Conformational Mechanics for Polymers Doubly-grafted to a

Homogeneous Substrate

by

Shengming Zhang

(Under the Direction of Michael Bachmann)

Abstract

Studies of polymer structures with their two ends anchored at a planar substrate can help to provide insights into conformational properties of biologically active systems such as molecular motors involved in cellular transport processes. In this dissertation, the conformational phase behavior of a coarse-grained flexible homopolymer model is investigated. For this purpose, extensive parallel tempering computer simulations were performed. Specific energetic and structural quantities were measured and used as indicator functions for the characterization of the conformational phases. Based on these results, phase diagrams, characterized by various temperatures, distances and substrate surface attraction strength values were constructed. Unique conformational phases and phase transitions are analyzed in detail.

Index words: Flexible Polymers, Polymer and Substrate, Polymer Substrate
                 Interaction, Coarse-grained Model, Canonical Analysis, Monte Carlo
                 Simulations, Metropolis, Parallel Tempering, Order Parameters,
                 Conformational phases, Phase Transitions, Finite Systems

CONFORMATIONAL MECHANICS FOR POLYMERS DOUBLY-GRAFTED TO A

HOMOGENEOUS SUBSTRATE

by

SHENGMING ZHANG

B.S., Northeast Forestry University, China, 2006

M.S., Northeast Forestry University, China, 2009

M.S., Indiana University of Pennsylvania, 2011

M.S., University of Georgia, 2018

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

CONFORMATIONAL MECHANICS FOR POLYMERS DOUBLY-GRAFTED TO A

HOMOGENEOUS SUBSTRATE

by

SHENGMING ZHANG

Major Professor:	Michael Bachmann
Committee:	David P. Landau
	Heinz-Bernd Schüttler
	Shan-Ho Tsai

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2021

To My Parents, My Beginning for Life

To My Wife, My Help in Times of Trouble

To My Son, My Enjoyment in Times of Affliction

To My Sages, My Enlightenment in Times of Hardship

To My Universe, My Purposes and Pursuits for My Eternal Life

Acknowledgments

I would like to thank my wife, Susannah, for her love, care, understanding, support and sacrifice during this Ph.D. process. I would like to thank my son, Alan, for his companionship. His presence motivates me to finish what I started. It has been a long and meaningful journey, and I could not have done it without them.

I would like to thank my parents, Mr. Zuqi and Mrs. Xianfen Zhang for their encouragement to persevere in my pursuit of the Ph.D. degree and financial support.

I extend my gratitude to my major professor, Dr. Michael Bachmann, who is a true example of an intelligent and diligent physicist. I will always remember his guidance in my academic training, his knowledgeable mind and enthusiasm in almost every aspect of physics.

I would like to thank Dr. Landau. His PHYS 6601 course was my first computational statistical physics course. At the time, I was not familiar with computational concepts nor statistical physics. Dr. Landau introduced me to the world of computational statistical physics. I also would like to thank Dr. Schüttler, who has provided me with endless discussions and valuable advice. I also would like to thank Dr. Tsai for answering my many questions in simulation physics and computational techniques. She has also coached me on how to use clusters in the Georgia Advanced Computing Resource Center (GACRC).

I also greatly appreciate Mike Caplinger, Jeff Deroshia and Dr. Zhuofei Hou for their timely and professional technical support. I extend my gratitude to Dr. Magnani, Dr.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Polymers have many valuable and irreplaceable uses in our modern world. They are of great interest to both academia and industry. Polymeric products are everywhere, such as plastic bags, Teflon-coated pans, synthetic fiber clothes, rubber polymers in automobiles, polymers-based paints on spaceships, silicone heart valves, etc.

Polymers are chains with an unspecified number of monomeric units. Polymer molecules can have a very high molecular weight (from 10 000 to 1 000 000 g/mol). A major genre of natural polymers is biological polymers, which include polysaccharide, nucleic acid (DNA, RNA) and proteins. There has been rapid development and research related to the properties of biological polymers. One application is imprinting biological polymers on sensor chips for surface plasmon resonance (SPR) to monitor low molecular weight analytes during real time biological interactions [1]. The state-of-art experimental physical techniques are applied to biological studies. For example, cryo-electron microscopy (cryo-EM) can determine complex molecular structures at a resolution as high as 4.5 Å and can investigate molecular complexes as small as 170 kilodaltons (kDa) [2].

Magnetic tweezers [3] and optical tweezers [4] can provide information of a single molecule (macromolecule) *in vitro* manipulation. With a bead attached to a macromolecular chain

such as nucleic acid (DNA, RNA) or proteins, the interaction between nucleic acid (DNA, RNA) and proteins can be investigated. The force of such interaction can be measured and the proteins and nucleic acid structures can be detected. Currently, experimental biophysicists are working hard to develop high throughput and real time imaging and manipulating techniques at single-molecule level [5, 6]. They are all state-of-the-art methods and valuable for solving their specific problems. However, they can only provide detailed information of either static structures or biological complex information of very specific structures in certain surrounding environments.

In contrast, properly selected computational simulation methods can obtain more general thermodynamic and structural properties of a biological polymer in a large variety of environments [7]. Due to complex molecular structures and the huge amount of degrees of freedom in bio-polymers, a coarse-grained model is chosen over an atomistic model [8]. Applying a properly selected polymer model is important for obtaining information to achieve a more general understanding of its structural and biomolecular functions.

Several efficient and effective computational methods have been developed, such as the microcanonical inflection point method [9, 10, 11], multicanonical Monte Carlo simulations [12, 13, 14], Wang-Landau sampling [15, 16, 17, 18], parallel tempering [19, 20, 21, 22] etc.

In this dissertation, a flexible coarse-grained homopolymer with its two ends attached to a homogeneous substrate is simulated using the replica-exchange (parallel tempering) Monte Carlo method. Thermodynamic and structural quantities are measured during the simulations. Considerable analysis was performed on these quantities and different conformations of this system under different thermal conditions are studied and presented. The phase diagram, parameterized by temperature and the substrate attraction strength of such a polymer-substrate system with a fixed end-to-end distance, is created. Another phase diagram, parameterized by temperature and the end-to-end distance of such a polymer-

substrate system with a fixed attraction strength, is also created. These phase diagrams provide generic structural information of such a system.

The thermodynamic and statistical physics concepts used in this dissertation are introduced in Chapter 2. The model used in this dissertation is presented in Chapter 3. Chapter 4 describes the simulation methods. Polymer-substrate systems with a fixed end-to-end distance with various temperatures and various substrate surface attraction strengths are studied in Chapter 5. Study of polymer-substrate systems with a fixed substrate surface attraction strength with various temperatures and various end-to-end distances is presented in Chapter 6. The summary of this dissertation is given in Chapter 7.

Chapter 2

Essential Thermodynamics and Statistical Mechanics

The thermodynamics and statistical mechanics concepts used in the dissertation are explained in this chapter. The connections and differences between the thermodynamics and statistical mechanics concepts of the same studied system are clarified and the thermal physics foundations are introduced in this chapter.

Thermodynamics studies temperature related questions of macroscopic systems. It is used to handle quantities involving heat, work, internal energy, temperature and structural properties of macroscopic systems. As an empirical theory, its four fundamental thermodynamics laws describe the behaviors of these thermal quantities of a system macroscopically, but not microscopically. These quantities are usually averaged at the thermal equilibrium of a macroscopic system and explained theoretically in terms of microscopic states by statistical mechanics, which will be introduced in Section 2.2.

