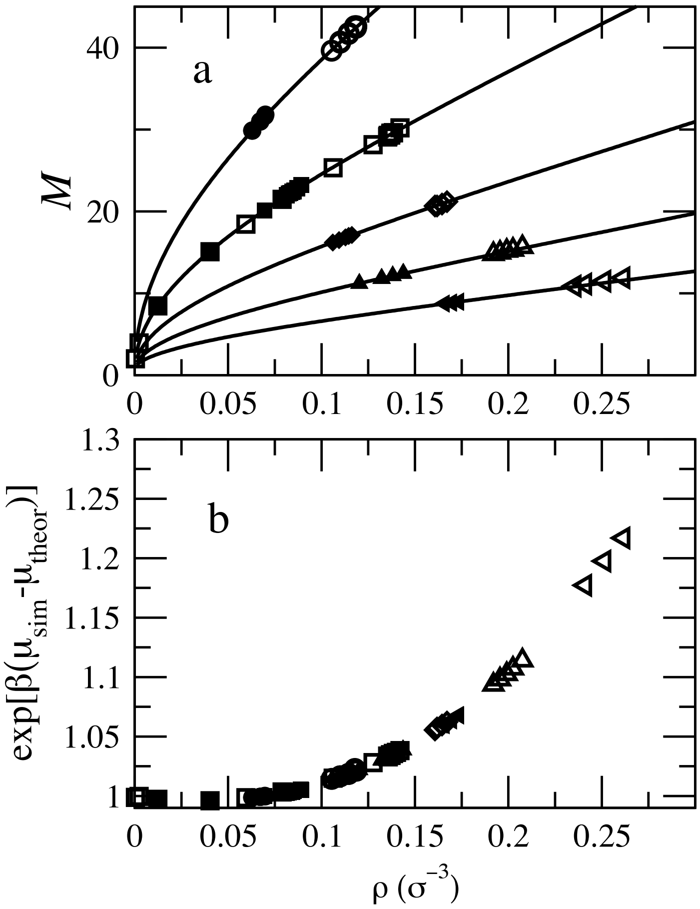


Figure 2.5: Comparison of theoretical predictions to simulation results for isotropic phases. Panel (a), mean chain length M versus total monomer concentration, showing simulation data and fits to Eq. (12). Panel b, ratio of simulation values of $\exp(\beta\mu)$ to values predicted using Eq. (15). Open symbols: $l_p = 100\ \sigma$; filled symbols: $l_p = 1000\ \sigma$. Circles, squares, diamonds, triangles, and left-facing triangles correspond to $K_{\text{assoc}} = 12500; 5000; 2000; 800;$ and 320, respectively. Parameters are $\kappa_{\text{I}} = 1.45\ \sigma^3$, $B = 0.85\ \sigma^3$.

Nematic Phase properties

The choice of κ_N in Eqs. (2.14) and (2.13) is somewhat arbitrary, as end effects on packing entropy are in fact dependent on the degree of orientational order ; here we fix κ_N using the the relation $\kappa_N = 1.5\kappa_I$ used in Ref. [27] and the value $\kappa_I = 1.45\ \sigma^3$ determined from the isotropic chain length distributions. (Other choices of κ_N do not significantly improve the global fit between simulation and theory as presented below.)

The theory does not attempt to include chain-length dependent orientational order, which was shown above to accompany an enhancement of short chains in the system relative to the limiting single exponential distribution for long chains. It is no surprise, therefore, that the theory over-predicts the overall mean chain length M in the nematic phase, as shown in Fig. 2.6 (points below the solid diagonal line.) Considering the long chains alone in the simulation does not yield much improved agreement; also shown in Fig. 2.6 is the ratio of the mean chain length for the long-chain limit, M_{long}, to the theoretically predicted M (points above the diagonal line). The theory appears to under-predict the degree of growth with respect to increasing ρ, and to fail in quantitatively predicting the influence of persistence length on the size distribution. Qualitatively, the simulation results and theoretical predictions are in agreement with the trends of increasing M with increasing ρ and l_p.

The chemical potential of the nematic phase, from the differentiation of Eq. (2.13)

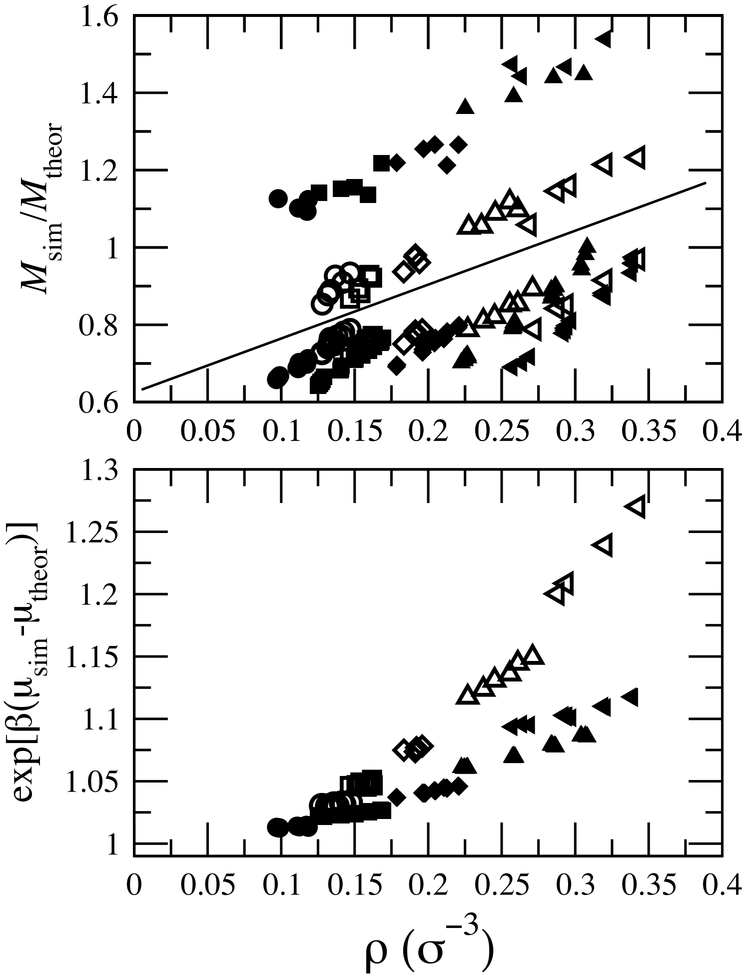


Figure 2.6: Summary of simulation results vs. theoretical predictions for nematic phases. Upper panel: ratio of observed to predicted values of mean chain length M using Eq. (14). Values below the line: MN/Mtheor; values above the line, $M_{\mathrm{long}}/M_{\mathrm{theor}}$. Lower panel: ratio of simulation values of $\exp(\beta\mu)$ to values predicted using Eq. (16). Symbols are defined as in Fig. 2.5.

with respect to ρ, is

$$\begin{aligned}\mu_{\text{nem}} &= -E(1-M^{-1}) + 2\left(\frac{4}{\sqrt{\pi\alpha}}B + \frac{\kappa_N}{M}\right)\rho + \left(\frac{1}{M}\ln\left(\frac{\alpha}{4}\right) + \frac{\alpha}{4l_p}\right) \\ &\quad + \frac{1}{M}\ln\frac{\rho}{M} + \frac{M-1}{M}\ln(M-1) - \ln M \end{aligned} \tag{2.16}$$

Figure 2.6 (lower panel) shows that the theory underestimates the chemical potential by a significant factor that grows as ρ increases. In contrast to the isotropic case, however, the deviation cannot be fixed by a simple ρ-dependent correction, as the degree of deviation is dependent on K_{assoc} and l_p as well as ρ. Quantitative prediction of the nematic phase chemical potential and chain length distribution would apparently require non-trivial modifications to the simple theory of Ref. [27].

Determination of coexistence region

The approximate boundaries of the isotropic, nematic, and two-phase coexistence concentration ranges at any given K_{assoc} and l_p can be determined with little difficulty through a series of simulations over a range of chemical potentials, to determine the limits of stability of isotropic and nematic phases. Due to hysteresis (i.e. over a range of values of μ, both phases are kinetically stable over the course of the simulation), the chemical potential data alone are insufficient to identify the unique pair of coexistence densities at each K_{assoc} and l_p, which are distinguished by common values of both pressure ($p_I = p_N$) and chemical potential ($\mu_I = \mu_N$). As we are able to determine the system pressure during our simulations, in principle we should be able to find the coexistence pair with little trouble. Unfortunately, pressure determination is particularly sensitive to the number, length, and configurations of the longest chains in the system through the center-of-mass displacement term r_{ij} in Eq. (2.10); these chains have long relaxation times that lead to large and uncertain statistical error in pressure results. Improving the statistical sampling for all simulation runs to reduce the error bars sufficiently to determine an intersection point between isotropic and nematic phases with some precision would be prohibitively time-consuming. Instead, a single nematic phase point, generally the lowest stable nematic concentration, was chosen for over-sampling (up to 108 Monte Carlo moves) for each choice of K_{assoc} and l_p investigated. The Gibbs-Duhem relation $\mathrm{d}p/\mathrm{d}\mu = \rho$ used to extrapolate from this (relatively) precisely determined pressure to obtain the pressure at higher densities:

$$p(\mu) \approx p(\mu_0) + (\mu - \mu_0)\rho + (\mathrm{d}\rho/\mathrm{d}\mu)(\mu - \mu_0)^2/2 \tag{2.17}$$

Pressure data for $K_{\mathrm{assoc}} = 5000$ is shown in Fig. 2.7, including the Gibbs-Duhem extrapolation from the over-sampled points in the nematic and isotropic phase. We found reasonable general agreement between simulation results and Eq. (2.17), although in some cases, as shown in the figure for $l_p = 1000\ \sigma$, the extrapolation from the over-sampled nematic phase point lies outside the error bars of the other points (note that this is an extrapolation, not a fit). Error bars are derived only from the error in fitting the weighted site-site correlation function, and therefore do not account for system fluctuations at large length- and time-scales that would not

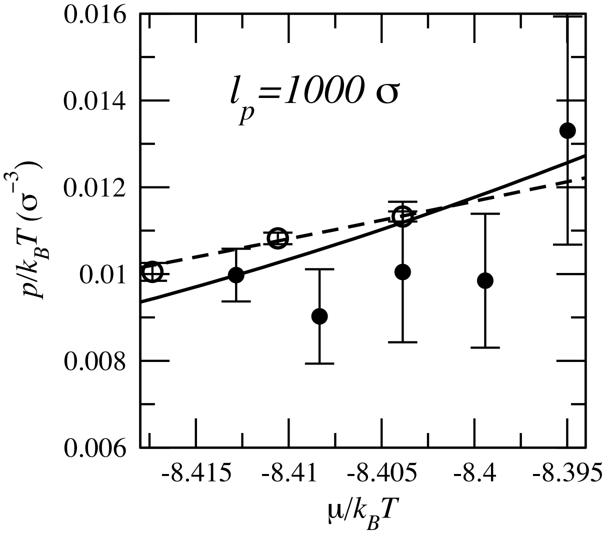


Figure 2.7: pressure versus chemical potential, from simulations of equilibrium polymers with $K_{\text{assoc}} = 5000$. Circles: simulation results for isotropic (open) and nematic (filled) phases. Curves: extrapolation from the best-measured simulation point using Gibbs-Duhem relation, Eq. (2.17), for isotropic (dashed curve) and nematic (solid curve) phases.

affect the smoothness of this function; in other words, we believe that the error bars are underestimated, especially for data points that were not oversampled. A different problem arises at $l_p = 100\ \sigma$: the combination of narrow coexistence regions (which, consistent with the Gibbs-Duhem relation, yield nearly superimposable graphs of p versus μ for isotropic and nematic phases near coexistence) and uncertain pressures made it impossible to even estimate a unique coexistence pair.

The intersections of Gibbs-Duhem curves were used to determine a pair of coexistence densities for the five systems with $l_p = 1000\ \sigma$ (with error bars based on the estimated uncertainty in calculated pressure). Due to the difficulty in choosing a unique pair of coexisting densities for systems with $l_p = 100\ \sigma$, only ranges are presented in the figure and table, with the limits determined by either the instability of the phase on the simulation timescale (e.g., the highest μ at which a spontaneous transition from nematic to isotropic was observed was used for the lower limit) or a clear difference in pressure (e.g., the lowest μ for which the pressure was clearly greater in the nematic phase than in the isotropic phase was used for the upper limit.) Figure 2.8 shows the phase boundaries determined from simulation with those predicted by the theory of van der Schoot and Cates, obtained by solving for the intersection of plots of chemical potential versus pressure for the free energy expressions of Eqs. (2.11) and (2.13). Additional data pertaining to the coexisting phases are collected in Table 2.1.

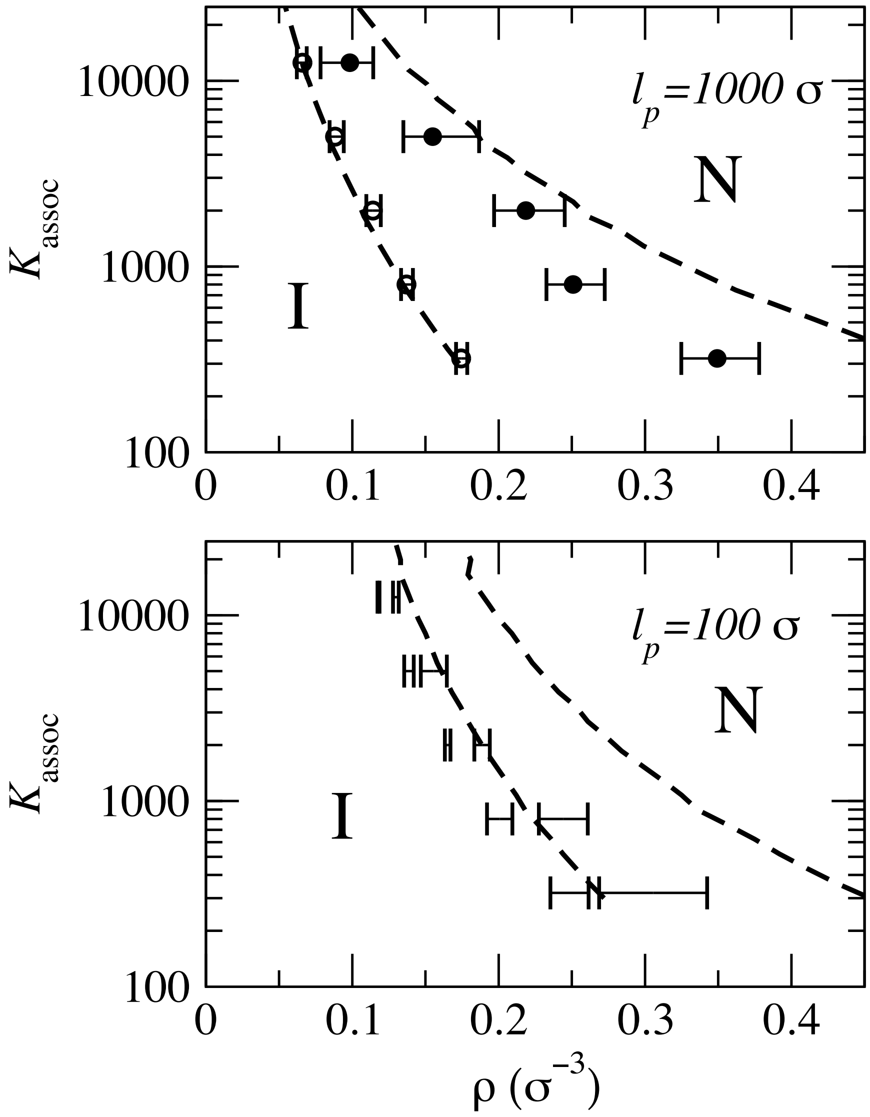


Figure 2.8: Comparison of phase diagrams from simulation and theory. Dashed lines: coexistence boundaries calculated using Eqs. (2.11-2.16), derived from Ref. [27]. Points/error ranges: coexistence boundaries determined from simulation as described in text.

$l_p(\sigma)$	K_{assoc}	G	$\rho_I(\sigma^{-3})$	M_I	$\rho_N(\sigma^{-3})$	M_N	M_{long}	M_{short}	S_N
1000	12500	1.08	0.066	30.6	0.10	73.4	121.8	9.5	0.865
	5000	1.12	0.088	23.2	0.16	87.2	138.2	6.2	0.932
	2000	1.17	0.114	16.9	0.22	92.2	144.0	3.8	0.957
	800	1.22	0.137	12.0	0.25	63.0	104.2	2.9	0.955
	320	1.32	0.175	9.0	0.34	69.9	107.6	1.7	0.970
100	12500	1.221	0.117	42.1	0.128	53.9	62.4	12.6	0.556
		1.225	0.119	42.6	0.132	57.1	66.6	10.7	0.601
	5000	1.265	0.135	29.1	0.147	36.8	44.1	10.8	0.547
		1.285	0.142	30.3	0.165	46.3	56.1	7.2	0.686
	2000	1.345	0.163	20.9	0.183	29.4	36.5	6.9	0.629
		1.360	0.167	21.1	0.194	32.9	42.3	5.9	0.687
	800	1.456	0.192	14.8	0.227	24.0	32.0	3.5	0.706
		1.520	0.209	15.8	0.261	32.2	41.4	4.9	0.795
	320	1.632	0.235	10.9	0.269	16.8	22.5	3.9	0.684
		1.792	0.262	11.9	0.343	32.1	41.5	2.0	0.856

Table 2.1: Properties of isotropic and nematic phases at coexistence. For systems with $l_p = 1000\ \sigma$, properties of isotropic and nematic phases at a common value of $G = K_{\text{assoc}} \exp(\beta\mu)$, selected based on intersection of Gibbs-Duhem extrapolations of pressure versus μ to give the best estimate of the phases at coexistence. For systems with $l_p = 100\ \sigma$, lower and upper limits of G and corresponding isotropic and nematic phase properties are given. Columns 4 and 5 contain data from isotropic phase simulations; columns $6 - 10$ from nematic phase simulations.

A few comments on methodology before discussing the results: we note that the long chains that complicate pressure determination via Eqs. (2.9) and (2.10) would similarly pose problems in constant pressure or Gibbs Ensemble simulation methods of determining coexistence that rely on volume re-scaling moves; the relative displacements associated with volume changes are proportional to center-of-mass displacement, so very small volume change moves would be required in dense systems of long chains. A more promising alternative to the present approach would be to use a density-of-states method [60], sampling throughout isotropic, nematic, and coexistence regions using addition and removal MC moves. It is not clear, however, whether the hysteresis observed in the present simulations would interfere with the calculation of a smooth and statistically meaningful density-of-states function across the transition. We did not make an effort to study the hexagonal phase, which has commonly been observed in experiments on equilibrium polymers [34] and was seen in simulations by Fodi and Hentschke [30]. Hexagonal ordering was not observed at any of the conditions tested; some possible explanations may be that the present simulation algorithm does not allow for periodically repeating unbroken chains, that the fixed box size was incommensurate with hexagonal ordering, or that the rate of the nematic-hexagonal transition was too slow to observe using our algorithm.

Qualitatively, the simulated phase diagrams' shapes are in good agreement with theoretical predictions. Following Eq. (2.13), the rotational free energy penalty for ordering in the nematic phase is inversely proportional to mean chainlength M, while the conformational free energy penalty is independent of M. For systems of predominantly short chains (roughly, $M < l_p$), which are plentiful when K_{assoc} is small, the position of the phase transition is strongly dependent on K_{assoc} through its effect on M; in the limit of large K_{assoc} (i.e., high M) the position of the phase transition becomes independent of K_{assoc} because of the M^{-1} dependence of the rotational free energy term, and is determined by the persistence length alone. This long-chain regime has not been achieved in the present simulations. In general, the breadth of the coexistence region is greater at low K_{assoc}, where the presence of short chains leads to strong coupling between chain growth and ordering as the monomer concentration increases.

The width of the coexistence region is generally overestimated by the theory, as expected given that the theory gives good agreement with simulation for the isotropic phase but underestimates of the chemical potential of the nematic phase. An important contribution to the discrepancy is that the theory only includes steric effects within the second virial approximation, and so overestimates the stability of nematic phases at high concentrations. Even within the second virial approximation, the orientationally disordered short chain component observed by simulation but not incorporated in the theoretical model may also contribute to this discrepancy.

Scaling of short chainlength $M_{\mathbf{short}}$

To guide future refinements of the theoretical model, we attempt to characterize the scaling behavior of the chainlength distribution for short chains. In an ideal system, the equilibrium concentration ratio of chains of length L to length $L+1$ equals $G = K_{\mathrm{assoc}} \exp(\beta\mu)$. For a system with excluded volume, the ratio is

$$\rho_{L+1}/\rho_L = G[1 - \gamma(\rho, L)], \tag{2.18}$$

where $\gamma(\rho, L)$ is the probability that addition of a monomer to the end of an L-mer results in a hard-sphere overlap. The number ratio of L-mers to $(L+1)$-mers within the population of orientationally disordered short chains is $exp(1/M_{\mathrm{short}})$. We find that when $G^{-1}\exp(-1/M_{\mathrm{short}})$ is plotted against ρ, as shown in Fig. 2.9, the data lies along two straight lines segregated by persistence length, corresponding to $(1 - 1.95\rho)$ for $l_p = 100\ \sigma$ and $(1 - 1.65\ \sigma)$ for $l_p = 1000\ \sigma$. For the short, isotropically oriented chains that contribute to the initial exponential decay in the nematic chain length distribution, γ appears to be independent of i and proportional to the total monomer density ρ to a first approximation, with a constant of proportionality that depends on the persistence length. (Previous work [27] has already served for a qualitative understanding of the chain length distribution in the long-chain limit, as expressed in Eq. (2.14); in terms of Eq. (2.18), the probability $\gamma(\rho, L)$ is lower for long chains because of their orientational order, giving a slower decay in the chain length distribution for large L.)

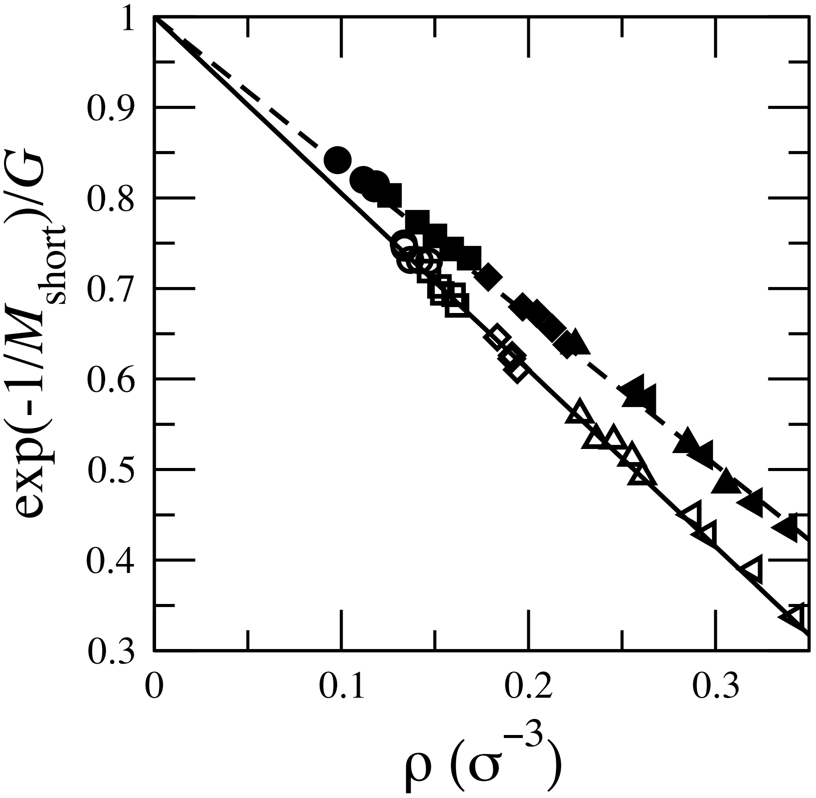


Figure 2.9: Scaling behavior of the short chain-length decay constant M_{short}^{-1} for nematic phases, with $G \equiv K_{\mathrm{assoc}} \exp(\beta\mu)$. Symbols are as in Fig. 2.5. Solid line: $y = 1 - 1.95x$; dashed line, $y = 1 - 1.65x$.

2.4.3 Disordering of short chains in the nematic phase

The nematic ordering transition of hard rigid and semirigid rods is strongly dependent on rod length [24]. In bidisperse and polydisperse mixtures of rods or semiflexible polymers, the length-dependent partitioning of chains between ordered and isotropic phases, and the dependence of order parameter on length, has been studied by theory and simulation [48, 61, 62]. In self-assembled systems in which the size distribution of either phase is determined by the forming and breaking of bonds at equilibrium, the length distributions of the two phases are expected to be quite different, as has been verified by simulation [30, 31, 63]. In the simulation study, we identified the length distribution of the nematic phase in semiflexible equilibrium polymers as being roughly bi-exponential, dropping off quickly for short chains and slowly for longer chains, and found that the division between the two regimes corresponds to a division between low and high order parameters. In this Note, we explore the qualitative effects of the length distribution by modifying the existing mean-field statistical thermodynamic treatment of the problem. The issues we will address are: does the phenomenon influence the coexistence boundaries of the I-N transition? And under what conditions is the disordered, short chain component of the distribution most prominent?

Theory

In simulations of semiflexible hard-sphere chains, the degree of orientational order in the nematic phase increases continuously with chain length, approaching a maximum value for long chains. Taylor and Herzfeld [64] have studied the limiting case of perfectly rigid spherocylinders allowing full optimization of the orientational distribution as a function of rod length. To simplify theoretical calculations, we will instead represent all short chains (those below a variationally determined crossover length L_c in units of monomer diameter σ) as isotropically oriented and all long chains (those with contour length greater than L_c) as having a Gaussian orientational distribution independent of length. The contributions of each population to the system free energy can then be treated as additive, except for a single cross term describing the excluded volume interactions between short and long chains. The free energy densities of the isotropic and nematic phases can be written [27, 63] ($k_{\mathrm{B}}T \equiv 1$)

$$f_{\mathrm{iso}} = -E\rho_{\mathrm{tot}} + \sum E\rho(i) + B_2\rho_{\mathrm{tot}}^2 + \sum \rho(i)\left[\ln\rho(i) - 1\right]; \tag{2.19}$$

$$\begin{aligned} f_{\mathrm{nem}} &= -E\rho_{\mathrm{tot}} + \sum_{i=1}^{L_c} E\rho(i) + \sum_{i=L_c+1}^{\infty} \rho(i)(E + \ln\frac{\alpha}{4}) \\ &\quad + B_2(\rho_s^2 + 2\rho_s\rho_l + \frac{4}{\sqrt{\pi\alpha}}\rho_l^2) + \rho_l\frac{\alpha}{4l_p} \\ &\quad + \sum_{i=1}^{L_c} \rho(i)[\ln\rho(i) - 1] + \sum_{i=L_c+1}^{\infty} \rho(i)[\ln\rho(i) - 1], \end{aligned} \tag{2.20}$$

where $E = \ln K_{\text{bond}}$ is the intrachain bond free energy and $B_2 = 0.85\ \sigma^3$ is the second virial coefficient obtained from simulation as in Ref. [63]. For simplicity, we ignore the steric end effect [27] here. Assumption of a Gaussian orientational distribution function $f_G(\theta) = (\alpha/4\pi)\exp(-\alpha\theta^2/2)$ gives the free energy penalty per chain [25–27] $\ln(\alpha/4)$, where α is related to the order parameter $S = (3\langle\cos^2\theta\rangle - 1)/2$ through $\alpha = 3/(1-S)$. The configurational entropy contribution is included as $\alpha/(4l_p)$, where l_p is the persistence length [25–27].

Equilibrium chain densities in the isotropic phase are described by $\rho(i) = \rho(1)q_1^{i-1}$, where $q_1 = \rho(1)K_{\text{bond}}$. The nematic phase is comprised of disordered and ordered components, therefore the total number density $\rho_{\text{tot}} = \rho_s + \rho_l = \sum_1^{L_c} i\rho(1)q_1^{i-1} + \sum_{L_c+1}^{\infty} i\rho(L_c)q_2^{i-L_c}$. The equilibrium condition is satisfied by equating monomer chemical potentials of the two components,

$$\mu_0 = \frac{1}{i}\frac{\partial f}{\partial\rho(i)} = \frac{\partial f}{\partial\rho(1)} = \ln\rho(1) + 2B_2\rho_{\text{tot}}; \quad i > L_c, \tag{2.21}$$

then we have

$$q_2 = \rho(1)K_{\text{bond}}\exp\left[\left(2 - \frac{8}{\sqrt{\pi\alpha}}\right)B_2\rho_l - \frac{\alpha}{4l_p}\right].$$

The value of α is numerically solved from the system of equations: $\rho_l = \sum_{L_c}^{\infty} i\rho(i)$ and $\rho_l = \rho_{\text{tot}} - \rho_s$.

The free energy is minimized with respect to the unbound monomer density $\rho(1)$ and the critical length L_c at any given ρ_{tot}, K_{bond}, and l_p. It is necessary to replace sums by integrals and sample L_c in a continuum space for the sake of accuracy and precision, especially at high densities. A method to obtain the phase coexistence range is that two equilibrated phases satisfy $\mu_{\mathrm{iso}}(\rho_{\mathrm{iso}}) = \mu_{\mathrm{nem}}(\rho_{\mathrm{nem}})$ and $p_{\mathrm{iso}}(\rho_{\mathrm{iso}}) = p_{\mathrm{nem}}(\rho_{\mathrm{nem}})$, from which we find the coexisting densities ρ_{iso} and ρ_{nem}.

Results and discussions

Figure 2.10 shows the optimized nematic length distribution near the phase boundary, which qualitatively agrees well with simulation. The deviation is increasing with higher association constants, or in other words, the number density. The average chain length of long chains, which is the negative inverse of the slope, is less than that of simulation, partly due to the omission of end effects and the simplification of the orientational distribution.

The phase diagram from the new "bi-exponential" length distribution model is shown in Fig. 2.11 along with the theory and simulation results from Ref. [63]. Allowing the short chains of the nematic phase to be orientationally disordered significantly shifts the nematic phase boundary to a lower density. Other factors that may contribute to the discrepancy between theory and simulation are the assumption of the Gaussian form for the orientational distribution and, especially at low K_{bond} and high ρ, the truncation of the virial series at the quadratic term.

The driving force for this isotropic-nematic phase transition is the competition between the orientational entropy and the translational (or packing) entropy. The

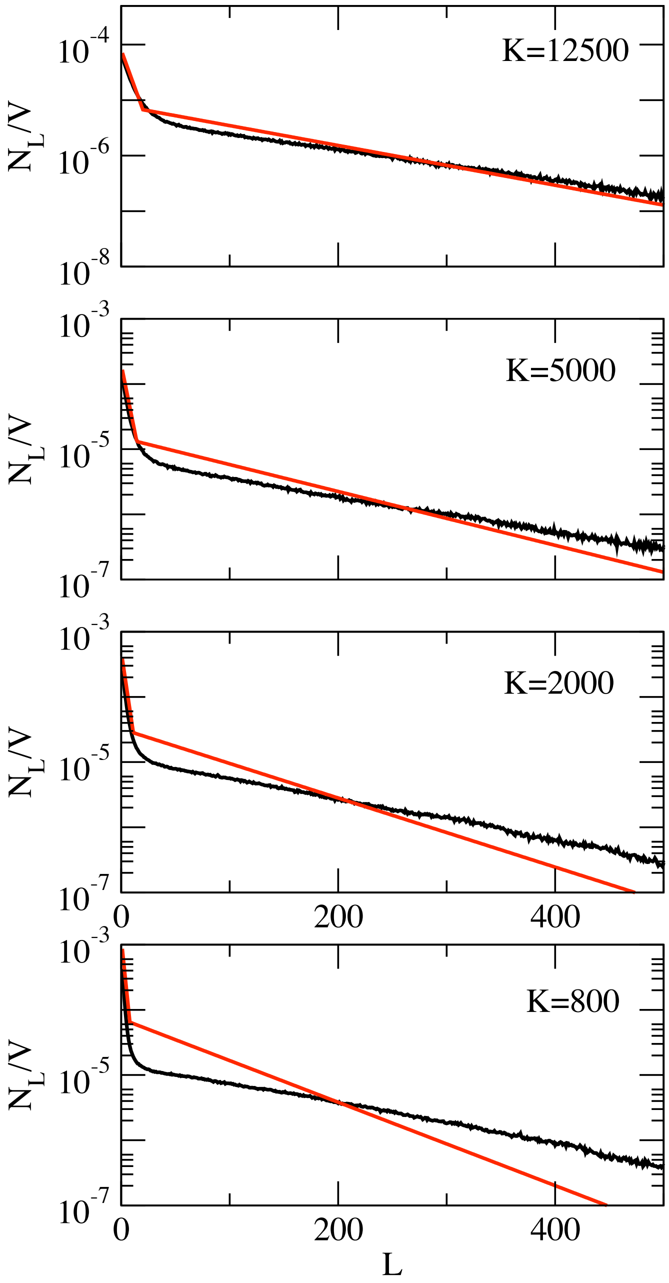


Figure 2.10: Semi-log plot of the chain length distribution: average number density of chains of contour length L versus L for K_{bond} and $l_p = 1000\ \sigma$ for the nematic phase at different constants. The top-to-bottom number densities are 0.118, 0.168, 0.221, and 0.304, respectively.

disordering of short chains maximizes the orientational entropy as well as increases the excluded volume interactions between short and long chains. The former stabilizes the system by decreasing the free energy while the latter results in higher packing and ordering of long chains. An optimal fraction of disordering short chains makes the phase transition occur at a lower density.

In principle, the critical length L_c is governed by the empty space between ordered long chains in the nematic phase, which depends on the density. In Fig. 2.12 the reciprocal of L_c is plotted against ρ and fitted by $1/L_c = a\rho$, where $a \approx 0.388$ is a fit coefficient independent of both the bond constant and the persistence length in a wide range. The universal relationship implies that the local structure of the nematic phase almost solely depends on the density, while varying K_{bond} and l_p only serves to change the average chain length. Not coincidentally, the same inverse relation between length and density is found in the Onsager theory for hard rods. When L_c approaches l_p in the extreme low density, L_c is underestimated systematically because the end-to-end distance is no longer linearly proportional to the contour length. Caution should also be taken if the persistence length is not long enough for a nematic phase to form.

To determine the conditions where the phenomenon of disordered short chains is most important, we plotted the fraction of monomers belonging to short chains over a range of bond association constants and densities in Fig. 2.13. At constant K_{bond}, this fraction drops rapidly with ρ as L_c decreases while the mean chain length

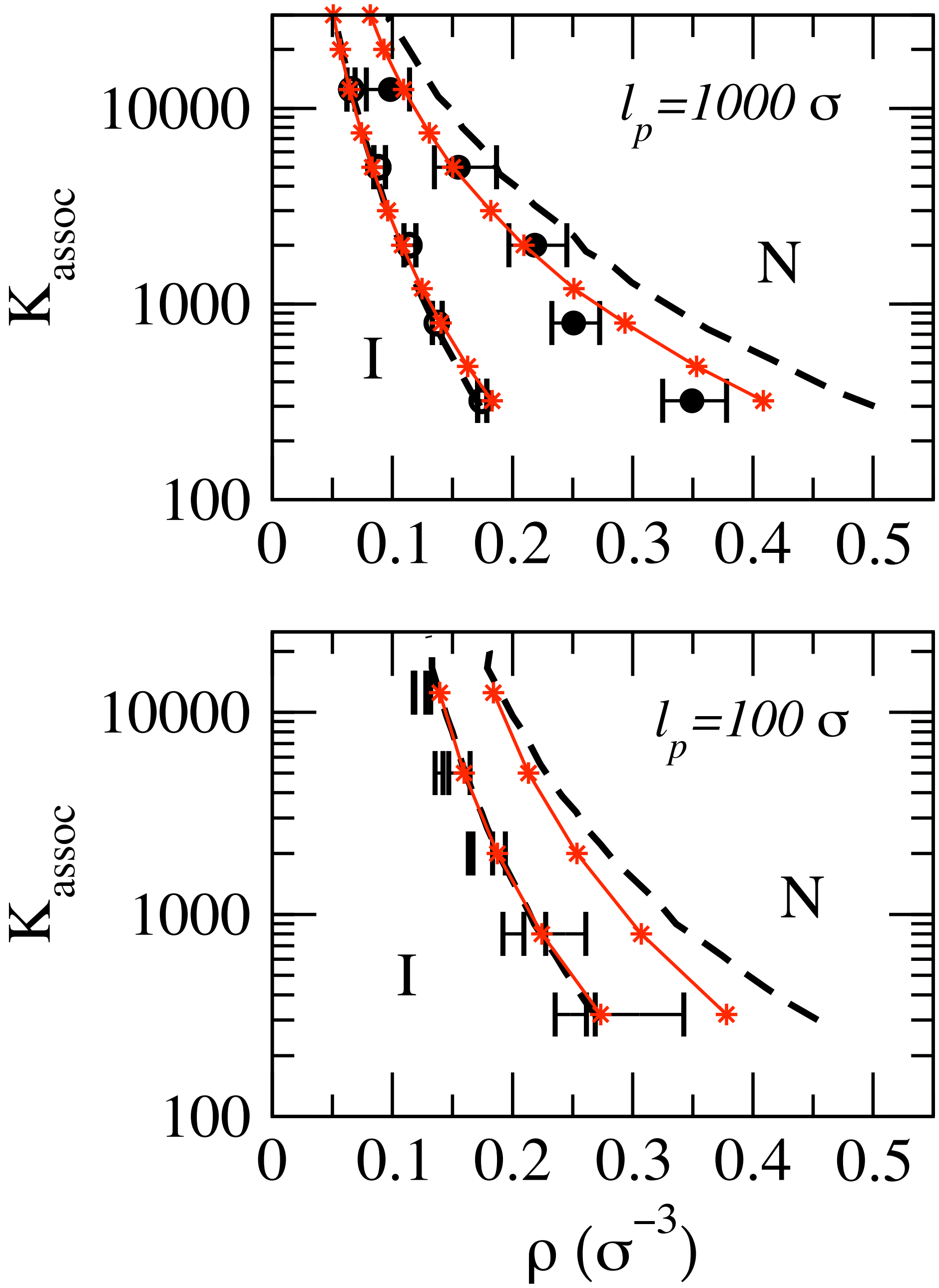


Figure 2.11: Comparison of phase boundaries obtained from both simulation and different theoretical models. Dots with error bars from simulation and dashed lines from the old model are as in Ref. [63]. Red star-dot curves are results of the new "bi-exponential" length distribution model.