2.1 Essential Thermodynamics

The first law of Thermodynamics is the conservation of energy law in the field of thermodynamic physics. For an *adiabatic* process[1], the change of internal energy U is:

$$\Delta U = \Delta W, \tag{2.1}$$

where $\Delta U = U_{\text{final}} - U_{\text{initial}}$ and ΔW is the work done to this system by other objects. The work being done to this system will change the thermal state of such a system. Note that the internal energy is a state variable to describe the current thermal state of the system, which will not be dependent on the paths between the initial state with internal energy U_{initial} and the final state with internal energy U_{final}. For a non-adiabatic process, the heat obtained from outside is: $\Delta Q = \Delta U - \Delta W$, or

$$\Delta U = \Delta Q + \Delta W, \tag{2.2}$$

which reflects the energy conservation nature of a general system. Note that heat and work are not state variables and they are dependent on the paths. $đQ$ denotes the infinitesimal change in heat with respect to the change in temperature dT. The heat capacity $C_{\text{V}} = đQ/dT$ describes how much heat needs to be added to the system to increase its temperature by an infinitesimal amount dT when the volume of such a system remains unchanged, therefore no work was done to the system. Then $dU = đQ$ and $C_{\text{V}} = (\partial U/\partial T)_{\text{V}}$.

The second law introduces a new state variable called entropy S. For a closed system connected with a heat bath [2], the entropy always tends to maximize itself, $dS \geq đQ/T$, where "=" is for a reversible process, and ">" is for an irreversible process. T is both

---

[1]An adiabatic process is a type of thermodynamic process which occurs without heat exchange between the system and its surroundings.

[2]The heat bath can only exchange heat with the closed system.

the temperature of the system and the heat bath for a reversible process. However, for an irreversible process, T is the temperature of the heat bath and does not have to be the temperature of the system. The thermodynamic definition of entropy is an empirical quantity, which can be used to measure the heat absorbed by a system connected to a heat bath with temperature T. In a *quasistatic, adiabatic* process $dQ = TdS$ [3]. This leads to the third law: a system will have its minimized entropy or zero entropy at a theoretical absolute zero temperature which may never be achieved in practice.

Finally, the zeroth law defines the thermodynamic equilibrium between systems. This postulate induces the "existence of temperature".

2.1.1   Thermodynamics Temperature and Free Energy

Simulation of replicas of polymer-substrate system at a range of specific temperatures has been performed. The temperature of each replica is not changed and equals an assigned temperature. These systems can be considered as a set of systems connected to a heat bath with a range of temperatures [4]. The expression of the 1st law of such a system is:

$$dU(S, V, N) = TdS - pdV + \mu dN ,$$
(2.3)

where

$$T = \left(\frac{\partial U}{\partial S} \right)_{N,V} ,$$
(2.4)

---

[3] The concept of entropy is the key to Thermodynamics. Its nature is not clear until its statistical mechanics definition, $S = k_{\mathrm{B}} \ln \Omega$, is given. Ω is the total number of the microstates accessible to the closed system. When S is large, the microstates of the system are large and each is in a disorder state. In contrast, if S is small, the system is in an order state.

[4] The system is essentially a canonical ensemble in statistical mechanics and it will be discussed in Section 2.2.

is the thermodynamics temperature, $p = -(\partial U/\partial V)_{S,N}$ and $\mu = (\partial U/\partial N)_{S,V}$. For a closed system, there are no changes in V and N; therefore the change of internal energy dU is only related to TdS. This means the heat will be transferred from or to the heat bath.

According to the second law, $T\Delta S \geq Q \geq 0$. Therefore, for a closed system connected to a heat bath, a new state variable can represent the energy which is able to be used for the work of such a system. This function can be introduced by applying Legendre transformation[5] to TdS. Plug it into Equation 2.3, resulting in $d(U - TS) = -SdT - pdV + \mu dN$. Therefore, the state variable is

$$F(T, V, N) = U(S(T, V, N), V, N) - TS(T, V, N) \, , \tag{2.5}$$

which is called the Helmholtz free energy. It minimizes itself in a thermal equilibrium while the entropy tries to be maximized. Since it is the only free energy considered in this dissertation, we will use free energy to refer to the Helmholtz free energy.

Its differential form is

$$dF = -SdT - pdV + \mu dN \, . \tag{2.6}$$

It has great uses for identifying phase transitions. Its statistical mechanics definition is also important and will be introduced in Section 2.2.

2.1.2 Thermodynamics Phases, Conformational Phases and Phase Transitions

When a system is under changes of external quantities, its macroscopic features could show similarity and the changes of these features could be smooth; hence, the macrostates of the system are in the same phase. When a system experiences a phase transition, a small change

---

[5]Legendre transformation: $xdy \equiv d(xy) - ydx$. Apply it to TdS, resulting in $TdS = d(TS) - SdT$.

of the external quantities could lead to a dramatic change in the macrostates of such a system. Usually C_V can be used as a response parameter to identify phase transitions. For a polymer-substrate system, its conformational parameters are usually used as response parameters to identify its phase transitions [23, 24]. Therefore these conformational parameters can be used to categorize the macrostates of the system into different conformational phases (e.g., gas-like random coils, liquid globule, crystalline). These conformational parameters will be introduced in Chapter 3.

Traditionally, two types of phase transitions are identified, according to the continuity of entropy of such a system at the phase transition temperatures. By taking a partial derivative of the free energy over T in Equation 2.6, we have the entropy

$$S(T, V, N) = - \left(\frac{\partial F}{\partial T} \right)_{N,V} . \tag{2.7}$$

If the entropy is also continuous, the second derivative of F can be taken with respect to T, resulting in

$$\left(\frac{\partial^2 F}{\partial T^2} \right)_{N,V} = - \left(\frac{\partial S}{\partial T} \right)_{N,V} = -\frac{1}{T} C_V(T) , \tag{2.8}$$

where $C_V = T(\partial S/\partial T)_{N,V} = (\partial U/\partial T)_{N,V}$ is the heat capacity. It represents the capacity of how a system responds to the heat with the amount, TdS, exchanged between such a system and its surroundings while its temperature changes.

First Order Phase Transition

If the entropy is discontinuous, as shown in Figure 2.1, a gap shows up on the entropy curve between the ordered and disordered phases. At temperature T_{tr}, the difference of the entropy is related to latent heat