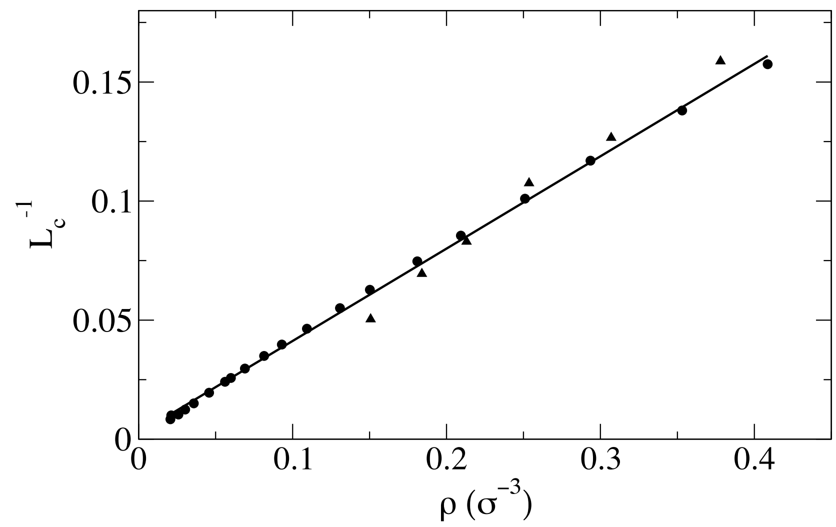


Figure 2.12: The critical length (reciprocal scale) versus the total number density at $l_p = 1000\ \sigma$. Dots are data at the nematic phase boundary for different association constants. The slope of the fit line is 0.388.

increases. The highest fraction can therefore be found at the phase boundary. As K_{bond} is increased and ρ decreases, a maximum fraction of around 4.6% is reached at $\rho \approx 0.06$. A qualitative explanation is that as the density increases, compact ordered long chains "squeeze" many not-that-short chains into the ordered state, i.e., L_c decreases as shown in Fig. 2.12, which results in a smaller fraction of monomers in short chains. At low density, although L_c is relatively large, high K_{bond}'s result in a great average chain length and a higher fraction of monomers entering the ordered long chains' component. The composition at the phase boundary is a result of the competition between the average chain length and L_c.

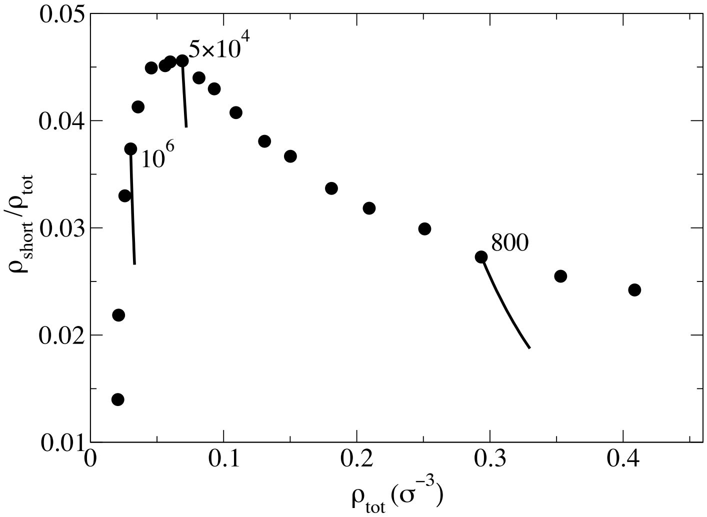


Figure 2.13: The fraction of monomers belonging to short chains at $l_p = 1000\ \sigma$. Dots are at the nematic phase boundary; three curve segments are in the nematic phase with $K_{\text{bond}} = 10^6, 5 \times 10^4,$ and 800, respectively.

2.4.4 Summary

We have used μVT Monte Carlo simulation to map out the phase diagram of a simple model for semiflexible equilibrium polymers. The PDIRR algorithm for changing aggregate size yielded a several-fold increase in efficiency in the equilibration of chain size distributions. We are able to roughly delineate not only the regions of stability of isotropic and nematic phases, but also the boundaries of the narrow coexistence region. The simplicity of the model affords a direct comparison with analytic theory of van der Schoot and Cates [27], giving quantitative agreement between theory (with two adjustable parameters) and simulation for chemical potential and mean chain length in isotropic systems over a wide concentration range in ten systems. Agreement between theory and simulation in the nematic phase is only qualitative, and simulation yields phase coexistence regions that are significantly narrower than predicted by theory. The most important novel observation from the simulations is that a strong dependence of the degree of orientational order on chain length for short chains leads to a bi-exponential chain length distribution. The scaling of the length M_{short} characteristic of the short chain component has been expressed in terms of the bond association constant, chemical potential, and total monomer concentration. We also shows that the presence of a short-chain disordered component in the nematic phase of a semiflexible equilibrium polymer lowers the phase transition point, and accounts for part, but not all of the discrepancy between mean field theory results and simulation. The analysis of the critical chain length L_c and the chain composition helps provide more insight to the isotropic-nematic phase transition of a self-assembled semiflexible polymer system.

Chapter 3

Monte Carlo simulation of actin self-assembly using a spherocylinder based coarse graining

Using grand canonical Monte Carlo simulations we study the equilibrium properties of actin self-assembly. The statistics of actin polymerization is described by a mechanism involving monomer activation and chain propagation with bond association constants derived from experimental free energy parameters. For efficiency in representing systems of extremely long, stiff chains we use a coarse-graining based on spherocylinders. We present results pertaining to the isotropic-nematic transition in this equilibrium biopolymer system.

3.1 Introduction

The actin protein abundantly exists in the cytoskeleton network of many living cells and plays important roles in cells' fundamental functions [10]. Actin has two structural states: unbound globular actin (G-actin) and polymerized long filamentous actin (F-actin). Under certain conditions, G-actin can reversibly self-assemble into F-actin through non-covalent forces [19, 20]. The actin self-assembly has been investigated extensively in many aspects such as the isotropic-to-nematic phase transition, the network structure and elastic behavior, etc. under *in vitro* conditions [23, 65, 66]. The actin filament is semiflexible and extremely polydisperse in length distribution (up to 70 μm) with a diameter of 8 nm and a persistence length of 10 $\sim$ 20 μm [65]. Particularly, the actin solution shows a exponential length distribution [20], and the isotropic-nematic phase transition of actin solutions has been examined in both theory and experiments. However, whether the transition is first-order or continuous is arguable in experiments [23, 67, 68], in contract with the first order prediction from the Onsager theory based on a long rod model. On the other hand, though the polymerization mechanism is not clear yet, it is believed to involve an unfavorable nucleation step followed by a favorable propagation step [16, 21]. The computer simulation is a convenient tool to study such a complicated system by using rather simple models to catch the main features of the actin self-assembly. In this article we use a hard spherocylinder chain to represent the actin filament in order to improve the computational efficiency as compared with the hard-sphere model. We mainly present the simulations results pertaining to the isotropic to nematic phase transition.

3.2 Simulation model

Our previous work [63] showed that the isotropic-to-nematic phase transition can be simulated by using a chain of tangent hard spheres to represent a polymer chain. However, the hard-sphere chain model simulation becomes computationally prohibitive when the polymer chain length is too long as in the case of actin filaments. In order to improve the simulation efficiency, in this work we model the actin filament as a series of jointed hard spherocylinders as shown in Fig. 3.1. The length of the cylinder rod is l_{rod} and the diameter is σ. Two successive spherocylinders are jointed by sharing the end sphere, which does not in fact exist and is represented by a dashed circle. The stiffness of the chain is subject to a square well potential function of the angle θ between two end jointed spherocylinders, where the maximum angle θ_m is related to the persistence length (l_p) through the cosine correlation function,

$$l_p = \frac{2 \times l_{rod}}{1 - \cos\theta_m}. \tag{3.1}$$

The formula is identical to that of the hard-sphere model except for that the bond length is replaced by the rod length (l_{rod}). If θ_m approaches zero gradually, the model transitions to the straight spherocylinder rod model.

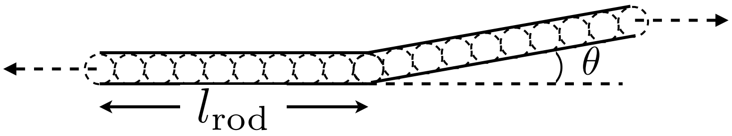


Figure 3.1: 2-D Schematic drawing of the coarse-graining model based on spherocylinders. θ is the angle between two adjacent spherocylinders. The dashed circles only serve to indicate the length relationship between the hard-spherocylinder model and the hard-sphere model, as well as the virtual end sphere shared by the two connected hardspherocyliders.

The association constant between spherocylinders is indirectly defined in terms of sphere monomers. The contour chain length L is related to the rod length by

$$L = n\ \sigma = (n' - 1)l_{rod} + 1, \tag{3.2}$$

where l_{rod} is practically set to be an integer multiple of σ and $n' - 1$ can be considered as the number of rods in a chain. In the hard-sphere chain model the weight function for a chain of length L is [63]

$$\omega_n = \frac{V K_{\mathrm{assoc}}^{n-1} \exp(n\beta\mu)}{N_{\mathrm{ch}} + 1}, \tag{3.3}$$

where V is the volume of the simulation box, K_{assoc} the association constant for two monomers, $\beta = (k_B T)^{-1}$, μ the chemical potential for a free monomer, and N_{ch} the number of chains existing in the system. For the spherocylinder model, the weight functions are rewritten accordingly as:

$$\omega_{1'} = \omega_1; \qquad \omega_{n'} = \sum_{i=(n'-2)l_{\mathrm{rod}}+2}^{(n'-1)l_{\mathrm{rod}}+1} \omega_i, \quad n = 2, 3, \cdots . \tag{3.4}$$

For example, if $l_{rod} = 20\ \sigma$, then $\omega_{2'} = \sum_{i=2}^{21} \omega_i$, $\omega_{3'} = \sum_{i=22}^{41} \omega_i$, and so on. That is to say, the contour length can be $1, 21, 41, 61 \cdots$ with an incremental length of l_{rod}.

In practice, the spherocylinder chain is determined by the coordinates of the two ends and joint monomers. The overlap check involves the sphere-spherocylinder and spherocylinder-spherocylinder checks. The former one can be done by computing the shortest distance between a line segment and a point. In the latter case we compute the shortest distance between two line segments in 3-dimensional space and compare it with the diameter (σ), of which a detailed account is given in Ref. [69]. In the following, we give a summary of the overlap check algorithm.

The problem is divided into two parts:

- Find the perpendicular distance ($d_\perp$) between two lines on which the two line segments lie on.
- Project the two lines onto a plane normal to the perpendicular distance vector. Then, find the shortest distance (d_{in}) within the plane between the two line segments.

The shortest distance between two line segments is then

$$d_{\mathrm{shortest}}^2 = d_\perp^2 + d_{\mathrm{in}}^2. \tag{3.5}$$

The two lines can be described by two parametric vector equations:

$$\vec{r}_i(t) = \vec{r}_{i,c} + t \times \vec{u}_i; \qquad i = 1, 2, \tag{3.6}$$

where $\vec{r}_{i,c}$ and $\vec{u}_i$ denote the center positions and directional unit vectors of the two line segments. The shortest distance vector $\vec{d}_\perp = \vec{r}_1(\lambda) - \vec{r}_2(\mu)$ is obtained from the fact that it is normal to both lines:

$$\vec{r}_\perp \cdot \vec{u}_1 = 0; \quad \vec{r}_\perp \cdot \vec{u}_2 = 0. \tag{3.7}$$

The minimum distance between the two lines is then $d_\perp = |\vec{d}_\perp|$.

The two lines are projected onto a plane normal to $\vec{d}_\perp$. The intersection point is given by (λ, μ). Taking this point as the origin, the two line segments are $\vec{r}_1 = \gamma\vec{u}_1$ and $\vec{r}_1 = \delta t_2 \vec{u}_2$, where $-l_{rod} < \gamma - \lambda < l_{rod}$ and $-l_{rod} < \delta - \mu < l_{rod}$. The constraints define a rectangle in the $\gamma\delta$-plane. The shortest in-plane distance $(d^2_\perp)$ is the minimum of $d^2(\gamma, \delta) = \gamma^2 + \delta^2 - 2\gamma\delta(\vec{u}_1 \cdot \vec{u}_2)$. The minimization algorithm is as follows: if the origin is inside the rectangle, that is, on both line segments, the shortest in-plane distance is zero. If not, find the end (γ_m or δ_m) closet to the origin and minimize the in-plane distance with respect to one of the parameters. With d^2_{in} we readily obtain the shortest total distance from Eq. 3.5. The relative positions of two line segments in plane can also be categorized into three cases as shown in Fig. 3.2.

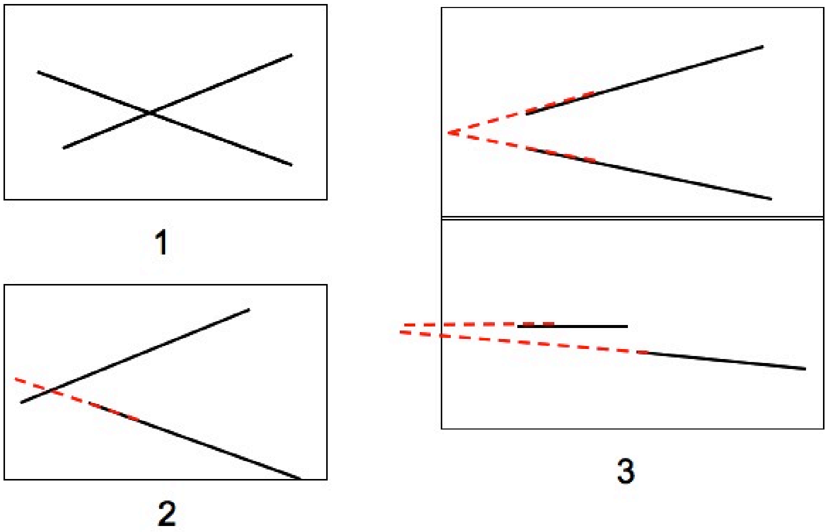


Figure 3.2: Three cases of the relative positions of two line segments in plane. (1) The intersection point is on both line segments, $d_{\text{in}} = 0$; (2) The intersection is on only one of the two line segments, d_{in} is the end-to-line distance; (3) The intersection is not on either line segment, d_{in} is the end-to-end distance.

3.3 Mechanism and thermodynamics of actin polymerization

Generally, the polymerization of actin consists of nucleation and propagation steps. The mechanism is usually studied from the viewpoint of kinetics [70], where rate constants are obtained from data fitting as a function of time. Recently, Niranjan et al [16, 17] showed that temperature can also be used as a measurement variable and the polymerization curve at equilibrium has been fitted effectively to give free energy parameters for a proposed reaction mechanism. In this Monte Carlo study, we follow the same basic reaction model with some necessary modifications.

Here we summarize the proposed mechanism and the derivation of thermodynamic parameters for actin polymerization in Ref. [16]

$$A_1 \rightleftharpoons A_1^* \qquad \text{(monomer initiation)}, \tag{3.8}$$

$$A_1^* + A_1^* \rightleftharpoons A_2 \qquad \text{(dimerization)}, \tag{3.9}$$

$$A_2 + A_1^* \rightleftharpoons A_3 \qquad \text{(trimerization)}, \tag{3.10}$$

$$A_i + A_1 \rightleftharpoons A_{i+1} \qquad i = 3, 4, \cdots \quad \text{(propagation)}, \tag{3.11}$$

where A_1 is a G-actin monomer, A_1^* an activated monomer, and A_i a polymerized actin i-mer. The equilibrium constants for the trimerization step and the propagation step are simplified to be identical so that only three different equilibrium constants (K_{init}, K_{dim} and K_{prop}) are needed to determine the equilibrium distribution of actin species at a given temperature T. Each K is related to the free energy change by

$$K = \exp[-\Delta G/RT] = \exp[-(\Delta H - T\Delta S)/T], \tag{3.12}$$

where ΔH and ΔS are the reaction enthalpy and entropy changes, respectively. Based on the Flory-Huggins theory, all K's can be determined from the equilibrium volume fraction (ϕ_i) for each actin species. The extent of actin polymerization here is defined as $\Phi = (\phi_1^0 - \phi_1 - \phi_1^*)/\phi_1^0$, where ϕ_1^0, the initial volume fraction of G-actin, is given by the conservation of mass

$$\begin{aligned} \phi_1^0 \;=\;& \phi_1 + K_{\text{init}}\phi_1 + 2K_{\text{dim}}\phi_1^{*2} \\ &+ \frac{K_{\text{init}}^3 K_{\text{dim}} K_{\text{prop}}(3 - 2K_{\text{prop}}\phi_1)\phi_1^3}{(1 - K_{\text{prop}}\phi_1)^2}. \end{aligned} \tag{3.13}$$

Three pairs of $(\Delta H, \Delta S)$ are obtained from the data fitting.