$$\Delta Q_{lat} = T_{tr} \Delta S , \tag{2.9}$$

where Q_{lat} is the energy released or absorbed by a chemical substance or a thermodynamic system during a change of state that occurs without a change in temperature. Q_{lat} can be used to distinguish first- and second-order transitions. It is nonzero in first-order phase transitions and zero in second-order phase transitions. Traditionally, if entropy is discontinuous and the latent heat is greater than zero, this type of phase transitions is classified as first-order transitions. The temperature at which the transition happens is denoted as transition temperature T_{tr}. In a first-order transition, from high temperature to low temperature, two phases coexist, heat will be released and the entropy of the system will decrease. This is equivalent to the Ehrenfest classification, which generally classifies the transitions by the first non-analyticity in the derivatives of the Gibbs enthalpy [25].
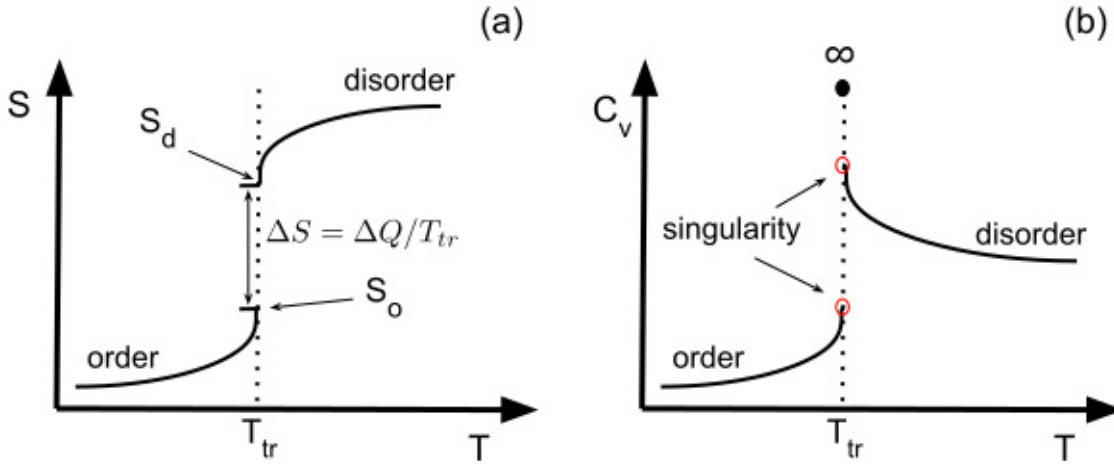
Figure 2.1: First-order phase transition. (a) shows the discontinuity of the entropy and (b) shows the singularity of the heat capacity.

Second-order Phase Transition

If entropy is continuous, but the derivative of entropy with respect to temperature is not continuous, then the transition is a second-order phase transition. Figure 2.2 shows a typical

second-order transition[6]. The transition temperature of a second-order phase transition is denoted as critical temperature T_c. When approaching the critical temperature, important quantities of the system such as heat capacity, correlation length, magnetic susceptibility etc. could show a series of critical power law behaviors. These power law behaviors are valid for many different physical systems. These series of critical power law behaviors shed light on the theory of *Universality*, which is a hot topic in statistical physics.
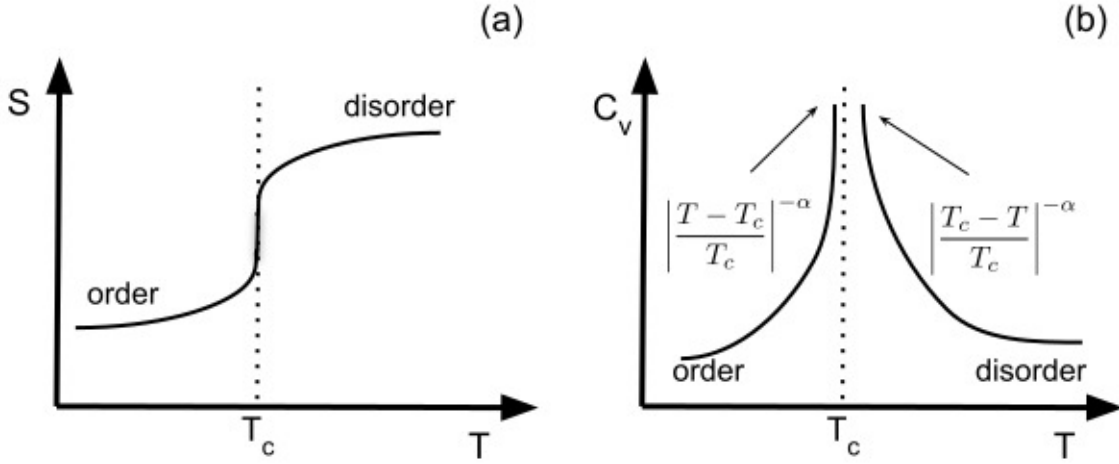
Figure 2.2: Second-order phase transition. This figure shows a general case of a second-order phase transition. (a) shows a smooth change of the entropy at T_c and (b) shows a discontinuity of C_V at T_c. When the system crosses over T_c, no heat is released, no two phases coexist and phase transition will happen immediately.

We try to define different phases and phase transitions because we want to classify different systems and their generic properties. Therefore, we can study the similarities of them and summarize the differences between them. Eventually we hope to find universal physics laws of different systems.

---

[6]The shape of C_V can be different, depending on the shape of entropy. There is also a λ type second-order phase transition, for which its C_V vs. T phase diagram looks like a λ shape, e.g., normal fluid helium (helium I) transitions to superfluid helium II.

Order Parameter

If a parameter is zero at one phase and not zero at another phase, it can be used to identify different phases. The magnetization M of a 2D Ising model is a good example of such a parameter because $M = 0$ when $T > T_c$ and $M \neq 0$ when $T < T_c$. Such a parameter is called an order parameter. It is an ideal case if an order parameter can be found to identify different phases and phase transitions. Often an order parameter is not easy to identify. In such cases, a comprehensive analysis of a set of energetic and conformational parameters[7] is necessary to identify different phases and phase transitions.

Thermodynamics Limits and Phase Transitions in Finite Systems

All of the above discussions are based on an assumption that the system is at its thermodynamic limit. However in nanotechnology and biological systems, the research subjects are of finite sizes. For example, the surface plasmon resonance (SPR) sensing chips are covered with polymers chains of limited length. The length of a protein is also limited, averaging around 300 residues [26]. For example, Hemoglobin subunit beta consists of 147 amino acids and has a molecular weight of 15,867 Da. Therefore, it is important to identify phases and "phase" transitions of finite size polymers.

Figure 2.3 shows the finite size effect of a 2D $L \times L$ Ising model with $L = 10,\ 20$ and 40, where L is the linear size of a square lattice. The peak widths of both C_V/N and χ/N become narrower and sharper as L increases, where $N = L \times L$ is the number of total spins, C_V is the heat capacity and χ is the thermal magnetic susceptibility. The locations of the peaks indicate the transition temperatures of the corresponding quantities. As the size increases, the transition temperatures indicated by the peaks of the C_V/N and χ/N curves approach the theoretical $T_c \approx 2.269$ of the 2D infinite Ising model, which is a good demonstration of the finite size effect. The discontinuities of response quantity curves at

---

[7]These parameters will be introduced in Chapter 3.
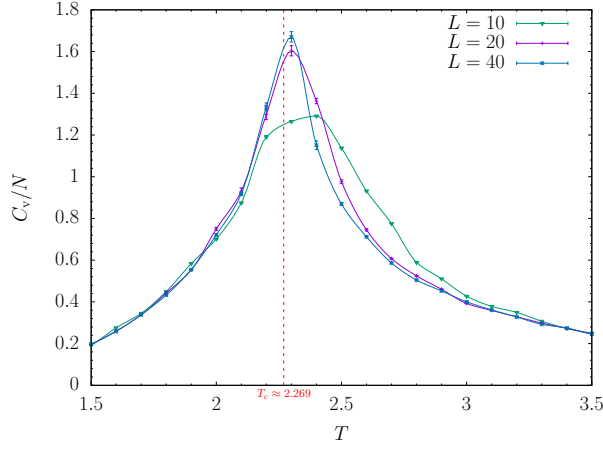
(a) C_V/N vs. T for 2D Ising model.
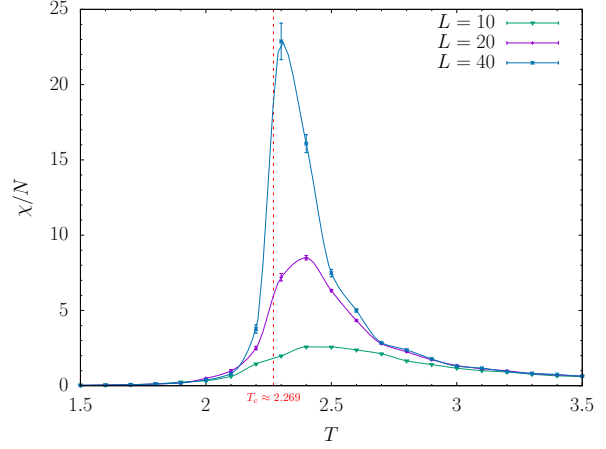
(b) χ/N vs. T for 2D Ising model.