In our grand canonical Monte Carlo simulation, equilibrium constants obtained from $(\Delta H, \Delta S)$ pairs are converted in accordance with the appropriate units. The activated monomers are treated the same as unactivated monomers as we have no intention to distinguish two kinds of monomers in our simulation. The resultant equilibrium constants are written as:

$$K'_{\text{dim}} = \frac{[A_2]}{[A_1][A_1]} = \frac{[A_2]}{[A_1^*][A_1^*]} \times \left(\frac{[A_1^*]}{[A_1]}\right)^2 = K_{\text{dim}} K_{\text{init}}^2; \tag{3.14}$$

$$K'_{\text{trim}} = \frac{[A_3]}{[A_1][A_2]} = \frac{[A_3]}{[A_2][A_1^*]} \times \frac{[A_1]}{[A_1^*]} = K_{\text{prop}} K_{\text{init}}; \tag{3.15}$$

$$K'_{\text{prop}} = K_{\text{prop}}, \tag{3.16}$$

where primed K's denote association constants used in simulation. We choose the physiological temperature $T = 310\ K$ and the total G-actin concentration 2 mg/ml, from which we obtain three association constants for simulation $K'_{\text{dim}} \approx 87\ \sigma^3$, $K'_{\text{trim}} \approx 5704\ \sigma^3$ and $K'_{\text{prop}} \approx 25507\ \sigma^3$, where σ is the monomer diameter.

The association constants are applied to produce a reasonable equilibrium distribution of actin chain length. As discussed in Ref. [16], the free energy parameters of initiation and dimerization are sensitive to the initial G-actin concentration, that is, the association constants depend on the initial G-actin concentration, which makes it practically impossible to reproduce the exactly same initial G-actin concentration and the extent of polymerization. The other thing might be of note is that in this Monte Carlo simulation, the temperature is incorporated into the K's and Boltzmann factor rather than defined directly.

3.4 Simulation results and discussion

In our simulation, the rod length $l_{\text{rod}} = 20\ \sigma$; the persistence length $l_p = 1000\ \sigma$, which has the same order of magnitude as the experimental measurement; the simulation cubic box edge length is $300\ \sigma$ with periodic boundary conditions. The resultant ratio of the mean chain length to the box size is the same magnitude as our previous simulation [63], where the relatively small box size does not severely affect the ordering transition.

The hard spherocylinder differs from the hard-sphere chain in both the actual occupied volume and the excluded volume. Nevertheless, we continue to use the number density in terms of monomers in this context. Also the spherocylinder should has a good approximation of a stiff hard-sphere chain.

For the isodesmic model, the mean-field free energy per unit volume can be minimized with respect to the mean chain length M [27, 63]. Based on Eq. (2.11) for the isotropic phase, if the fraction of monomers and dimers is negligible, the free energy density is written as

$$\begin{aligned} f_{\text{iso}} &= -\frac{\rho}{M}[E_1 + E_2 + E_3(M-3)] + \left(B + \frac{\kappa_I}{M}\right)\rho^2 && (3.17)\\ &\quad + \frac{\rho}{M}\left(\ln\frac{\rho}{M} - 1\right) + \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M, && (3.18)\end{aligned}$$

where E_1, E_2 and E_3 correspond to three different association constants in the actin polymerization, respectively; ρ is the number density of monomers; B is the second virial coefficient; κ_I reflects the the chain end effect and is negligible in this case. The physical meanings of the formula is elucidated in Ref. [63] or Chapter 2.

$$M_{\text{iso}}(\rho) = 0.5 + 0.5\sqrt{1 + 4\rho\exp(3E_3 - E_1 - E_2 + \kappa_I\rho)}, \tag{3.19}$$

where we have the relationship

$$\exp(3E_3 - E_1 - E_2) = K_3^3/(K_1K_2). \tag{3.20}$$

On the other hand, the relationship between M and ρ can also be derived through a kinetic analysis. The number densities of different actin species are constrained by the association constants similar to the analysis of the extent of polymerization in section 3.3. At equilibrium, from $i\frac{\partial f}{\partial \rho_1} = \frac{\partial f}{\partial \rho_i}$ we obtain the total monomer number density (ρ) and the chain number density (ρ_{ch}):

$$\rho = \rho_1 + 2\rho_1^2 K_1 + \frac{3 - 2\rho_1 K_3}{(1-\rho_1 K_3)^2} \cdot \rho_1^3 K_1 K_2, \tag{3.21}$$

$$\rho_{\mathrm{ch}} = \rho_1 + \rho_1^2 K_1 + \frac{\rho_1^3 K_1 K_2}{(1-\rho_1 K_3)^2}, \tag{3.22}$$

where ρ_i denotes the number density of i-mer. The average chain length $M = \rho/\rho_{ch}$ is obtained numerically from Eqs. 3.21 and 3.22 at a given concentration of the free monomer (ρ_1). In fact, the result is a simple exponential distribution except for the over-populated monomers and dimers. Therefore, we also compute the average chain length excluding the free monomers and dimers:

$$\begin{aligned} M_{\mathrm{long}} &= \frac{\rho - \rho_1 - 2\rho_1^2 K_1}{\rho_{\mathrm{ch}} - \rho_1 - \rho_1^2 K_1} \\ &= \frac{3 - 2\rho_1 K_3}{1 - \rho_1 K_3}. \end{aligned} \tag{3.23}$$

The results from simulation and theory are shown in Fig. 3.3. For M_{long}, the free energy analysis and the kinetic theory agree very well in a certain range of low density. The simulation data are also qualitatively in line with the theoretical prediction. However, for the overall chain length, we need to more accurate on the approximation in the free energy equation.

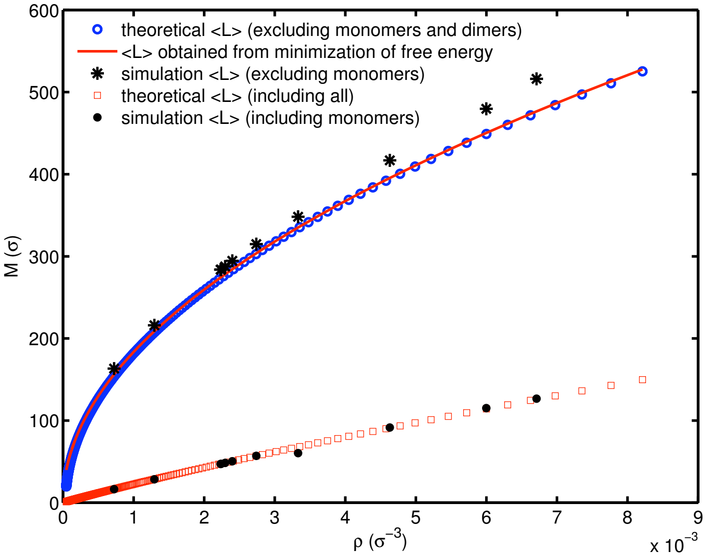


Figure 3.3: Comparison of theoretical predictions to simulation results for isotropic phases in low density regime: the mean chain length as a function of the monomer number density.

3.4.1 Comparison with the sphere-chain model

As discussed in Ref. [63], for a certain association constant, the isotropic-nematic phase transition only occurs when the persistence length is long enough, or in other words, the chain is stiff enough. The phase diagrams for $l_p = 100\ \sigma$ and $1000\ \sigma$ showed that the phase coexistence range becomes much narrower with increasing bonding energy, but the transition remains first-order in our simulation. Although we can explore much larger association constants, the current spherocylinder model still cannot overcome the limit of expensive computational cost with increasing K_{assoc}. Therefore, we have no conclusive results for the question whether the nature of the isotropic-nematic phase transition in the limit case of K_{assoc} is affected by the chain length polydispersity.

3.4.2 The phase transition of actin filaments

In experiments the threshold concentration for the isotropic-nematic transition is about 2 mg/ml [22]. As a benchmark, 1 mg/ml is equal to 24 μM [10], which can be converted to the number density 0.0074 as used in our model. In our simulations the I-N phase transition is first-order and occurs in the extremely low density regime. With the three association constants deduced from section 3.3, the transition number density is around 0.012, which is in the same magnitude of the experimental observations. The coexistence region is relatively narrow, which is in line with the prediction of

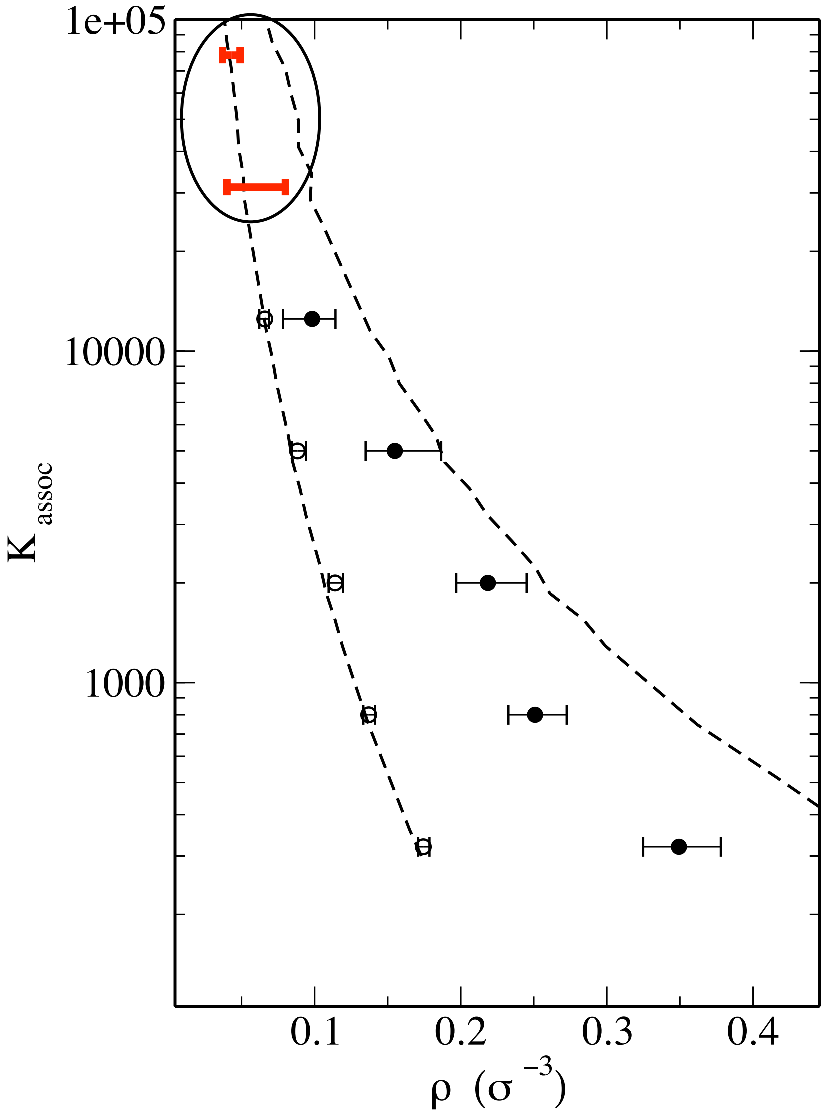


Figure 3.4: The isotropic-nematic phase transition regions for extremely large association constants. Ovaled red bars are the coexistence regions on the phase diagram taken from Fig. 2.8 $l_p = 1000\ \sigma$ in Chapter 2. No phase transition points as accurate as in Chapter 2 are obtained using the spherocylinder model.

the isodesmic model that the coexistence region narrows down with increasing K_{assoc}, i.e., the mean chain length. The order parameter of the nematic phase is ~ 0.7, which is consistent with experimental results of semiflexible actin filaments in a quite dilute solution. Compared to Ref. [16], we were not able to reproduce the extent of polymerization under similar conditions such as concentration and temperature.

3.4.3 The length distribution

In general, the observed length distributions are very similar to the simulation results in Chapter 2 or the isodesmic model in section 3.4.1. The difference lies in the overpopulated free monomers, which results from the aggregation barrier of the dimerization step. Beyond that, the propagation step dominates the behavior of the chain system (the mediocre trimerization step does not contribute much to the length distribution, Ref. to section 3.3). Therefore, the rest part of the length distribution roughly obeys the same rule as observed in Chapter 2: an exponential distribution in the isotropic phase and a "biexponential" distribution in the nematic phase. In the latter case, the two distinct regimes for ordered long chains and disordered short chains are not as profound as the isodesmic sphere-chain model, which can be attributed to the facts such as the nonuniform association constants and the extremely low density. Compared to the isodesmic model, the barrier of the dimerization results in a much greater mean chain length with a relatively low bond energy for the propagation step, which is the main reason for the resulting low transition point. Showed in Figs. 3.5 and 3.6 are the isotropic and nematic length distributions, respectively. In these simulation results, we did not include free monomers for the sake of geometrical consistency in the model, that is, the constituent monomer is a spherocylinder instead of a sphere (an extension from the original model).

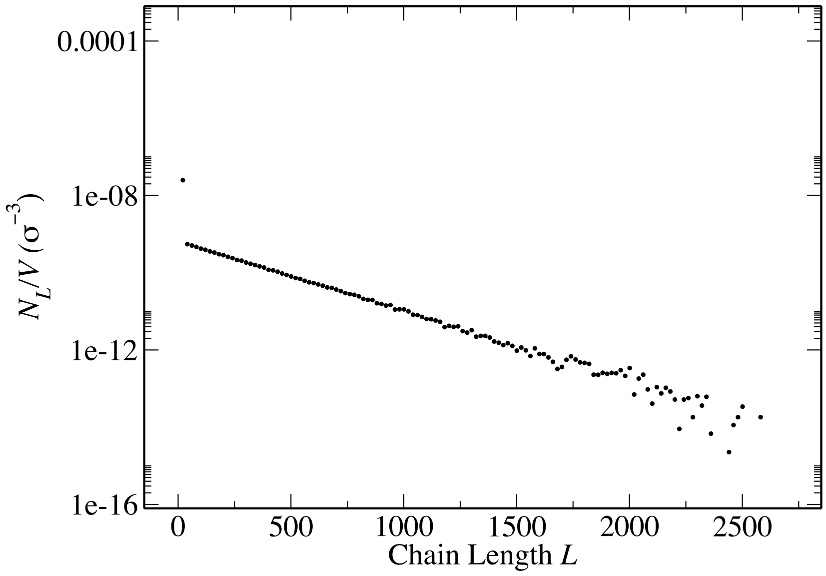


Figure 3.5: The length distribution in the isotropic phase. The simulation box is $(300\ \sigma)^3$. $l_p = 1000\ \sigma$. The number density is about 0.00073.

3.5 Summary

We have used a spherocylinder based coarse grained model to represent extremely long polymer chains in a grand canonical Monte Carlo simulation. First, we

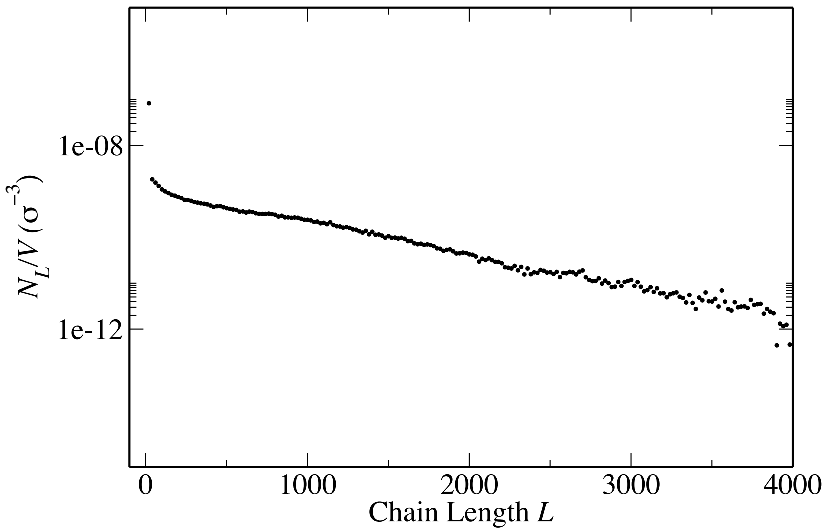


Figure 3.6: The length distribution in the nematic phase. The simulation box is $(300\ \sigma)^3$. $l_p = 1000\ \sigma$. The number density is about 0.00124.

have performed simulations using the isodesmic model for the limit case of large association constant. We observed the I-N coexistence range consistently narrows down with increasing K_{assoc}, however questions remains whether the I-N phase transition of self-assembled chain polymers remains first-order as in the case of monodisperse infinite long rods, or becomes continuous due to complications such as polydispersity. Second, we incorporated nonuniform association constants deduced from an experimental study into our model to simulate the actin filaments. The simulation results are considerably consistent with experimental observations.

Chapter 4

Monte Carlo simulation of self-assembled polymer chains with inter-chain attractions

4.1 Introduction

Actin filaments (F-actin) are stiff chains formed by the reversible self-assembly of globular actin protein monomers (G-actin); the filaments themselves organize into networks and bundles that constitute a major component of they cytoskeleton [11, 32]. In the high density regime, actin filaments undergo the isotropic-nematic phase transition in *in vitro* experiments [22, 23]. However, the *in vivo* biological system is usually abundant in attraction agents such as crosslinking proteins and polyvalent counterions, which gives rise to the inter-chain interactions. With this inter-chain interaction, the actin filaments can form a network of bundles, rings, etc. [14, 71, 72] This behavior has also been observed with a wide range of biological polymers such as DNA and microtubules [73–75]. *In vitro* experiments have shown that short-range inter-chain attractions narrow the length distribution, which was interpreted as indicating the attraction induced filament bundling [76]. However, to directly observe the bundles of actin filaments using techniques such as confocal microscopy and TEM, most authors focused instead on phalloidin stablized F-actin solution [72, 77]. For example, Kwon et al. [77] utilized polycations as attraction agents in a F-actin solution with a quenched actin length distribution. The authors measured the bundling threshold as well as the bundle thickness and length, and proposed an anisotropic nucleation-growth mechanism based on the analysis of the bundle's length and diameter growth.

Recent efforts to understand the morphologies and structures of this type of system include both theoretical and simulation works. Monte Carlo simulations have been applied to study the isotropic-nematic phase transition of equilibrium self-assembled polymers with only excluded volume interaction [63], and whereof the unique biexponential length distribution in the nematic phase (instead of an exponential distribution in the isotropic phase) was also analyzed theoretically [78]. However, due to the anisotropy of actin filaments, parallel chains have the maximized inter-chain attraction energy gain, which by forming aligned bundles wins out in the competition with the translational and orientational entropy. Depending on the physical properties of the system, the linker-assisted fixed-length filament aggregation can take different structures or phases such as cubatic and bundle phases [79], of which the theory was modified to explain the narrowed length distribution to some extent [76]. Kierfeld et al. used both theory and Monte Carlo simulations to show that the unbinding of fixed-length filament bundles is discontinuous and the bundle tends to separate into several sub-bundles due to slow kinetics [80]. Sayer et al. [81] performed Monte Carlo simulations of network formation from free monomers through chain self-assembly and bundling. Even with an effective bond fluctuation method, the authors did not observe complete phase separation as expected in thermodynamic equilibrium. Additionally, the length distribution issue was not addressed quantitatively.

In this work, we focus on the short range attraction induced bundle formation of self-assembled equilibrium semi-flexible chain polymers using canonical ensemble Monte Carlo method. The coupling between the reversible chain polymerization and the attraction energy gain from aligned chains strongly alters the length distribution. The resultant bundle is analyzed both qualitatively and quantitatively. Grand canonical ensemble simulation results are also presented as a comparison, showing that for the current model and conditions, the formation of discrete bundles is unstable with respect to condensation of a macroscopic dense aligned phase.

4.2 Model

In this work, we use the same model as in Ref. [63]. A polymer chain is modeled as a string of tangentially connected hard spheres with diameter σ, which is equal to the bond length in this case. The reversible monomer association and dissociation on the end of a chain is controlled by an association constant $K_{\mathrm{assoc}} = \exp(\beta E_{\mathrm{bond}})$, where E_{bond} is the bond energy and $\beta = 1/(K_{\mathrm{B}}T)$. The polymer flexibility is controlled independently by an infinite square well potential function of bond angle θ: $U(\theta) = 0$ for $\theta_{\min} < \theta \leq \pi$; $U(\theta) = \infty$ for $\theta \leq \theta_{\min}$. Additionally, we apply a square-well isotropic attraction potential function to phenomenologically describe the inter-chain attraction's dependence on the monomer-monomer separation r,

$$u(r) = \begin{cases} \infty & r \leq \sigma \\ \epsilon & \sigma < r \leq \sigma + \delta \\ 0 & \sigma + \delta < r. \end{cases} \tag{4.1}$$

Here, note that an appropriate δ can exclude the adjacent monomer-monomer attractions within a chain, though the self-attraction of a polymer chain is intrinsically allowed, especially when a chain is flexible enough to coil up.

Our off-lattice canonical ensemble Monte Carlo simulations are performed in a cubic cell with standard periodic boundary conditions, more specifically, the number of monomers (N), the volume (V), and the temperature (T) are fixed. We apply two main types of Monte Carlo moves:

1. Single monomer move

   In each step there are two possibilities. 1) An end monomer of a chain becomes a free monomer; 2) a free monomer is attached onto another randomly selected chain. The exception is the dimerization step and its reverse one, which involves two free monomers. This conventional move involves chain-end bond formation or breaking, which guarantees the equilibration process in simulation. If N_ch is defined as the total number of chains including free monomers, the acceptance probability for bond formation is $\min[1, W]$ ($\min[1, 1/W]$ for bond breaking), where W is obtained from the condition of detailed balance:

   $$W = \frac{1}{V} K_{\text{assoc}} \exp[-\beta\epsilon\Delta n_{\text{attr}}] N_{\text{ch}} 2\pi(1 + \cos\theta_{\text{min}}). \tag{4.2}$$

   Here $\Delta n_{\text{attr,j}}$ is the change of the number of attractions in this step. The acceptance probability is the same for the dimerization step.

2. Multiple monomers move

   This move allows multiple monomers to be transferred from one chain to the other in one single Monte Carlo step, in the same spirit of the PDIRR algorithm as in Ref. [63], First, two chains ($\geq$ 2-mers) are randomly selected, and one end is randomly chosen out of each chain's two ends. Monomers are then sequentially transferred from one end to the other and vice versa until the transfers are interrupted due to the excluded-volume interaction or a dimer is reached (to simplify the calculation of the acceptance probability). The weight factor w_i for an i-mer move is computed after each successful monomer move:

$$w_i = \sum_{j=1}^{i} \exp(-\beta\epsilon\Delta n_{\mathrm{attr},j}), \tag{4.3}$$

where $\Delta n_{\mathrm{attr,j}}$ is the change of the number of isotropic attractions after the j-th monomer is transferred. Second, the probability of the transfer is given by $\min[1, W]$, where W is defined by

$$W = \sum_{i=1}^{i_{\max}} w_i. \tag{4.4}$$

Then, the number of transferred monomers and the direction of the transfer are determined by the acceptance probability ratio w_i/W.

As shown in the formulae, the success rate of this multiple monomers move depends only on the inter-chain attraction energy change and the excluded volume interaction since the number of bonds stays constant. This Monte Carlo move speeds up the equilibration of existing polymer chains while the number of chains remains constant, i.e., the mean chain length does not vary.

The simulation starts from a box of randomly inserted unbound hard spheres (or whatever appropriate initial configuration), eventually will form a equilibrium polydisperse chain system. With a carefully chosen simulation box size and periodic boundary conditions, the finite size effect can be reduced to be negligible, that is to say, if the aggregate size is relatively small compared with the simulation cell's dimension. With inter-chain attractions, the simulation can generate a variety of phases with different combinations of physical parameters: the polymer chain flexibility (or stiffness), the bonding association constant, the concentration of monomers, and the attraction strength. The general procedure is that we first adjust $\exp(-\beta\epsilon)$ to a relatively high value for the aggregation to occur, then we gradually lower $\exp(-\beta\epsilon)$ to achieve equilibrium bundles and obtain the transition boundaries. Starting from right above the transition boundary, we can also increase $\exp(-\beta\epsilon)$ incrementally and compare the results to that of the decremental procedure.

4.3 Results and discussions

Throughout this study, we set $\sigma = 1$. The attraction range $\delta = \sqrt{2} - 1 \approx 0.414$, which is the distance between two monomers separated by one monomer within a chain when the bond angle is $\pi/2$. This value is chosen so to exclude the the intra-chain monomer-monomer attraction. We performed simulation with a wide range of physical parameters. The simulated concentrations (or volume fraction $\phi = (1/6)\pi\sigma^3 N/V$) are far below the isotropic-nematic phase transition concentrations determined from chains with only excluded volume interactions [63]. In the following, the simulation cube size is $(160\ \sigma)^3$, except where noted.

4.3.1 Phase behavior and diagram

For a constant volume fraction $\phi = 0.082\%$, we performed extensive simulations for three persistence lengths, $l_p = 10\ \sigma$, $100\ \sigma$, and $1000\ \sigma$, respectively. With small association constants or weak attractions, the chain system is in the isotropic phase. Increasing either K_{assoc} or $\exp(-\beta\epsilon)$ leads to a transition to a microphase separated phase, where multiple chains align together to form a cluster at equilibrium with surrounding dilute isotropic chains. The transition boundaries from the simulation are summarized in Fig. 4.1. The squares are the upper boundary where the isotropic phase stays stable before condensation with incrementally increasing attraction. While the attraction strength is decreased, the bundle will not fall apart until the values of $\exp(-\beta\epsilon)$ reach the circle points. This strong hysteresis can be attributed to the fact that the transition involves a high energy barrier, therefore, both phases are kinetically stable in the region between the two boundaries. Besides direct observation of the bundling from snap shots, the most convenient criterion is the average number of attractions per monomer $\langle n_{\text{attr}} \rangle$, which is nearly zero in the isotropic phase and increases dramatically with nascent bundles.

Figure 4.2(a) is a snapshot of the isotropic phase for $l_p = 1000\ \sigma$. At this persistence length, the mean chain length in our simulation is far less than than l_p, therefore, a well-defined bundle can be observed, as shown in Fig. 4.2(b), with strong enough inter-chain interactions. With increasing chain flexibility, or in other words,

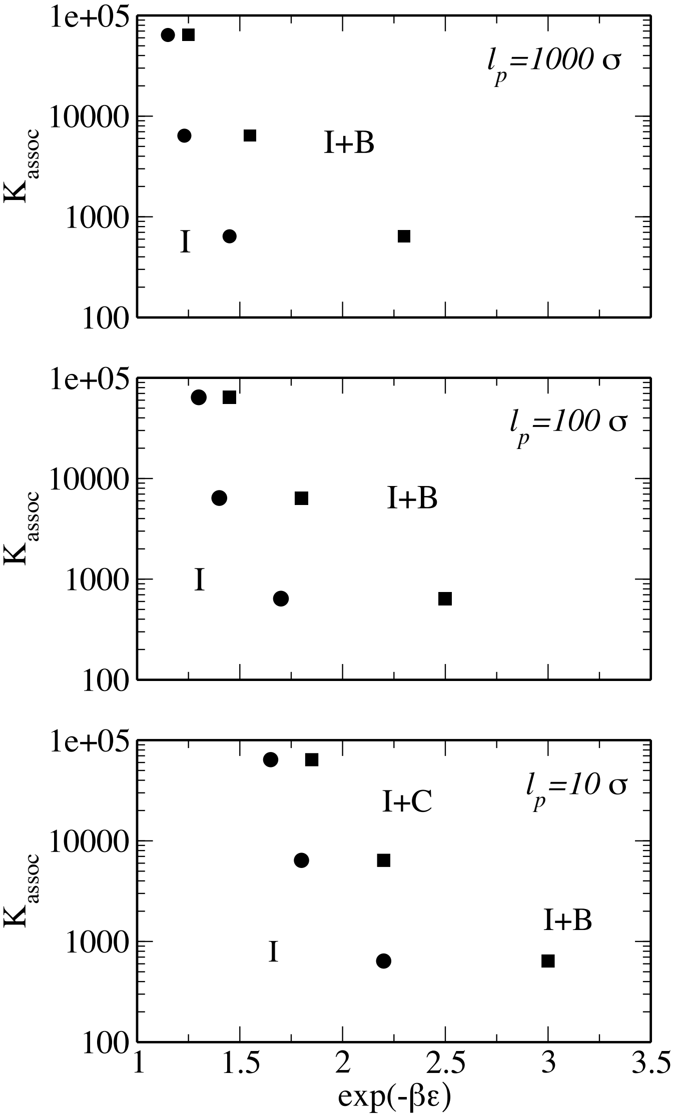
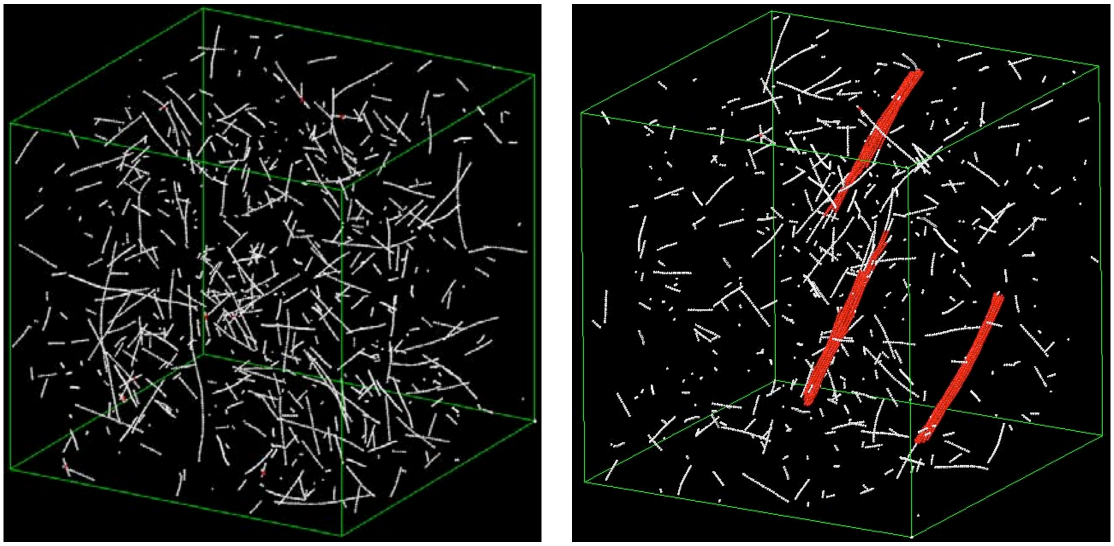


Figure 4.1: Phase diagrams of the phase separation with different persistence lengths and attraction strengths. The volume fraction is 0.082%. I stands for isotropic, B for bundle, and C for cluster (globule or toroid). Circles are the lower boundary; squares are the upper boundary.

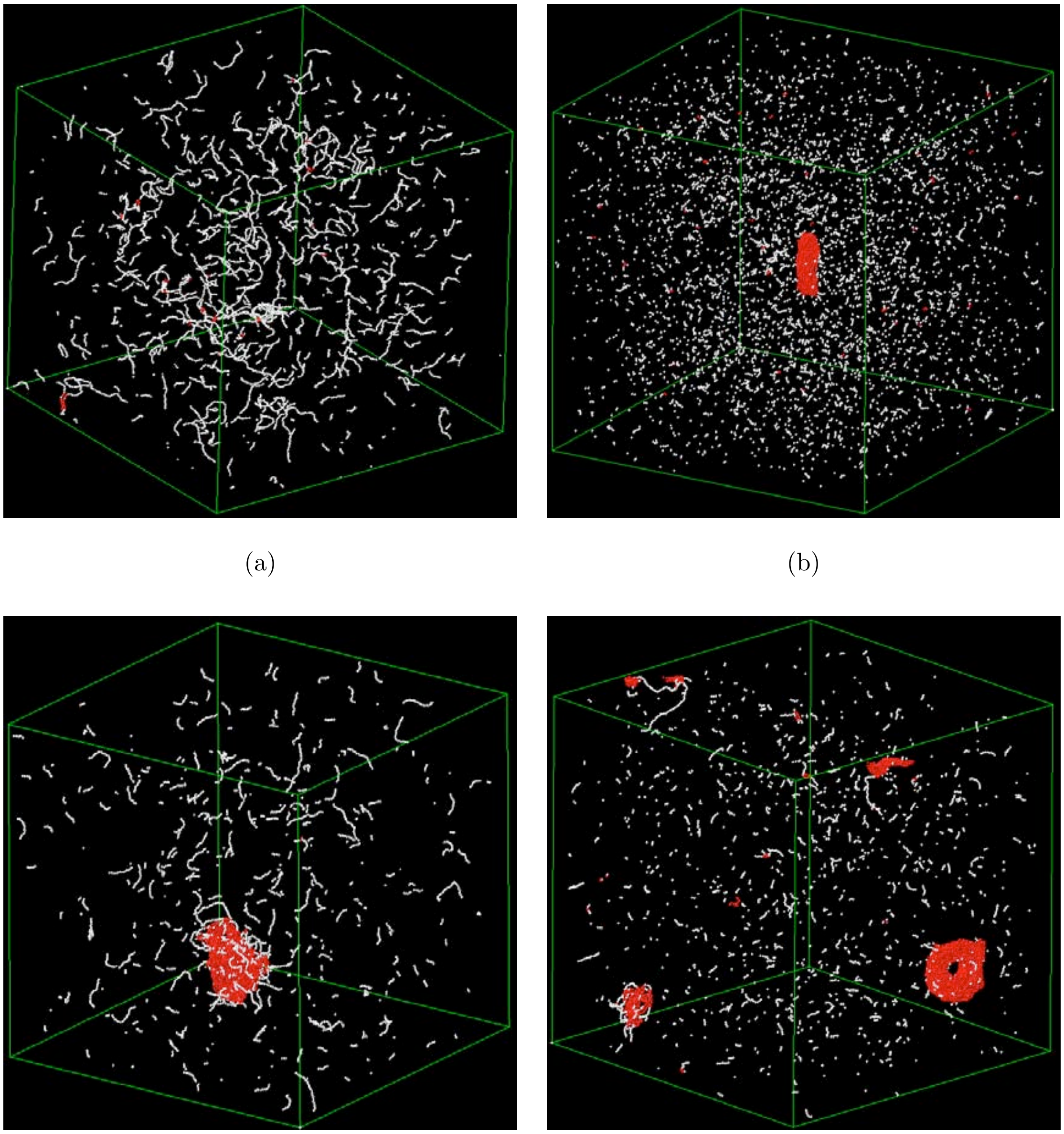
decreasing persistence length, the transition boundaries shift to the higher values of $\exp(-\beta\epsilon)$.



(a) (b)

Figure 4.2: Snapshots from the simulation, where monomers with non-bonded attractions are highlighted in red. (a)Isotropic phase with $\exp(-\beta\epsilon) = 1.15$. (b) Bundled phase with $\exp(-\beta\epsilon) = 1.16$, the bundle wraps across the periodic boundaries twice. For both (a) and (b), $l_p = 1000\ \sigma$ and $K_{\text{assoc}} = 64000$.

The phase behavior becomes different when the mean chain length is comparable with the persistence length as in the case $l_p = 10\ \sigma$. The isotropic phase for flexible chains is almost identical with the stiff chain case. If the association constant is so small that chains much longer than 10-mers seldom form, a short bundle can be observed. It is not the case anymore with large K_{assoc}'s as shown in Fig. 4.3, where the cluster can take shapes such as globule and toroid, depending on the simulation details, such as the



(a) (b)

(c) (d)

Figure 4.3: Snapshots from the simulations, where monomers with non-bonded attractions are highlighted in red. (a) Isotropic phase with $\exp(-\beta\epsilon) = 1.8$ and $K_{\text{assoc}} = 6400$. (b) Short bundle with $\exp(-\beta\epsilon) = 2.3$ and $K_{\text{assoc}} = 640$. (c) Globule with $\exp(-\beta\epsilon) = 1.7$ and $K_{\text{assoc}} = 64000$. (d) Toroid: an intermediate to convergence with $\exp(-\beta\epsilon) = 2.0$ and $K_{\text{assoc}} = 6400$. For all snapshots, $l_p = 10\ \sigma$.

physical parameters, the initial configuration and the number of Monte Carlo steps . Figure 4.3(c) is a globule where polydisperse chains irregularly coil up and form a cluster with flexible chain ends or segments sticking out. Figure 4.3(d) is an intermediate stage of a toroid, where finally the other smaller free cluster will disappear and coalesce into the big toroid. These phenomena were also observed in the coarse-grained simulation work by Noguchi et al. [82, 83] for the collapse of a single homopolymer chain, as a model of DNA toroid formation [73, 74]. The difference is that in our simulation the toroid or globule contains multiple chains rather than a single chain coil. The flexible chains can be attracted to themselves, which is different from the stiff chain case. Due to the relatively irregular packing of flexible chains compared with bundles of stiff chains, the value of $\langle n_{\mathrm{attr}} \rangle$ is significantly less than that of bundles.

The bundle formation is essentially a process of nucleation and cluster formation. Analogous to simulations of gas-liquid nucleation, multiple aggregates formed in the early stage relax into a single bundle (cluster) in equilibrium with a limited number of chains. Similarly, bundles with branched ends, as shown in Fig. 4.4, only appear in the intermediate stage. In the end the branches were always observed to merge into one unbranched bundle. This annealing process might be prevented due to slow kinetics if branched bundles were to fuse together in a network. Under these dilute conditions, such behavior was not seen. It may not be the case if the bundle can interact with itself across the periodic boundary, which is carefully avoided in this particular simulation in order to observe the equilibrium microphase. Limited by the computational capacity, we have not explored extensively in the high density regime, which might produce long-lived connected bundles due to slow kinetics.