Figure 2.3: The finite size effect of a 2D Ising model with $L = 10, \ 20, \ \text{and} \ 40$.

phase transitions can only be identified for infinite systems. For finite size systems, there are no discontinuities on such curves. Sometimes the peak, valley or shoulder patterns of response quantity curves could indicate a phase transition, as shown in [7]. Usually for finite size systems, the indicated T_{tr} is not a single value but rather a range of the possible T_{tr} values. In such cases, a transition temperature band can be obtained instead of a single transition temperature at the thermodynamic limit.

Pseudophases

A notable pattern of finite polymer phase transitions is that these conformational transitions are smooth processes comparing with dramatic thermodynamic phase transitions at infinite scale. Therefore these conformational phase transitions do not behave like real thermodynamic phase transitions, but are rather described as conformational pseudophase transitions since the polymers simulated in this dissertation are finite sized. Similarly, these conformational phases of a polymer are defined as pseudophases. In this dissertation, for short,

phases and phase transitions will be used as the pseudophases and pseudophase transitions for finite length polymers.

2.2 Canonical Analysis

Section 2.1 gives the essential concepts of thermodynamics used in this dissertation. In this section, essential statistical mechanics related to this dissertation is given. Statistical mechanics provides a microscopic perspective to the macroscopic properties of a thermodynamic system and uses concepts of thermal fluctuations to analyze quantities of such a system. The probability of a microstate of a macroscopic thermal system is the key to statistical mechanics. Thermal systems can be studied using different ensembles, in which the microstates can be interpreted with different probability distributions. Microcanonical ensembles, canonical ensembles and grand canonical ensembles are the commonly used ones [7, 27].

2.2.1 Canonical Ensemble

The system, embedded into a heat bath at a specific temperature, is represented by a canonical ensemble. This ensemble can be considered as a closed system connected to a large heat bath with temperature T. The number of particles (N), the volume of the system (V) and the temperature T of such a system do not change. The temperature of this system is the temperature of the heat bath. Therefore it can be referred to as an NVT ensemble. Energy could exchange between the system and the heat bath even in thermal equilibrium via thermal fluctuation. The probability of a microstate s in the canonical ensemble is

$$p_s = \frac{1}{Z}e^{-\beta E_s} \, , \tag{2.10}$$

where E_s is the energy of the system at the microstate s, $\beta = \frac{1}{k_B T}$ is the inverse temperature and k_B is the Boltzmann constant. T is the canonical temperature, which is equal to the temperature of the heat bath. Z is the partition function

$$Z(T) = \sum_s e^{-\beta E_s} \ , \tag{2.11}$$

which is the summation over all the possible microstates of the system. Z can also be obtained by summing over all the different energy,

$$Z(T) = \sum_E g_E e^{-\beta E} \ , \tag{2.12}$$

where g_E is the number of microstates with energy E. In this way, the probability of a canonical ensemble with energy E is

$$p(E) = \frac{1}{Z} g_E e^{-\beta E} \ . \tag{2.13}$$

In a simulation, an energy histogram can be obtained and an approximate energy probability distribution can be calculated by normalizing the energy histogram. This energy histogram can be used for thermodynamic studies. In the canonical ensemble, knowing Z and $p(E)$, the average energy $\langle E \rangle$ of the system at equilibrium with temperature T is

$$\langle E \rangle = \sum_E E p(E) = \frac{1}{Z} \sum_E E g_E e^{-\beta E} = k_B T^2 \left(\frac{\partial}{\partial T} \ln Z \right)_{T,V,N} \ . \tag{2.14}$$

The Helmholtz free energy is

$$F = -k_B T \ln Z \ . \tag{2.15}$$

14

From Equation 2.7 the entropy is

$$S(T, V, N) = -\left(\frac{\partial F}{\partial T}\right)_{N,V} = k_B \ln Z + \frac{\langle E \rangle}{T} . \tag{2.16}$$

2.2.2 Fluctuation

The thermal fluctuation is the derivative of an observable with respect to T. The heat capacity can also be expressed as

$$
\begin{aligned}
C_V &= \frac{d \langle E \rangle}{dT} \\
&= \frac{d}{dT}\left(k_B T^2 \left(\frac{\partial}{\partial T} \ln Z\right)\right) \\
&= \frac{1}{k_B T^2}\left[\frac{1}{Z}\frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2}\left(\frac{\partial Z}{\partial \beta}\right)^2\right] \\
&= \frac{1}{k_B T^2}\left[\frac{1}{Z}\sum_E E^2 g_E e^{-\beta E} - \left(\frac{1}{Z}\sum_E E g_E e^{-\beta E}\right)^2\right] \\
&= \frac{1}{k_B T^2}\left[\langle E^2 \rangle - \langle E \rangle^2\right] ,
\end{aligned}
\tag{2.17}
$$

which could also be considered as a responsive value of the system to its energetic fluctuation. Similarly, for any observable O, its energetic fluctuation responding to a temperature change is

$$
\begin{aligned}
\frac{d \langle O \rangle}{dT} &= \frac{d}{dT}\frac{1}{Z}\sum_E O g_E e^{-\beta E} \\
&= -\frac{1}{k_B T^2}\left[\frac{1}{Z}\left(\frac{\partial}{\partial \beta}\left(\sum_E O g_E e^{-\beta E}\right)\right) + \left(\frac{\partial}{\partial \beta}\left(\frac{1}{\sum_E g_E e^{-\beta E}}\right)\right)\sum_E O g_E e^{-\beta E}\right] \\
&= \frac{1}{k_B T^2}\left[\frac{1}{Z}\sum_E O E g_E e^{-\beta E} - \left(\frac{1}{Z}\sum_E O g_E e^{-\beta E}\right)\left(\frac{1}{Z}\sum_E E g_E e^{-\beta E}\right)\right] \\
&= \frac{1}{k_B T^2}\left[\langle OE \rangle - \langle O \rangle \langle E \rangle\right] .
\end{aligned}
\tag{2.18}
$$

$d \langle O \rangle / dT$ could be used to discern phase transitions. The observables used in this dissertation will be introduced in Chapter 3.
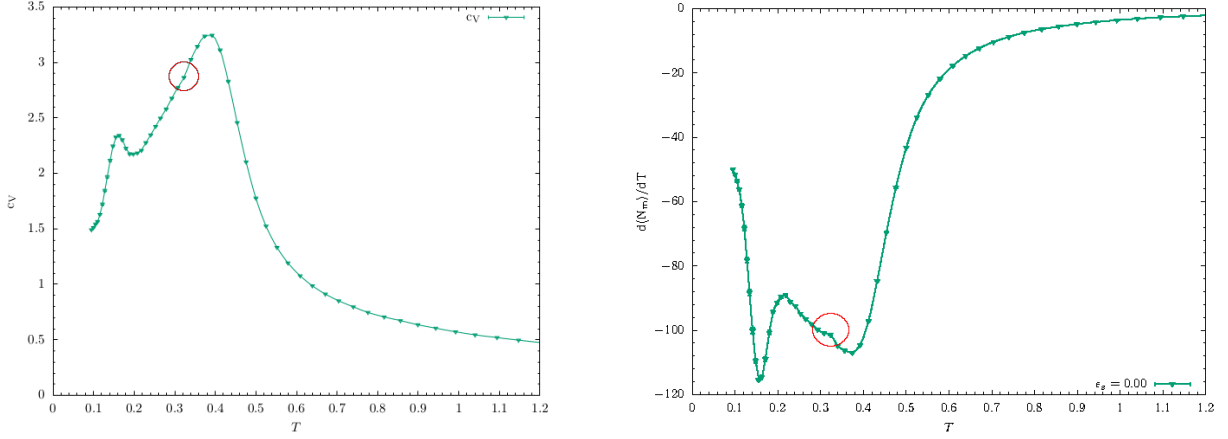
(a) c_V vs. T for a 40mer-substrate system.

(b) $d \langle N_m \rangle / dT$ vs. T for a 40mer-substrate system.

Figure 2.4: The specific heat c_V vs. T curve and the thermal fluctuation of the number of monomer-monomer contacts $d \langle N_m \rangle / dT$ vs. T for a 40mer polymer-substrate system.

As discussed in Section 2.1.2 and shown in Figure 2.2, $d \langle O \rangle / dT$ most likely will experience a singularity and a power law behavior when approaching T_{tr} in the thermodynamic limit. However for a finite system, the responsive quantities will never experience a singularity, as shown in Figure 2.3. Sometimes a shoulder shape on a c_V curve could also be a hint for a phase transition. Other responsive quantities may show more apparent patterns on their curves thus indicating the same phase transition to match with the specific heat $c_V = C_V / N$ curve as shown in Figure 2.4.

Since no singularity will be observed, it is difficult to distinguish between the first- and second-order "phase" transitions. As mentioned in Section 2.1, at the transition temperature T_{tr} of a first-order transition, two phases will coexist and Q_{lat} will be released. Therefore we can use the energy of the system as an order parameter to obtain an energetic histogram of different temperatures. Equation 2.13 gives the normalized form of such a histogram, which is proportional to $e^{-\beta E}$ and should only have one peak. One of the histograms represents

16

one of the macroscopic phases. If two peaks are present, this means two macroscopic phases coexist at the transition temperature. This two-peak pattern of a energy histogram indicates a first-order phase transition. For a second-order phase transition, the energy histogram is unimodal. The difference between the two energy peaks of the distribution is a measure for the latent heat [28]. Figure 2.5 gives an example of such a transition. There are double peaks at $T = 0.31$ and 0.316. Therefore, T_{tr} is in between them. Note that two peaks are located approximately at $E_1 \approx 231.5$ and $E_2 \approx 215.5$, and the energy between E_1 and E_2 is the latent heat.
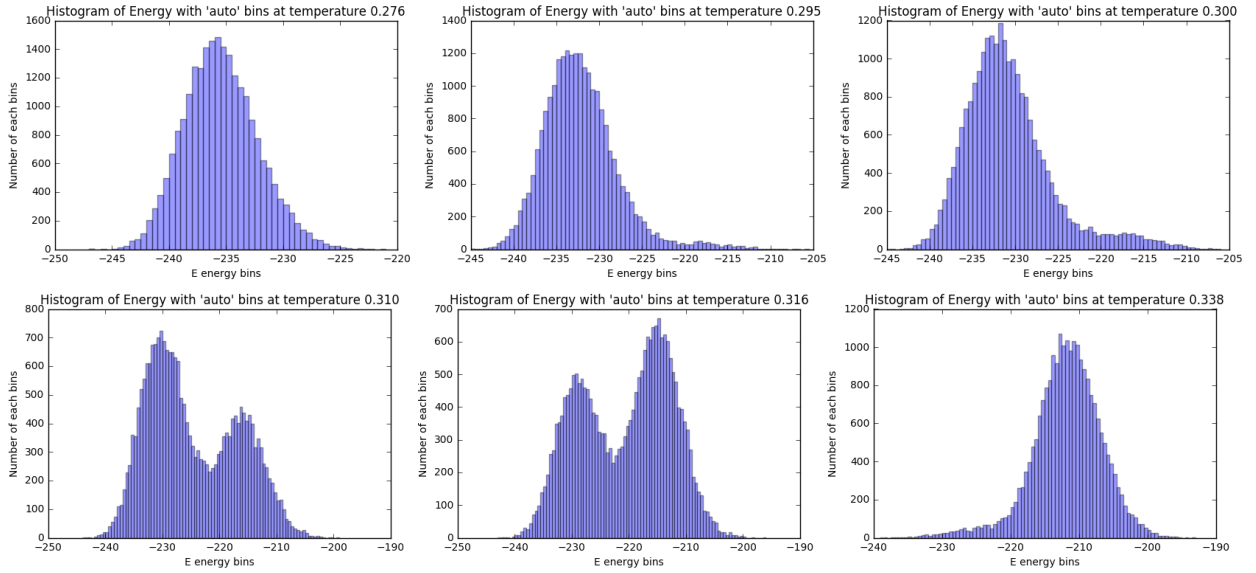
Figure 2.5: Energy histograms of a 55mer free polymer at different temperatures.

2.2.3 Internal Energy U, Energy E, Free Energy F and Potential Energy V

The internal energy U is a macroscopic quantity. It is the summation of all the energy of the subsystems of a given system. According to Equation 2.3, it is constant and the entropy is maximized in thermal equilibrium.

For a given system, an energetic Hamilton operator $\hat{H}$ can be applied to its current microstate and the corresponding value is the energy E. E can be separated into kinetic energy E_k and potential energy V [8]. Therefore E can be decoupled into the microstate of the coordinates in the coordinate space $\mathbf{X}$ and momentum space $\mathbf{P}$ of such a system. E is fluctuating even in thermal equilibrium. Therefore E always fluctuates and it is not a constant. However in equilibrium the average of energy $\langle E \rangle$ satisfies $\langle E \rangle = U$, which is a connection between thermodynamics and the canonical ensembles. At equilibrium, the average kinetic energy is

$$\langle E_k \rangle = \frac{3Nk_BT}{2} \ , \tag{2.19}$$

where N is the number of particles in the system. The kinetic energy's contribution to C_V is $(C_V)_{K.E.} = 3Nk_B/2$, which is a constant and will not be useful in identifying the transition. Therefore we can discard the kinetic energy (Hamiltonian in $\mathbf{P}$ space) part of E when identifying the phase transitions. This is also the reason that the phases are defined as conformational phases. Therefore in this dissertation, the E in the following sections refers to the potential energy related to the conformational $\mathbf{X}$ coordinate. Details of the decoupling of Hamiltonian can be found in Section 2 of [29].

2.2.4 Limitation of Canonical Analysis on Finite Systems and the Microcanonical Inflection Point Method

Canonical analysis is straight forward to apply to a simulation and provides an easy-to-follow connection between thermodynamics and its statistical mechanics nature. As a result, its effectiveness is limited in finite systems. It has the following limitations:

- patterns are not very noticeable;

- it is not easy to separate 1st and 2nd order "phase" transitions;

---

[8]Please notice that V here is not used for volume as before, it presents the potential energy here.