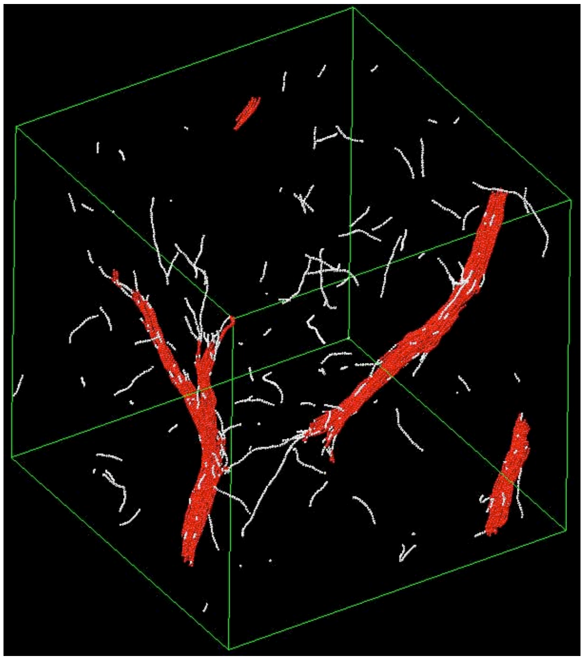


Figure 4.4: An intermediate branched bundle with monomers with attractions highlighted in red. The volume fraction $\phi = 0.123\%$, $l_p = 50$, $K_{\mathrm{assoc}} = 640000$, and $\exp(-\beta\epsilon) = 1.4$.

Figure 4.5 shows the relationship between the attraction strength and the volume fraction at the transition boundary. The simulation box size is 160 σ. The transition attraction strength decreases and finally levels off with increasing volume fraction. For very weak attractions, the transition density is so high that bundling is superseded by bulk nematic ordering, which was not studied in this work. Also plotted on the figure are the theoretical predictions for the transition boundary. If sphere chains are approximated by perfectly rigid rods and the rod length is set to be the mean chain length of the polydisperse system, based on an Onsager-like theory [24], the leading terms in the second virial coefficient can be written as [76, 79]

$$w = \frac{1}{2}\pi l^2 d + 2\pi l d^2 + \frac{1}{2}\pi l^2 \delta[1 - e^{-4\beta\epsilon}] + 4\pi l d \delta \left[1 + (e^{-\frac{2l}{d}\beta\epsilon} - 1)\frac{d}{2l\epsilon}\right], \qquad (4.5)$$

where l is the rod length and d is the rod diameter (equal to σ in this study). The rod length is equal to the mean chain length $\langle L \rangle = 0.5 + 0.5\sqrt{1 + 4\rho K_{\text{assoc}} \exp(k_I \rho)}$, where ρ is the monomer number density and the end effects parameter k_I is obtained from fitting simulation results [78]. The first two terms of w are the hard-core repulsion. The remaining terms are introduced by the short-range attractions. Approximately, two chains crossing at large angles have four pairs of attractions, and two parallel chains have $2l\epsilon$. At the transition point, the repulsion and attraction contributions cancel off, that is, $w_c = 0$, which is then solved numerically for the critical $\exp(-\beta\epsilon)$. In Figure 4.5, the transition boundary for the stable isotropic phase is not shown due to the issue of consistency. With increasing volume fraction, it becomes increasingly difficult to pinpoint the transition points for the condensation process, which is due in large part to the computational expense incurred by the geometrically increased number of monomers. To some extent, the general trend is captured by the current crude approximation. A more precise theoretical account should include the effects of the chain flexibility, the polydispersity, etc.

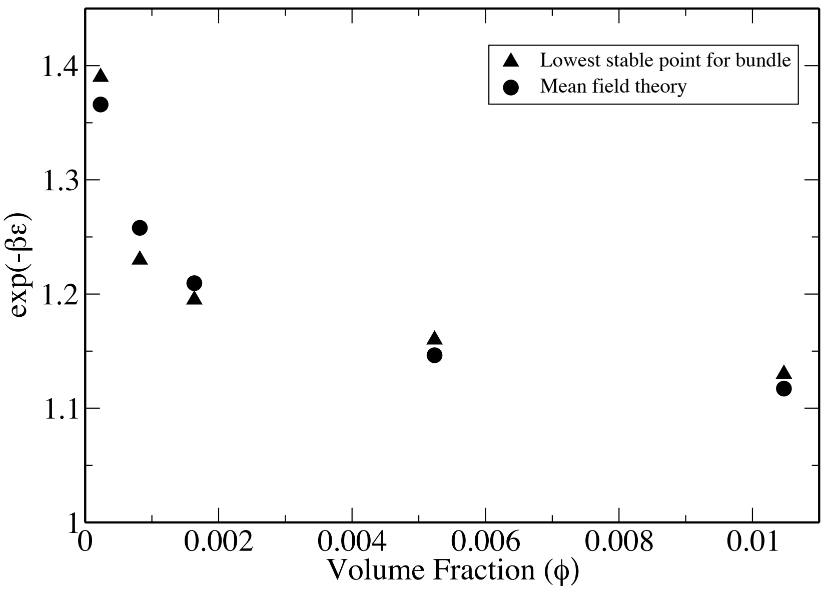


Figure 4.5: The inter-chain attraction strength versus the volume fraction at the transition boundary of microphase separation. $K_{\text{assoc}} = 6400$, $l_p = 1000\ \sigma$. The solid upper triangles are the lower boundary in simulation. The dots are critical values for condensation obtained from a theoretical approximation.

4.3.2 Chain length distributions

Even with the inter-chain attractions, the chain length distribution in the isotropic phase is still an simple exponential distribution and almost identical to the simulation results without inter-chain attractions in Ref. [63]. Only with strong enough attractions, the chain growth coupled with the attraction energy gain from clustering gives rise to more long chains. Especially in bundling, the distribution develops a peak around the magnitude of the bundle contour length. This phenomenon is in partial agreement with the simulation work [81] (because their results are completely in the intermediate regime, we did not make further comparisons) and the experimental results in Ref. [76], where the authors measured the coefficient of variance of the length distribution, $r_\sigma \equiv \sigma_l / \langle l \rangle$, which is the standard deviation divided by the mean chain length. As a result of actin filaments bundling, the authors argued, the value of r_σ changes from $r_\sigma = 1$ of an exponential distribution to $r_\sigma < 1$ of a Gaussian-like distribution; while in the intermediate regime, the length distribution is bimodal and $r_\sigma > 1$. In our simulation, we have observed a dramatic increase of r_σ because the length distribution is divided into two parts: the isotropic phase part and the cluster part, but not the reduction of r_σ to be less than 1. This discrepancy arises from factors: first, our NVT ensemble simulation result is unstable with respect to the macroscopic closely packed phase, which will be discussed in the section of μVT simulation; second, as the authors discussed, short filaments cannot be measured experimentally so that this part of contribution is excluded. As for the numerical prediction shown in Ref. [76], we were not able to reproduce the same length distribution, but have applied the same theory to predict the condensation boundaries in the previous section.

Figure 4.6 shows the chain length distribution in a well equilibrated bundled system. Since the system size is reasonably small, the simulation can efficiently explore the configurational space and achieve a smooth histogram curve. We can *roughly* assign the straight line with short chains to the isotropic solution, and the Gaussian-like peak to the bundle. (In fact, the two parts overlap considerably in the valley region.) We can fit the peak to a Gaussian function

$$f(l) = c \exp\left[\frac{(x-L_b)^2}{2b^2}\right], \tag{4.6}$$

where the first term describes the isotropic chains and the second term for the bundle. The method of defining and analyzing the bundle will be presented in section 4.3.3. To elucidate the question whether the Gaussian shape of the peak results primarily from the fluctuation of bundle itself or the chain length polydispersity within the bundle, we computed the standard deviation of chains within each bundle, 4.8 which was averaged over 101 snapshots. This value is significantly ($\sim 40\%$) less than $b = 7.8$, the standard deviation of the full length distribution of the bundle. This suggests that the distribution's Gaussian-shape can be attributed to in the bundle fluctuation, which has a standard deviation of 4.1 in this case. The bundle fluctuation is related to the fluctuation of the total monomers within the bundle, which stems from the equilibrium between the bundle and the isotropic solution.

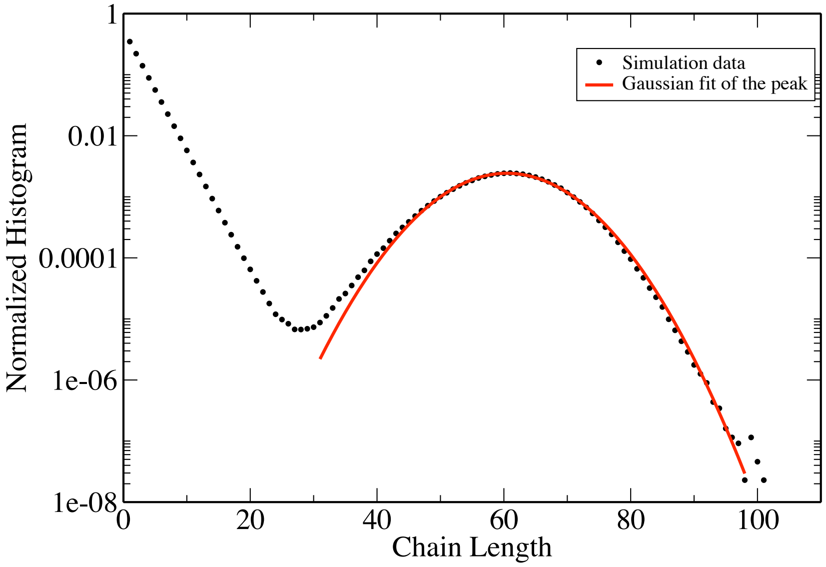


Figure 4.6: Semi-logarithmic plot of the chain length distribution. The simulation box size is 80 σ, $\phi = 0.082\%$. $l_p = 1000\ \sigma$, $K_{\text{assoc}} = 6400$, and $\exp(-\beta\epsilon) = 1.30$. The red curve is a Gaussian fit in the length range of $[31, 98]$: the mean is 60.5 and the standard deviation is 7.88.

Figure 4.7 shows the chain length distribution for a system with a thick and long bundle with $l_p = 10000$, where chains (much shorter than l_p) are so stiff that the bundle consists of almost "perfectly" aligned polymer chains rather than possible interwoven chains as in the case of small l_p values. For such a system, the distribution is further divided into two far separated parts: the isotropic chains and the bundle. The peak for the bundle becomes no longer symmetric (or Gaussian) as in the small bundle case. Here we propose one hypothesis that surface and bulk chains in the bundle become more and more distinct as the bundle grows. The surface chains are attracted to fewer surrounding chains than the bulk chains, thus the two types of chains have different environments and follow different growth laws. If the bundle is approximated by a cylinder of length L_b, then the surface chains' distribution (assuming $l < L_b$) is proportional to $(L_b - l)/l \exp(al)$ with a fitting parameter a. For an ideal system $a = \ln[\rho(1) K_{\text{assoc}} \exp(-\beta\epsilon\langle n_{\text{attr}}\rangle)]$, where $\rho(1)$ is the number density of free monomers and $\langle n_{\text{attr}}\rangle$ is the average number of attractions per surface chain monomer. In the formula $(L_b - l)$ and $1/l$ account for the surface chain's translational and orientational entropy's dependence on the chain length, respectively. On the other hand, the bulk chains' distribution is a Gaussian-like distribution. Therefore, the distribution function used for fitting can be written as

$$f(l) = c_1 \frac{1}{2} \left[(L_b - l) + |L_b - l|\right] \frac{1}{l} \exp(al) + c_2 \exp[-(l - L_b)^2/(2b^2)], \qquad (4.7)$$

where c_1 and c_2 are coefficients for the two components, containing some physical constants such as π. Here we optimize the fit to $\ln f(l)$. The fitting results are: $c_1 = 2.76 \times 10^{-11}$, $c_2 = 2.62 \times 10^{-4}$, a=0.10021, $b = 11.21$, and $L_b = 203$. With $\langle\rho(1)\rangle = 1.05 \times 10^{-4}, K_{\text{assoc}} = 6000$ and $\exp(-\beta\epsilon)$, we obtain the average number of attractions per surface monomer $\langle n_{\text{attr}}\rangle$ is 3.08. As a comparison, for the whole bundle (including both bulk and surface monomers) $\langle n_{\text{attr}}\rangle \sim 7$ with the reasonable assumption that dilute isotropic chains have no attractions. The values of $\langle n_{\text{attr}}\rangle$ can be related to how chains pack in the bundle, which will be further discussed in the section of bundle structure analysis. The irregularity of the fitting curve near the peak can be attributed to the assumption of a fixed cutoff L_b while L_b itself fluctuates around the peak value.

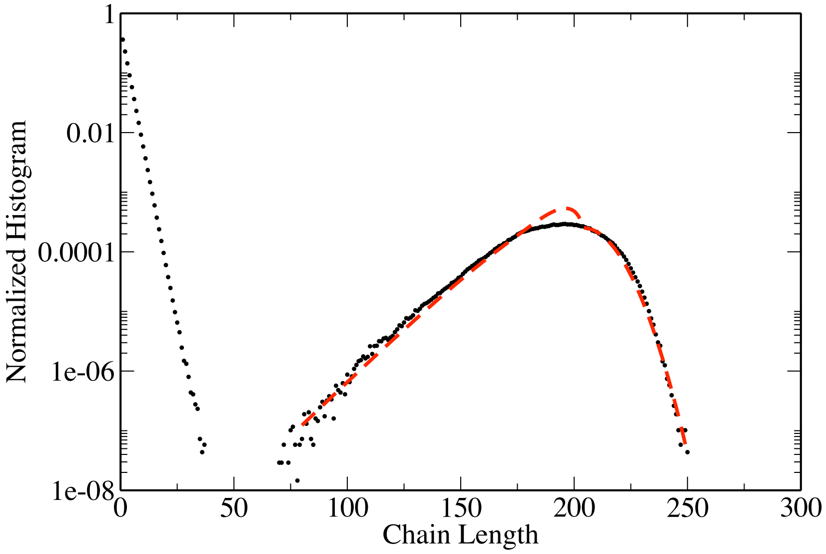


Figure 4.7: Semi-logarithmic plot of the chain length distribution. The simulation box size is 160 σ, $\phi = 0.082\%$. $l_p = 10000\ \sigma$, $K_{\text{assoc}} = 6000$, and $\exp(-\beta\epsilon) = 1.20$. The red curve is a fit of the chain length distribution of the bundle.

In order to explain the difference for the surface chain behavior between small and big bundles, we neglect the internal conformational entropy of a chain of length l. The free energy of such a (straight) chain in the dilute surrounding space can be written as

$$\beta F_{\text{iso}} = -\ln(V - V_{\text{b}}) - E_{\text{bond}}(l-1) \approx -\ln V - E_{\text{bond}}(l-1), \tag{4.8}$$

where the bundle volume V_{b} is tiny compared to V. The orientational entropy is zero from the definition

$$S_{\mathrm{or}} = -k_{\mathrm{B}} \int f(\Omega) \ln[4\pi f(\Omega)] \mathrm{d}\Omega, \tag{4.9}$$

because $f(\Omega) = 1/4\pi$ for the isotropic chains. While the chain resides on the surface of a bundle, the free energy will be

$$\beta F_{\mathrm{bundle}} = -\ln[\pi D(L_b - l)] - \ln\left(\pi \frac{D}{l}\right) + \epsilon l - E_{\mathrm{bond}}(l-1). \tag{4.10}$$

Here the first term accounts for the translational entropy of a chain of length l on the surface of a cylinder of width D and length L_b. Here we neglect the contribution of the surface chains longer than L_b because of too high energy penalty. In the Section 4.3.3 we show that surface chains are indeed shorter than bulk chains. The second term is the orientational entropy if we assume the surface chain can freely rotate within $\theta \approx D/l$ without any energy penalty. (This is a very crude approximation since the actual angle can be much smaller than D/l. It should suffice here since the contribution of this term is both constant and negligible.) The third term is the energy gain from the inter-chain attractions and the last term the bond energy. For a small value of L_b, F_{bundle} is larger than the F_{iso} for all $l < L_b$, which means that surface chains are not stable on a small bundle. Only when the L_b is large enough, the curve of F_{bundle} as a function of l intersects that of F_{iso}, which means that a long chain becomes more stable on a big bundle than on a small bundle. We have plotted the relative free energy $F' = F_{\mathrm{bundle}} - F_{\mathrm{iso}}$ as a function of the chain length l for given bundle contour lengths L_b in Fig. 4.8. Therefore, the answer to the question why the length distribution of a small bundle system is less affected by surface chains than that of a big bundle is that stable surface chains requires the bundle size to be above a critical value. For example, the critical bundle contour length is approximately 110 σ while $D = 3\ \sigma$ is assumed. This simple analysis also explains the crossover length between the two length distribution parts, namely, the bundle and the isotropic solution. Given $D = 3\ \sigma$ and $L_b = 200\ \sigma$, we have the crossover length close to 80 σ, which is in agreement with the length distribution as shown in Fig. 4.7. Noticeably, the upswing trend of the free energy curve when the chain length approaches the bundle contour length indicates that for chains of length comparable with L_b, we can not neglect the translational entropy resulting from partial overlapping between surface chains and the bundle. To further understand the mechanism of actin bundling, we may need further simulation work on the nucleation and growth steps. We can a cylinder composed of straight chains as a growth front similar to the model by Liu et al. [84], then allow self-assembling chains to attach onto or detach from the cylindrical growth front.

4.3.3 Analysis of the bundles

The bundle is determined by a connectivity matrix **P** defined by [81]

$$P_{ij} = \begin{cases} 1 & i = j \\ 1 & \text{monomers i and j are either attracted or bonded} \\ 0 & \text{otherwise,} \end{cases} \tag{4.11}$$

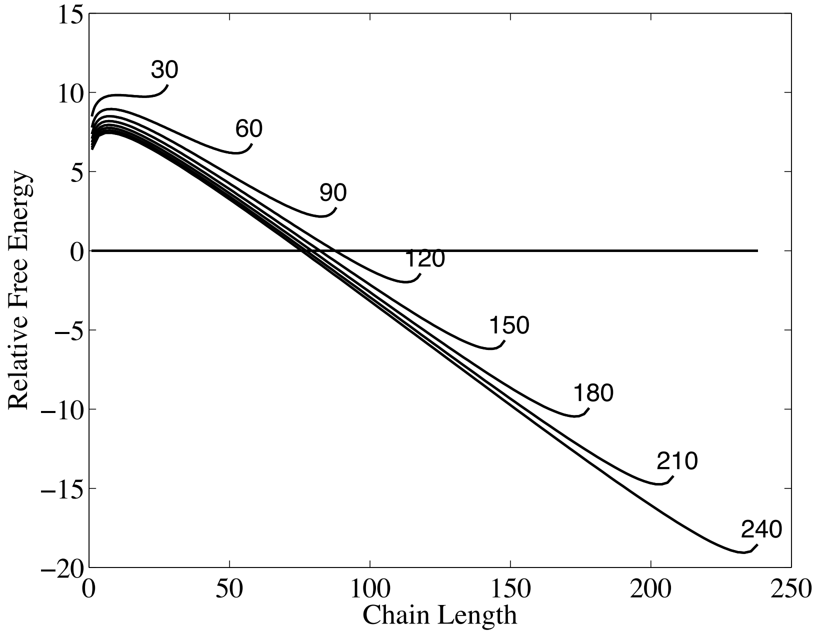


Figure 4.8: Relative free energy ($F' = F_{\text{bundle}} - F_{\text{iso}}$) as a function of the chain length l and the bundle contour length L_b. The number labels are L_b.