- most importantly, the canonical can only provide a transition band, rather than a transition point, for a given thermal condition.

In recent years, a method based on the microcanonical inflection point is proposed and applied to finite systems such as Ising model, Potts model and polymers models [30, 31, 32, 33, 34, 35]. The essential idea is to define an microcanonical inverse temperature

$$\beta_E = \left(\frac{dS(E)}{dE} \right)_{E=E_{tr}} \tag{2.20}$$

where $S(E) = k_B \ln g(E)$ is the microcanonical entropy and $g(E)$ is the density of states with respect to energy E. By analyzing the pattern of β and the higher order derivatives of $S(E)$ with respect to E, the transition temperatures and different orders of transitions can be identified.

Chapter 3

Models

Polymers are linear structures connected via multiple similar chemical units and can form complex spatial structures. One simple type is that all the similar chemical units are the same. For example, polymerthylene is a long chain chemical structure only consisted of methylene groups $-CH_2-$. In contrast, complex biological polymers consist of different chemical units with similar chemical structures. For example, DNA consists of 4 types of phosphate groups and proteins contain 20 amino acid groups. The atomic model of the exact representation of a polymer could have hundreds of thousands of degrees of freedom. Therefore it is very difficult to simulate such a polymer with current computational resources[1].

3.1   Coarse-Grained Model

Fortunately, polymer structures are not exclusively determined by the atomistic details of each residue but rather are results of cooperative interactions. For example, the cooperative interactions of a protein include interactions of its amino acid side chains and interactions of the side chains and the surrounding solvent particles. The protein folding processes are also highly cooperative processes and could follow many possible pathways [34, 38]. The

---

[1]Current developments of the atomic polymer model simulations can be found in [36, 37].

protein folding processes are governed by the free energy of a set of generalized coordinates and the native structures of proteins possess the minimum free energy among all possible conformations [39]. Therefore, for studying the conformational phases and "phase" transitions, the atomisic details can be omitted and a simplified coarse-grained model with fewer macroscopic degrees of freedoms can be used. As discussed previously, we can introduce *united* atoms, or *monomers*, to replace the atomistic details of a polymer. Then classical potential interactions between the momomers can be used to replace the cooperative quantum and electrical many-body interactions between atoms and their surrounding solvent particles approximately. Note that a united monomer does not have to represent a single residual of polymers. The representation of a united monomer depends on the potentials applied to it and it could represent one residual, partial residual or several residuals. Mathematical details of physics can be found in [7]. Applying the coarse-grained models can not only simulate a polymer more efficiently but also give proper representations of the conformational phases statistically and qualitatively [7].

The effective conformational potentials can be described by the free energy under a relevant set of degrees of freedom [40, 39]. An example of such degrees of freedom is the coordinates of the monomers of a polymer chain. Depending on the aspects of the research problems, different models [7, 41] were developed such as lattice models, in which the polymer can only move at fixed lattice vertices [42, 43, 44]; off-lattice models [45], in which the polymer can move freely in a continuous space [46]; homopolymer models [7], which treat all amino acid residues as identical united monomer; HP, HOP lattice models [47, 48] and AB off-lattice model [49, 50, 51], all of which classify 20 amino acid residues into different groups.

An example of various structural levels of a protein is given in Figure 3.1 and will be explained below. The primary and secondary structures are created by PDBsum [52] and the tertiary and quaternary structures are created by Pymol [53]. This example is fetched from Protein Data Bank (PDB) with id: 6kyi, which is the rice ribulose bisphosphate carboxylase

21

complex [54]. It is used for photosynthesis in rice. From left to right of Figure 3.1, the black box of the left figure indicates the primary structure which is the sequence of a protein. The red box of the left figure and the center figure indicate a secondary structure. The green circles of the center figure and the right figure present a tertiary structure. Finally, the quaternary structure consisted of multiple protein subunits is shown in the right figure. The center figure shows the small chain of such a complex.
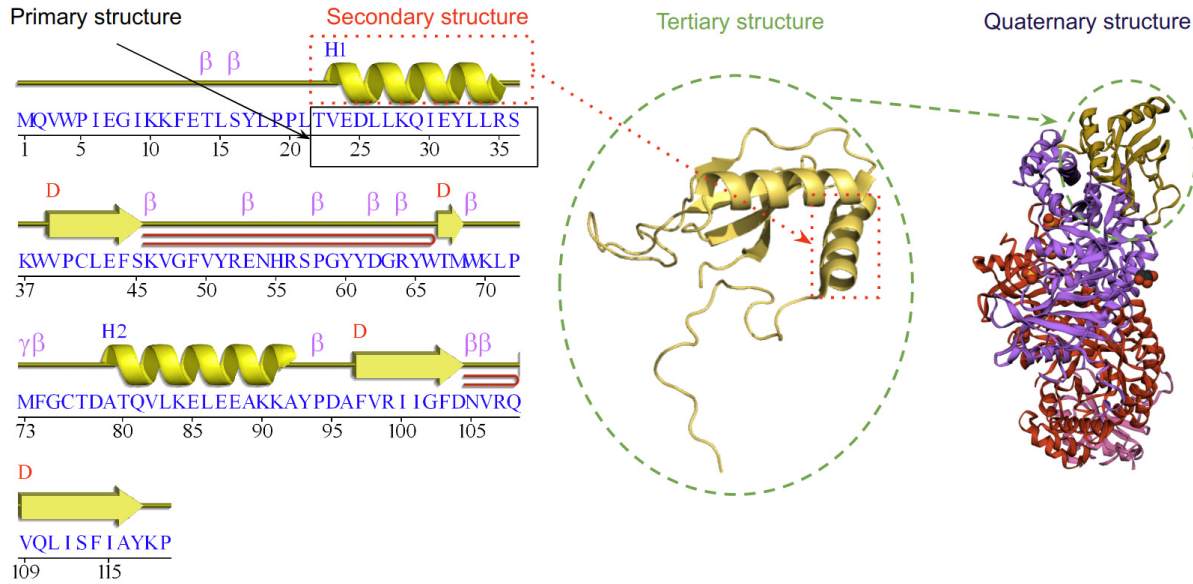
Figure 3.1: Different structure levels of a protein.

Primary structure is simply the order of the amino acid sequence of a protein. The amino acids are connected by peptide bonds. A peptide bond is an amide type of covalent bond formed between two amino acids when the carboxyl group of one molecule reacts with the amino group of the other molecule, releasing a molecule of water.

It requires $8-16$ kJ/mol of Gibbs free energy to break a peptide in a water solvent [55]. In room temperature, $T_{\text{room}} = 25$ °C, the thermal fluctuation energy $k_B T$ is 2.479 kJ/mol. Therefore, the peptide bonds of a protein in a water based solvent will not be broken by thermal fluctuations around room temperature. Such a bonded potential will be introduced

in Section 3.2. However, the bonded potential in a coarse-grained model is not necessarily representing a covalent or peptid bond, but rather an unbreakable effective bond between united monomers; since the interactions between bonded monomers are cooperative. There are also non-bonded interactions such as volume exclusion effect and Coulomb interaction (dipole moment and polarization effect), which can be represented by Lennard-Jones (LJ) potential [56]. Such a potential will also be introduced in Section 3.2.

(a)   (b)
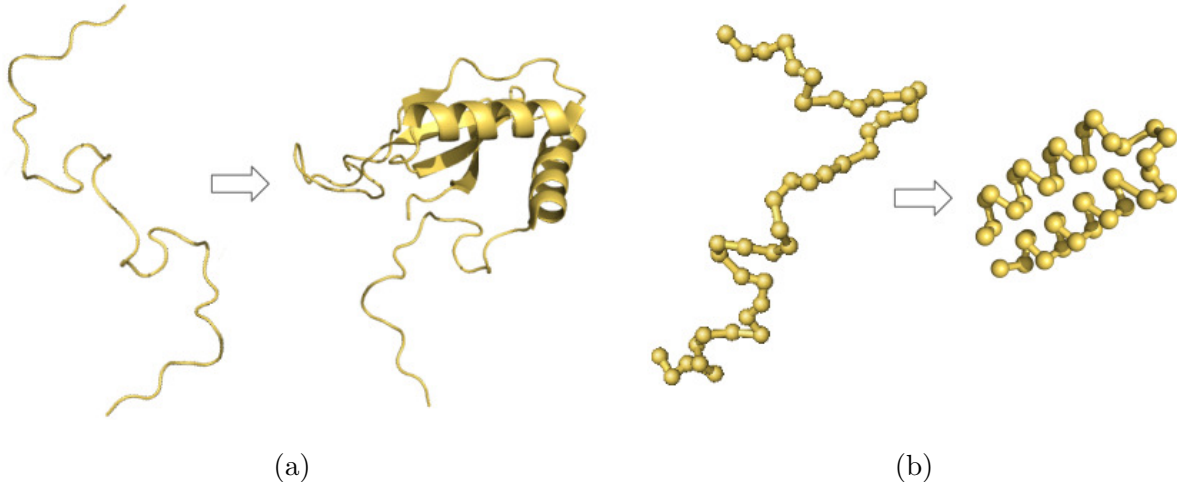
Figure 3.2: The transition of a protein from a secondary random coil to a tertiary structure. (a) shows a cartoon representation of the ribulose bisphosphate carboxylase small chain from a secondary structure to a tertiary structure. (b) shows a coarse-grained model representation of a 40mer homopolymer changing from a random coil to a compact structure with two local helix bundles [57].

As discussed previously, the protein folding and its native structures are cooperative effects and primarily governed by their primary backbone structures. Therefore, a homopolymer with proper potentials could give generic ideas of a protein folding process. An example of such an idea is shown in Figure 3.2, (a) shows a cartoon representation of the ribulose bisphosphate carboxylase small chain changing from a random coil structure to a tertiary structure and (b) shows a coarse-grained model representation of a 40mer homopolymer changing from a random coil to a compact structure with two local helix bundles. The energy of this coarse-grained model can be calculated by $E(\mathbf{X}) = S_{\text{FENE}} \sum_i V_{\text{FENE}}(r_{i,i+1}) +$

$S_{\text{LJ}} \sum_{i>j+1} V_{\text{LJ}}(r_{i,j}) + S_{\tau} \sum_{l} V_{\text{tor}}(\tau_l) + S_{\theta} \sum_{k} V_{\text{bend}}(\theta_k)$, where $\mathbf{X} = \{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ is the conformational coordinates of the monomers of the polymer, V_{FENE} is the FENE potential, V_{LJ} is the LJ potential, V_{tor} is the torsion potential and V_{bend} is the bending potential; S_{FENE}, S_{LJ}, S_{τ} and S_{θ} are the co-factors of the corresponding potentials. Details of these potentials can be found in reference [57, 58, 59]. Figure 3.2 shows that the coarse-grained model with proper potentials is able to represent a protein-like polymer and can provide generic and statistical information of its conformational phases. The tertiary structures are determined by the potentials between the united monomers on the polymer backbones [7]. Depending on the problems, different potentials can be switched on or off. For example, in Figure 3.2, the $S_{\tau} \sum_{l} V_{\text{tor}}(\tau_l)$ and $S_{\theta} \sum_{k} V_{\text{bend}}(\theta_k)$ are switched on to show these local helix structures. If exploring a fairly novel polymer structure, a minimalistic energy using less potentials is desired. Figure 3.3 shows an example of a model with minimalistic energy of a polymer with its two ends anchored into a homogeneous substrate. In this model, the essential problem is how a polymer will behave with its two ends anchored into a substrate. Local details of such a polymer are not primary considerations. Therefore, the $S_{\tau} \sum_{l} V_{\text{tor}}(\tau_l)$ and $S_{\theta} \sum_{k} V_{\text{bend}}(\theta_k)$ are switched off and only the bonded and unbonded potentials are kept to study the phases and phase transitions of a flexible polymer. The minimalistic energy will be introduced in Section 3.2.
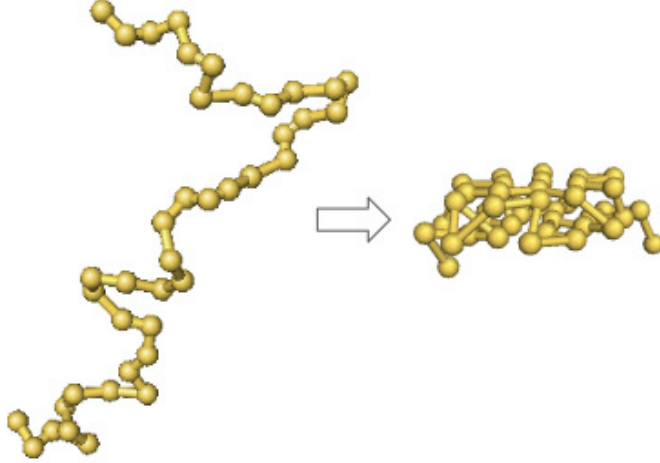
Figure 3.3: Coarse-grained model representation of a process from random coil (disordered structural phase) to a compact crystalline structure (ordered structural phase). Note that the potentials used in Figure 3.2 (b) and the potentials used in this figure are different.

3.2 Potentials

The minimalistic energy of such a polymer is

$$E(\mathbf{X}) = \sum_{i>j+1} V_{NB}(r_{i,j}) + \sum_{i} V_B(r_{i,i+1}) + \sum_{i} V_S(z_i) \qquad (3.1)$$

where $\mathbf{X}$ represents the current conformation of the polymer with energy E, $V_{NB}(r_{i,j})$ is for non-bonded monomers, $V_B(r_{i,i+1})$ is for bonded monomers and $V_S(z_i)$ is for the monomer and substrate potential. z_i is the distance between the substrate surface to the i-th monomer and $z = 0$ for grafted monomers. It is also the energy model used throughout this dissertation. Figure 3.4 shows a general bead-stick representation of such a coarse-grained polymer with its two ends grafted to a homogeneous substrate.
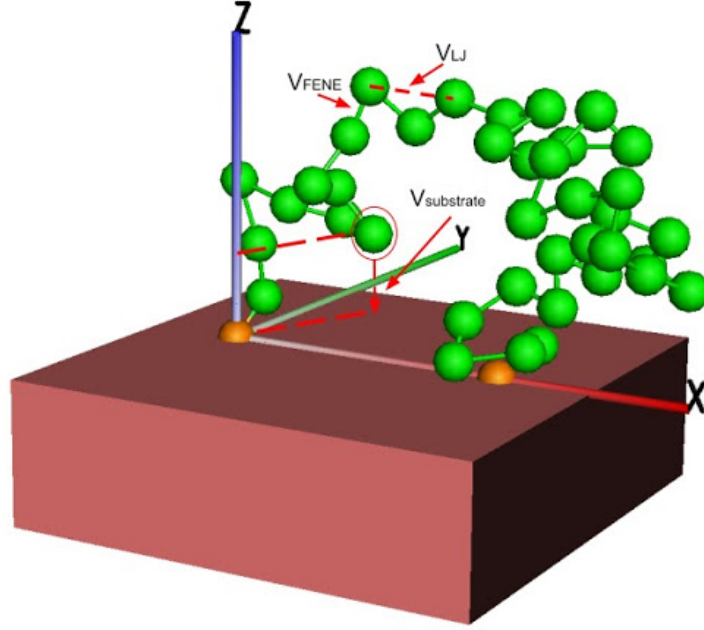
Figure 3.4: Coarse-grained model representation with substrate potential. Bead-stick representation of a coarse-grained polymer with two ends grafted to a homogeneous substrate. The beads represent the monomers. The grafted monomers are colored in orange and the others in green.