In principle, $\mathbf{P}^{N-1}$ generates the global connectivity matrix $\mathbf{G}$ where

$$G_{ij} = \begin{cases} \neq 0 & \text{at least one path exists between i and j} \\ \\ = 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

But this matrix exponentiation is computationally expensive. In practice, we search the matrix $\mathbf{P}$ recursively and find the largest number of connected monomers, then record their coordinates and other relevant information at regular intervals.

The bundle takes a cylindrical shape if the persistence length is longer than the bundle contour length, as in the case $l_p = 1000\ \sigma$. The ends of the bundle are not flat like a true cylinder, but rough. As shown in 4.3.2, the chain length distribution in the bundle is highly peaked around the bundle contour length. In Fig. 4.9 we count the number of chains through the cross-section area along the contour the bundle. The plot shows that the number of chains in the buddle's middle part is almost constant while the two ends taper off. The difference between red triangles and black stars arises from different procedures, namely decremental and incremental, to obtain the data as explained in the MODEL section. The extent of smoothness in the middle suggests where component chains are continuous or broken. Generally, The smoother the distribution, the better equilibrium the system achieves. This is also true for the smoothness of the chain length distribution. The nonequilibrium problem will be further discussed later.

Within the bundle, we can construct a discrete central axis by slicing the bundle along its contour and then computing the center of mass of monomers in each resulting

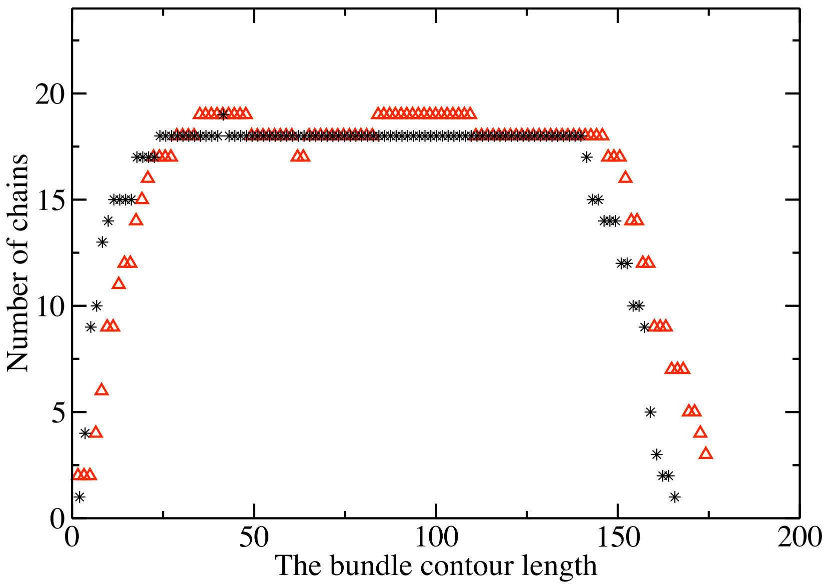


Figure 4.9: The number of chains through the cross-section versus the bundle contour length. which is averaged over 101 samples. The simulation system is a 160 × 160 × 160 box with 6400 monomers. Both colors are for $l_p = 1000\ \sigma, K_{\text{assoc}} = 64000$, and $\exp(-\beta\epsilon) = 1.17$. Red triangles are results from a decremental procedure; black stars are from an incremental procedure.

bin. In Fig. 4.10 we plotted the mean monomer displacement from the center axis against the length of the chain which the monomers belong to. We can observe the trend that the displacement decreases with increasing chain length, which implies that surface chains are shorter than bulk chains, consistent with the model of section 4.3.2. Both small and big bundles show this feature, although the latter is more profound. We also compare two bundles of similar size with different persistent lengths, 1000 σ and 10000 σ, respectively. More flexible chains in a bundle tend to be somewhat interwoven such that the descending slope is not as steep as that of stiffer chains. In both cases, very short chains (roughly $l < 25\ \sigma$) have significantly greater displacements than long chains do. These irregular points suggest that the bundle defined in this work can include short chains occasionally. One possible scenario is that a relatively short chain is in contact with and perpendicular to the bundle. This phenomenon is also confirmed by both the length distribution analysis of the bundle and the intermediate snapshots.

Chains are packed hexagonally within a bundle. Technically, the local order of packing can be estimated by slicing the bundle using the method described in previous paragraphs. For simplicity, it will suffice to project the whole bundle of monomers onto a plane perpendicular to the director of the bundle. Similar to the nematic phase, the director is determined through solving for the order parameter of the entire bundle as defined by

$$S = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{3}{2} \cos^2 \theta_i - \frac{1}{2} \right), \tag{4.13}$$

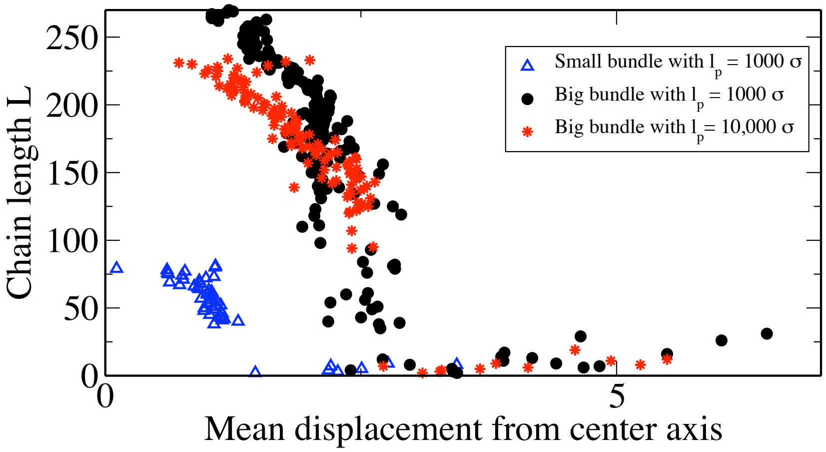


Figure 4.10: The distribution of chains as a function of the mean displacement from the central axis of the bundle, which is averaged over 101 bundle samples. Black dots are $l_p = 1000\ \sigma$, $K_{\text{assoc}} = 64000$ and $\exp(-\beta\epsilon) = 1.16$. Blue upper triangles and red stars are the same systems as in Fig. 4.6 and Fig. 4.7, respectively. The volume fraction is 0.082% for all three data sets, but the simulation box is $(80\ \sigma)^3$ for the small bundle instead of $(160\ \sigma)^3$ for others.

where θ_i is angle between the ith bond and the director. All monomers are projected onto a new $x-y$ plane with one arbitrary monomer as the origin as shown in Fig. 4.11 and the director as z axis. For perfectly packed straight chains, it should show only N (the number of chains) data points because all monomers on a same chain overlap. While in our simulation, even with a persistence length as long as 10000 σ, a projected chain traces a irregular curve on the $x-y$ plane. An approximate method is to obtain the average position of each chain. The hexagonal order parameter, S_{hex}, is defined by [30]

$$S_{\text{hex}} = \frac{2}{N(N-1)} \sum_{i}^{N} \sum_{j>i}^{N} \cos(\mathbf{g} \cdot \mathbf{r}_{ij}), \tag{4.14}$$

where $\mathbf{r}_{ij}$ is a vector from i to j in $x-y$ plane, and $\mathbf{g}$ is an arbitrary reciprocal lattice vector. In practice, we simply use one of the reciprocal basis vectors $\mathbf{b}_i$. For a perfect hexagonal lattice (i.e., vertices with periodic boundary conditions), the value of S_{hex} is 1 (independent of the choice of $\mathbf{g}$). The magnitude of the basis vector, from which $\mathbf{b}_i$ is obtained, is approximated by the average nearest pair distance for N data points. The directions of the basis vectors were determined arbitrarily from one of the closet pair points as shown in Fig. 4.11. One way to refine the current S_{hex} estimation is to obtain basis vectors through maximizing S_{hex} instead of choosing basis vectors *a priori*. Interestingly, figure 4.11 also shows the twist pattern of the whole bundle, where all chains in the bundle winds roughly in the same direction (counterclockwise in the figure).

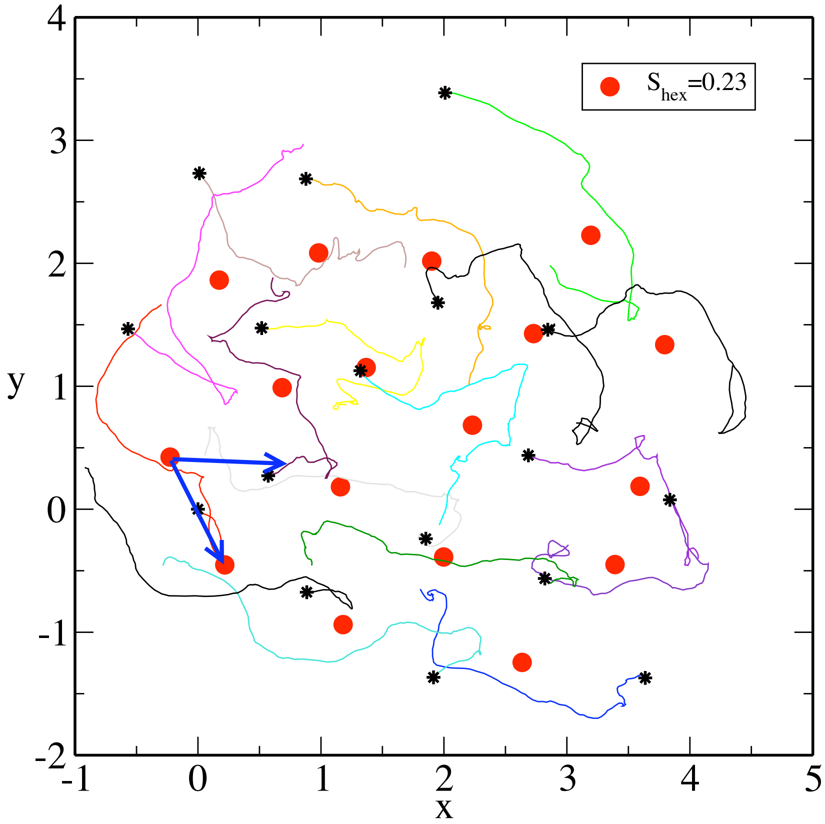


Figure 4.11: Estimation of S_{hex} by projecting chain monomers onto the $x-y$ plane perpendicular to the nematic director (positive z) of the bundle chains. The bundle sample is extracted from a simulation snapshot with $K_{\text{assoc}} = 4000$ and $l_p = 10000\ \sigma$. Each curve of different color represents the projection of a different chain. Stars represent the chains' ends on one side of the bundle. Each red dot is the mean position of one curve. The two blue vectors are the arbitrarily chosen basis vectors of the lattice. The estimated S_{hex} of this snapshot is 0.23.

The hexagonal structure of bundle chains further explains why not surprisingly, the attraction number per surface monomer $\langle n_{\text{attr}} \rangle$ is very close to 3, which is straightforward if a monomer is on the side of a hexagonal lattice. This comparison confirms that the fitting result of the proposed length distribution formula is reasonable and agrees well with the bundle structure. On the other hand, in our simulations, a monomer is allowed to be attracted to two consecutive monomers of another chain due to the range constraint on the short-attraction, which is the reason that the estimated $\langle n_{\text{attr}} \rangle \sim 7$ for all monomers in a bundle is greater than 6, which is straightforward for a perfect bulk hexagonal lattice.

4.3.4 Aspect ratio of the bundle

Figure 4.12 shows the relationship between the percentage of monomers in the bundle and the inter-chain attraction strength. The bundle size (N_b) increases very fast with increasing $\exp(-\beta\epsilon)$ above the transition point, but then gradually levels off when the bundling approaches a degree as high as $70 \sim 80\%$. At equilibrium, the monomer chemical potential of the bundle surface is equal to that of the surrounding isotropic solution, which logarithmically approaches negative infinity with decreasing monomer concentration as we can ignore the volume of the bundle. This implies that the equilibrium $\exp(-\beta\epsilon)$ also exponentially increases with the decreasing monomer concentration in the solution, i.e., increasing percentage of bundled monomers N_b/N.

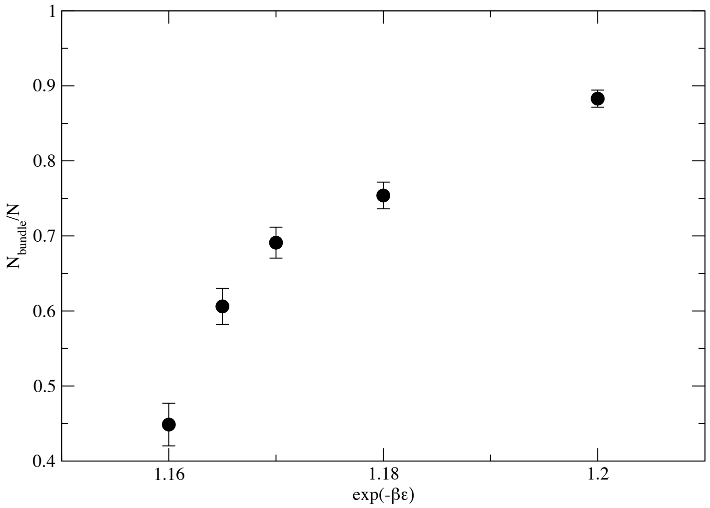


Figure 4.12: The simulation box is $(160\ \sigma)^3$ with 6400 monomers ($\phi = 0.082\%$). $l_p = 1000\ \sigma$ and $K_{\text{assoc}} = 64000$.

For a fixed volume fraction, we also study an interesting property of the bundle: aspect ratio, an important quantity which can be directly measured in experiments [77]. If the bundle is considered as an ideal cylinder, the free energy can be expressed as

$$F = \gamma S + g\frac{V}{L_b}, \tag{4.15}$$

where γ is the surface energy per unit area, S is the surface area of the actin bundle, g is the end-cap energy per cross-section area, and V is the bundle volume. γ and g correspond to the microscopic parameters $-\epsilon$ and $K_B T \ln K_{\text{assoc}}$, respectively. The cross-section area is given by V/L_b. If the diameter of the cylinder is denoted as D, then we have

$$S = \pi(D^2/2 + DL_b); \qquad V = \pi D^2 L_b/4. \tag{4.16}$$

Then we can obtain the optimal aspect ratio (L_b/D) through free energy minimization with respect to either L_b or D:

$$L_b/D = 1 + \frac{g}{2\gamma}. \tag{4.17}$$

This formula suggests that the stronger the inter-chain attraction is, the thicker the bundle is; while the association constant mainly controls the bundle length. However, this simple result does not agree with our NVT ensemble simulation results, where we do not have ideally packed straight chains, neither guaranteed is the equilibration of the bundled system. The combination of bundle aggregation and geometrical optimization complicates the relationship between the aspect ratio and physical parameters as shown in Fig. 4.13, where we plot the aspect ratios and their corresponding error bars for the procedure of decreasing $\exp(-\beta\epsilon)$ progressively. Above the transition boundary, the aspect ratio initially increases with increasing attraction, then decreases with further stronger attraction. It suggests that in the early stage the bundle aggregation prefers the longitudinal growth. When the aggregation levels off, the axial growth of actin bundles begin to win out in the competition with the longitudinal growth. One remaining problem here is how to improve the efficiency of Monte Carlo sampling, which may play a critical role in this interesting phenomenon. For example, in the incremental procedure, it is extremely slow for a closely packed bundle to relax back to a shorter one (in terms of L/D), although the results for relatively small $\exp(-\beta\epsilon)$'s (≤ 1.18) are identical to that of the decremental procedure.

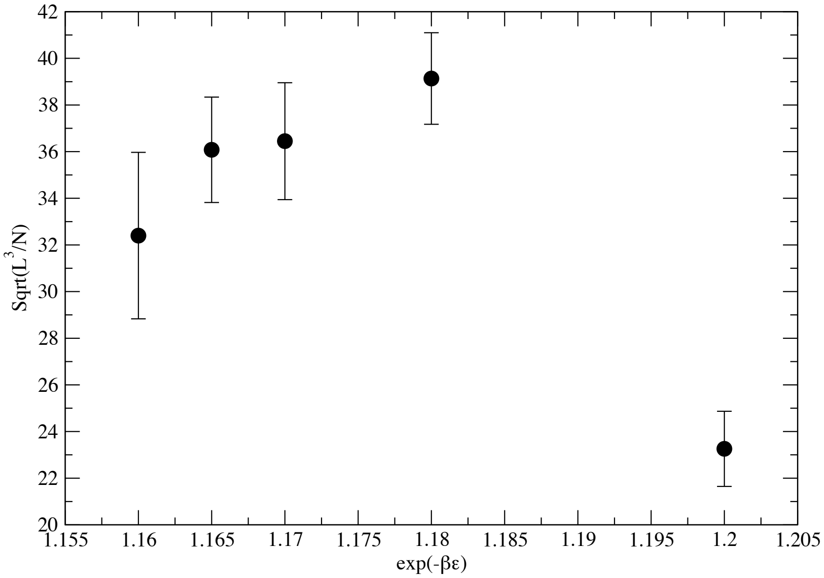


Figure 4.13: Aspect ratio varies with the attraction strength. The simulation box is $(160\ \sigma)^3$ with 6400 monomers ($\phi = 0.082\%$). $l_p = 1000\ \sigma$ and $K_{\text{assoc}} = 64000$. The error bars indicate standard deviation.

4.3.5 Convergence issues

For practical purposes, we use two quantitative criteria to follow the convergence of a system to equilibrium. A parameter for fast examination is the average number of isotropic attractions per monomer $\langle n_{\text{iso}} \rangle$. In an isotropic phase, $\langle n_{\text{iso}} \rangle$ is almost zero at low densities as in our study; when the bundle microphase emerges, $\langle n_{\text{iso}} \rangle$ increases dramatically to be greater than one. When $\langle n_{\text{iso}} \rangle$ is stable, we need to examine the length distribution carefully. An equilibrated bundle system always has one single peak around the magnitude of the bundle contour length in the chain length distribution. Multiple peaks can be observed in the intermediate stage.

The dissolution of a bundle depends sensitively on whether the system is well equilibrated. Figure 4.14 shows that a not-well-equilibrated bundle quickly disappears while a well-equilibrated bundle lasts significantly longer. Starting from the same initial configuration at $\exp(-\beta\epsilon) = 1.30$, the system transitions to the isotropic phase within 2×10^7 steps if $\exp(-\beta\epsilon)$ is reduced to 1.25. On the other hand, if the attraction strength is reduced by decrements of 0.01 and each reduction is followed by 3×10^7 Monte Carlo steps, the system continues to stay stable at the same ending point of $\exp(-\beta\epsilon) = 1.25$. For which, the main reason is that the procedure of decrementally reducing $\exp(-\beta\epsilon)$ allows both the bundle and chains in the isotropic solution to relax into a more stable state than the $\exp(-\beta\epsilon)$ quenching method.

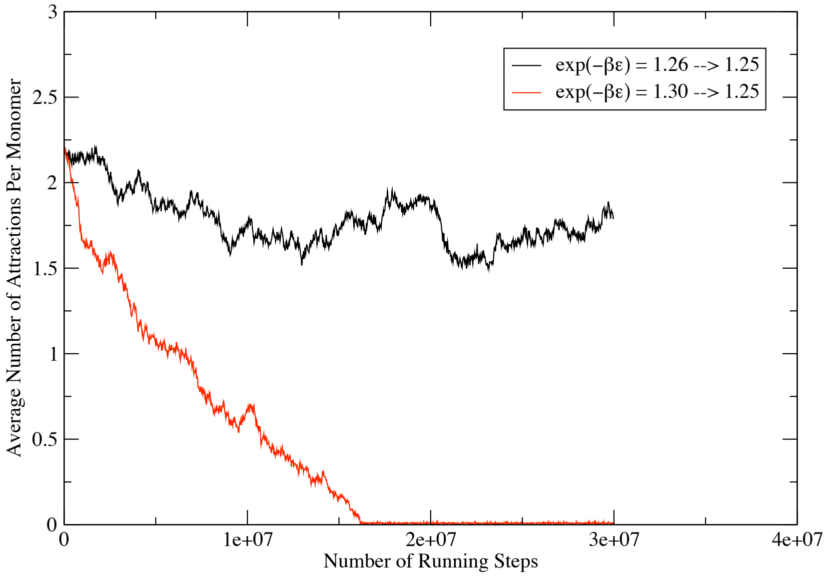


Figure 4.14: Comparison of the stability of the same initial bundle responding to different reduction rates of $\exp(-\beta\epsilon)$. Red curve: A not-yet-equilibrated system at $\exp(-\beta\epsilon) = 1.30$ transitions to an isotropic phase quickly after $\exp(-\beta\epsilon)$ is quenched to 1.25. Black curve: The bundle is still stable at $\exp(-\beta\epsilon) = 1.25$. The starting configuration of $\exp(-\beta\epsilon) = 1.26$ is obtained by decrements of 0.01 from the $\exp(-\beta\epsilon) = 1.30$ configuration. Note that the shown black curve is the fifth segment, each equals 3×10^7 Monte Carlo steps.

4.3.6 The effect of the simulation box size

The system undergoes a complete microphase separation in our simulation. The simulation box size has strong influence on the simulation results. Intuitively, the simulation box must be spacious enough to accommodate the formed bundle, otherwise a straight enough bundle can easily interact with itself across the box boundary. Different simulation box sizes with the same volume fraction result in different transition boundaries because the microphase size increases with increasing number of monomer as the volume of the system is increased. We tested two typical cubic side lengths 80 σ and 160 σ with the volume fraction $\phi = 0.082\%$, for which the transition value of $\exp(-\beta\epsilon)$ for a stable bundle to fall apart is 1.29 and 1.23, respectively. Approximately, the number of monomers in the bundle is octupled as expected; while the aspect ratios of the bundle differ significantly, which may be attributed to the non-equilibrium issue as discussed in the previous two sections.

4.3.7 Comparison with grand canonical Monte Carlo simulations

We have also performed Monte Carlo simulations of the same system using grand canonical ensemble (μVT) [63, 85], where we fix the chemical potential μ instead of the total number of monomers N. Figure 4.15 shows a snapshot of such a typical GCMC simulation. The result shows that the the bundles are thermodynamically unstable with respect to the bulk phase separation, where chains form a macroscopic dense aligned phase. The comparative simulation is carried out as follows. In practice, we set a maximum number of monomers, for example, 20000 in a simulation box $(40\ \sigma)^3$. Then, first we choose a value of $\exp(-\beta\epsilon)$ as small as 1.00 to achieve an isotropic phase with a density roughly equal to the NVT simulation parameter. Second, we increase the attraction strength to observe the transition value of $\exp(-\beta\epsilon)$. To observe the phenomenon of closed packed simulation box, we tested 20000 monomers in three boxes of 160, 80, and 40 within our computational capability. Due to the fact that the chain length greatly exceeds the box size, the self-interaction of chains becomes profound with small simulation boxes. Therefore, the transition $\exp(-\beta\epsilon)$ decreases with increasing simulation box size. Overall, for the transition point, the μVT simulation results agree well with the NVT simulation results obtained from the increasing procedure as shown in Fig. 4.1.

4.4 Conclusion

We have performed comprehensive canonical Monte Carlo simulations of self-assembled linear polymers with inter-chain attractions. We observed that the phase behavior considerably depends on the volume fraction and association constant. The polymer chain flexibility plays a role in the microphase morphology, which is determined by the relative length scale of the persistence length and the mean chain length.

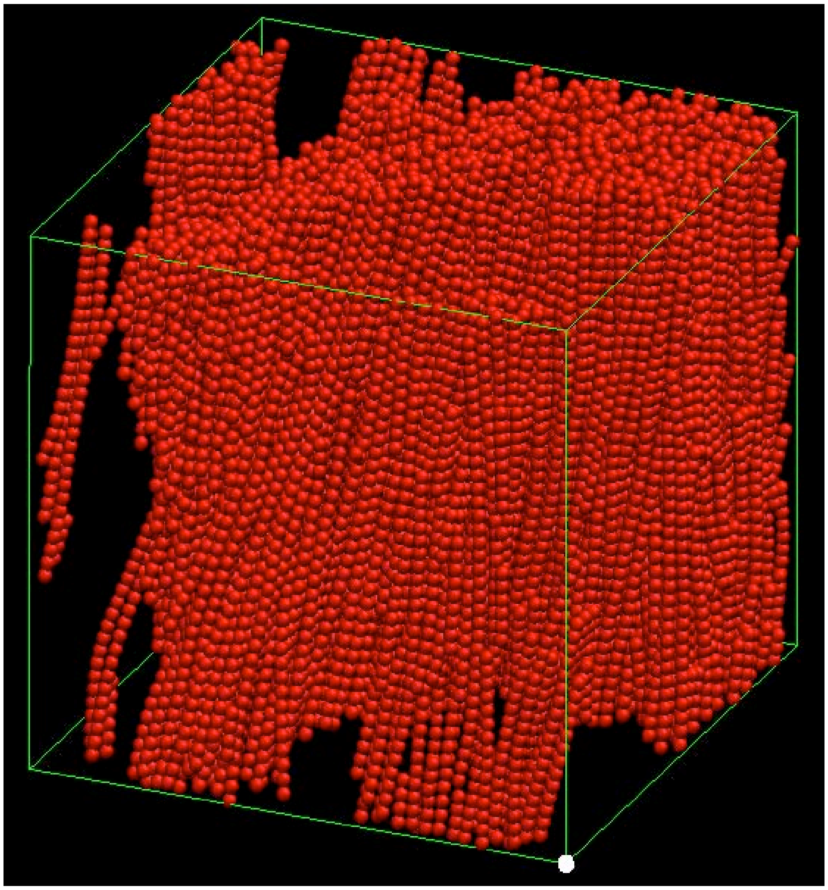


Figure 4.15: A snapshot of a highly dense system. The simulation box size is 40 σ, $l_p = 1000\,\sigma$, $K_{\text{assoc}} = 6400$, $\exp(-\beta\epsilon) = 1.53$, $\mu = 1.14\times10^{-4}$. The volume fraction is 16.36% (a maximum imposed in the simulation code). Here all monomers are in red whether they have attractions or not.

We proposed a simple statistical thermodynamical model to interpret the chain length distribution of a bundled system. The bundle aspect ratio is not monotonically proportional to the inter-chain attractions, but it is still unclear how much this phenomenon is affected by the limited computational capacity. The grand canonical Monte Carlo simulations have been performed to show that the microphase bundles are thermodynamically unstable with respect to the macroscopic condensed bulk phase of aligned chains.

Bibliography

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 1996).

[2] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[3] A. F. Bielajew, *Fundamentals of the Monte Carllo method for neutral and charged particle transport* (2001).

[4] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, 2000).

[5] J. I. Siepmann and D. Frenkel, Mol. Phys. **75**, 59 (1992).

[6] M. N. Rosenbluth and A. W. Rosenbluth, J. Chem. Phys. **23**, 356 (1955).

[7] P. H. Verdier and W. H. Stockmayer, J. Chem. Phys. **36**, 227 (1962).

[8] F. T. Wall and F. Mandel, J. Chem. Phys. **63**, 4592 (1975).

[9] O. F. Olaj, T. Petrik, and G. Zifferer, J. Chem. Soc. Faraday Trans. **91**, 2551 (1995).

[10] D. H. Boal, *Mechanics of the Cell* (Cambridge University Press, New York, 2002).

[11] J. A. Cooper, Annu. Rev. Physiol. **53**, 585 (1991).

[12] P. A. Janmey, J. X. Tang, and C. F. Schmidt (1999), previously published in the Biophysics Textbook Online (BTOL).

[13] J. X. Tang and P. A. Janmey, Biol. Bull. **194**, 406 (1998).

[14] A. Cēbers, Z. Dogic, and P. A. Janmey, Phys. Rev. Lett. **96**, 247801 (2006).

[15] J. X. Tang, J. A. Käs, J. V. Shah, and P. A. Janmey, Eur. Biophys. J. **30**, 477 (2001).

[16] P. S. Niranjan, J. G. Forbes, S. C. Greer, J. Dudowicz, K. F. Freed, and J. F. Douglas, J. Chem. Phys. **114**, 10573 (2001).

[17] P. S. Niranjan, P. B. Yim, J. G. Forbes, S. C. Greer, J. Dudowicz, K. F. Freed, and J. F. Douglas, J. Chem. Phys. **119**, 4070 (2003).

[18] S. C. Greer, J. Phys. Chem. B **102**, 5413 (1998).

[19] S. C. Greer, Annu. Rev. Phys. Chem. **53**, 173 (2002).

[20] D. Sept, J. Y. Xu, T. D. Pollard, and J. A. McCammon, Biophys. J. **77**, 2911 (1999).

[21] D. Sept and J. A. McCammon, Biophys. J. **81**, 667 (2001).

[22] J. Viamontes and J. X. Tang, Phys. Rev. E **67**, 040701 (2003).

[23] J. Viamontes, P. W. Oakes, and J. X. Tang, Phys. Rev. Lett. **97**, 118103 (2006).

[24] L. Onsager, Ann. N.Y. Acad. Sci. **57**, 627 (1949).

[25] T. Odijk, Macromolecules **19**, 2313 (1986).

[26] G. J. Vroege and H. N. W. Lekkerkerker, Rep. Prog. Phys. **55**, 1241 (1992).

[27] P. van der Schoot and M. E. Cates, Europhys. Lett. **25**, 515 (1994).

[28] Z. Y. Chen, Macromolecules **26**, 3419 (1993).

[29] A. Milchev and D. P. Landau, Phys. Rev. E **52**, 6431 (1995).

[30] B. Fodi and R. Hentschke, J. Chem. Phys. **112**, 6917 (2000).

[31] A. Chatterji and R. Pandit, Europhys. Lett. **54**, 213 (2001).

[32] C. G. dos Remedios, D. Chhabra, M. Kekic, I. V. Dedova, M. Tsubakihara, D. A. Berry, and N. J. Nosworthy, Physiol. Rev. **83**, 433 (2003).

[33] M. Hase and K. Yoshikawa, J. Chem. Phys. **124**, 104903 (2006).

[34] H. Hoffmann, G. Oetter, and B. Schwandner, Prog. Colloid Polym. Sci. **73**, 95 (1987).

[35] T. Oda, K. Makino, I. Yamashita, K. Namba, and Y. Maéda, Biophys. J. **75**, 2672 (1998).

[36] P. J. Flory, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **234**, 60 (1956).

[37] A. Y. Grosberg and A. R. Khokhlov, Adv. Polym. Sci. **41**, 53 (1981).

[38] A. N. Semenov and A. R. Khokhlov, Sov. Phys. Usp. **31**, 988 (1988).

[39] M. Dijkstra and D. Frenkel, Physical Review E **51**, 5891 (1995).

[40] F. A. Escobedo and J. J. de Pablo, J. Chem. Phys. **106**, 9858 (1997).

[41] W. E. McMullen, W. M. Gelbart, and A. Benshaul, J. Chem. Phys. **82**, 5616 (1985).

[42] T. Odijk, J. Physique **48**, 125 (1987).

[43] E. M. Kramer and J. Herzfeld, Phys. Rev. E **58**, 5934 (1998).

[44] A. Matsuyama and T. Kato, J. Phys. Soc. Jpn. **67**, 204 (1998).

[45] J. T. Kindt and W. M. Gelbart, J. Chem. Phys. **114**, 1432 (2001).

[46] Y. Rouault, Eur. Phys. J. B **6**, 75 (1998).

[47] N. B. Wilding and P. Sollich, J. Chem. Phys. **116**, 7116 (2002).

[48] F. A. Escobedo, J. Chem. Phys. **118**, 10262 (2003).

[49] J. T. Kindt, J. Phys. Chem. B **106**, 8223 (2002).

[50] J. T. Kindt, in *Mesoscale Phenomena in Fluid Systems*, edited by F. Case and P. Alexandridis, American Chemical Society (Oxford University Press, 2003), vol. 861 of *ACS Symposium Series*, pp. 298–312.

[51] D. J. Tildesley and W. B. Streett, Mol. Phys. **41**, 85 (1980).

[52] R. Eppenga and D. Frenkel, Mol. Phys. **52**, 1303 (1984).

[53] B. O'Shaughnessy and D. Vavylonis, Phys. Rev. Lett. **90**, 118301 (2003).

[54] J. Dudowicz, K. F. Freed, and J. F. Douglas, J. Chem. Phys. **119**, 12645 (2003).

[55] J. P. Wittmer, A. Milchev, and M. E. Cates, J. Chem. Phys. **109**, 834 (1998).

[56] A. Milchev, J. P. Wittmer, and D. Landau, Phys. Rev. E **61**, 2959 (2000).

[57] J. P. Wittmer, P. van der Schoot, A. Milchev, and J. L. Barrat, J. Chem. Phys. **113**, 6992 (2000).

[58] W. M. Gelbart, A. Ben-Shaul, W. E. McMullen, and A. Masters, J. Phys. Chem. **88**, 861 (1984).

[59] A. R. Khokhlov and A. N. Semenov, Physica A **112**, 605 (1982).

[60] Q. L. Yan, R. Faller, and J. J. de Pablo, J. Chem. Phys. **116**, 8745 (2002).

[61] K. Ghosh and M. Muthukumar, Phys. Rev. Lett. **91**, 158303 (2003).

[62] A. Speranza and P. Sollich, Phys. Rev. E **67**, 061702 (2003).

[63] X. Lü and J. T. Kindt, J. Chem. Phys. **120**, 10328 (2004).

[64] M. P. Taylor and J. Herzfeld, J. Phys.: Condens. Matter **5**, 2651 (1993).

[65] F. C. MacKintosh, J. Käs, and P. A. Janmey, Phys. Rev. Lett. **75**, 4425 (1995).

[66] M. L. Gardel, J. H. Shin, F. C. MacKintosh, L. Mahadevan, P. Matsudaira, and D. A. Weitz, Science **304**, 1301 (2004).

[67] A. Suzuki, T. Maeda, and T. Ito, Biophys. J. **59**, 25 (1991).

[68] C. M. Coppin and P. C. Leavis, Biophys. J. **63**, 794 (1992).

[69] M. P. Allen, G. T. Evans, D. Frenkel, and B. M. Mulder, Adv. Chem. Phys. **86**, 1 (1993).

[70] F. Oosawa and S. Asakura, *Thermodynamics of the Polymerization of Protein* (Academic Press, New York, 1975).

[71] J. X. Tang and P. A. Janmey, J. Biol. Chem. **271**, 8556 (1996).

[72] O. Pelletier, E. Pokidysheva, L. S. Hirst, N. Bouxsein, Y. Li, and C. R. Safinya, Phys. Rev. Lett. **91**, 148102 (2003).

[73] V. A. Bloomfield, Curr. Opin. Struct. Biol. **6**, 334 (1996).

[74] V. A. Bloomfield, Biopolymers **44**, 269 (1997).

[75] D. J. Needleman, M. A. Ojeda-Lopez, U. Raviv, H. P. Miller, L. Wilson, and C. R. Safinya, Proc. Natl. Acad. Sci. **101**, 16099 (2004).

[76] D. Biron, E. Moses, I. Borukhov, and S. A. Safran, Europhys. Lett. **73**, 464 (2006).

[77] H. J. Kwon, Y. Tanaka, A. Kakugo, K. Shikinaka, H. Furukawa, Y. Osada, and J. P. Gong, Biochemistry **45**, 10313 (2006).

[78] X. Lü and J. T. Kindt, J. Chem. Phys. **125**, 054909 (2006).

[79] I. Borukhov, R. F. Bruinsma, W. M. Gelbart, and A. J. Liu, Proc. Natl. Acad. Sci. USA **102**, 3673 (2005).

[80] J. Kierfeld, T. Kühne, and R. Lipowsky, Phys. Rev. Lett. **95**, 038102 (2005).

[81] M. Sayar and S. I. Stupp, Phys. Rev. E **72**, 011803 (2005).

[82] H. Noguchi and K. Yoshikawa, J. Chem. Phys. **109**, 5070 (1998).

[83] H. Noguchi, Int. J. Bifurcation Chaos **12**, 2003 (2002).

[84] C. Liu and M. Muthukumar, J. Chem. Phys. **109**, 2536 (1998).

[85] L. S. Pam, L. L. Spell, and J. T. Kindt, J. Chem. Phys. **126**, 134906 (2007).