3.2.1   Non-Bonded Potential

The non-bonded potential $V_{NB}(r_{i,j})$ is essentially the Coulomb potential between atomic or molecular units. The system is usually electric neutral but there could be dipole-dipole interactions due to uneven distributions of electrons caused by thermal fluctuations in equilibrium.

Assuming there are many charges in the system, its total charge is zero[2]. For a charge q_i, its radius vector is $\mathbf{r_i}$. The potential of the field created by all the charges at $\mathbf{R}$ is

$$V(\mathbf{R}) = \sum_i \frac{q_i}{|\mathbf{R} - \mathbf{r_i}|} \tag{3.2}$$

_____

[2]The electric field is actually created by charge distributions. However for simplicity, only point charges are considered.

where $\mathbf{R} - \mathbf{r_i}$ are the distances between the i-th charge q_i to where the potential is. Assuming $R \gg r_i$, e.g., R is much larger than the Bohr radius in a hydrogen atom [60], the potential can be expanded in powers of $\mathbf{r_i}/R$ using the vectorial Taylor expansion of a scalar function[3]. Equation 3.2 becomes

$$V(\mathbf{R}) = \frac{\sum_i q_i}{R} - \sum_i q_i \mathbf{r_i} \cdot \nabla \frac{1}{R} + \dots \, , \tag{3.3}$$

where the first term is zero since the net charge is zero and the second term corresponds to the dipole-dipole interactions. The higher-order terms which represent higher-order multipole interactions could be omitted, because they are much smaller than the first and second terms. Then Equation 3.3 becomes

$$V(\mathbf{R}) = \frac{(\sum_i q_i \mathbf{r_i}) \cdot \mathbf{R}}{R^3} \, . \tag{3.4}$$

For a pair of particles, Equation 3.4 becomes

$$V(\mathbf{R}) = \frac{q_1 q_2}{R^3} \left(x_1 x_2 + y_1 y_2 - 2 z_1 z_2 \right) \, . \tag{3.5}$$

Equation 3.5 gives the perturbation part of the Hamiltonian of the two particles. The ground-state wave function of a pair of non-interacting particles is $U_0^{(0)} = U_{100}^{(0)}(\mathbf{r_1}) U_{100}^{(0)}(\mathbf{r_2})$, where $U_{n\,l\,m}^{(0)}(r_i)$ is the wave function of the i-th particle at its 0-th energy state and n, l and m are the quantum numbers. According to the Rayleigh-Schrödinger perturbation theory[4], the first-order perturbation energy correction value is approximately zero due to the spherical symmetry of the ground state function; in other words, angular momentum (l) of the i-th

---

[3] $f(\mathbf{R} - \mathbf{r}) = f(\mathbf{R}) - \mathbf{r} \cdot \nabla f(\mathbf{R}) + \dots$ [7, 61]

[4] The perturbation eigenstate is $|n\rangle_\lambda = |n^{(0)}\rangle - \lambda \sum_{k \neq n} \frac{\delta H_{kn}}{E_k^0 - E_n^0} |k^{(0)}\rangle + \mathcal{O}(\lambda^2)$ and the energy with perturbation correction is $E_n(\lambda) = E_n^{(0)} + \lambda H_{nn} + \lambda^2 \sum_{k \neq n} \frac{|\delta H_{kn}|^2}{E_k^0 - E_n^0} + \mathcal{O}(\lambda^3)$ for the Hamiltonian $H(\lambda) = H^{(0)} + \lambda \delta H$, where $\delta H_{kn} = \langle k^{(0)} | \delta H | n^{(0)} \rangle$ [62].

particle equals zero, $l_i = 0$. The second-order energy perturbation correction

$$E^{(2)}(r) = \frac{(q_1 q_2)^2}{R^6} \left(\sum_{k \neq 0} \frac{|\langle k^{(0)}|x_1 x_2 + y_1 y_2 - 2z_1 z_2|0^{(0)}\rangle|^2}{E_0^{(0)} - E_k^{(0)}} \right) \tag{3.6}$$

will not be zero and $E_0^{(0)} < E_k^{(0)}$. Equation 3.6 shows that the interaction[5] between dipoles and charges varies with a function of their distance as $-1/R^6$.

Due to the excluded volumes of each particle, there must be a repulsive term of the interaction between the two particles. For computational efficiency, $1/R^{12}$ has been commonly used, since it is the square of $1/R^6$. The repulsive part could be modified according to different materials [63, 64]. This yields the well-known Lennard-Jones (LJ) potential [56]

$$V_{\mathrm{LJ}}(r) = 4\epsilon_{\mathrm{LJ}} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] , \tag{3.7}$$

where σ is the van der Waals distance which sets the basic length scale and ϵ_{LJ} fixes the energy scale of the interaction. LJ potential will converge towards 0 rapidly with the increase in the distance, but it will never be 0 as shown in Figure 3.5a. For computational simplicity, a cut-off distance r_c of the LJ potential can be set, the LJ potential can then be shifted accordingly. r_c is indicated by the red dashed line in Figure 3.5a. Therefore, the non-bonded potential can be written as the shifted Lennard-Jones (LJ) potential:

$$V_{\mathrm{NB}}(r) = \begin{cases} V_{\mathrm{LJ}}(r) - V_{\mathrm{LJ}}(r_c) , & r < r_c \\ 0 , & r \geq r_c \end{cases} \tag{3.8}$$

where the LJ potential is given in Equation 3.7. $\epsilon_{\mathrm{LJ}} = 1$ is the potential well depth, which sets the energy scale in this simulation. $\sigma = r_0/2^{1/6}$ is the van der Waals radius and $r_0 = 1$,

---

[5]Notice this is derived for interactions between dipoles and charges. Interactions between dipoles and dipoles also vary with a function of their distances as $-1/R^6$ and the derivation is presented in Section 1.6.3 of [7].

28

which fixes the distance scale between particles. $V_{\mathrm{LJ}}(\sigma) = 0$. $r_c = 2.5\sigma$ is the cut-off distance and $V_{\mathrm{LJ}}(r_0) = -1$ is the energy minimum. Notice that ϵ_{LJ} is the energy scale, which can be changed and measured for different systems; r_0 sets the length scale in this simulation. The details of the units of r_0 and ϵ_{LJ} are discussed in Section 5.1.1, where a specific simulation is conducted and the parameters of this simulation are provided.

3.2.2 Bonded Potentials

Bonded potential including both LJ potential and FENE (Finitely Extensible Nonlinear Elastic) potential [58, 65] is

$$V_{\mathrm{B}}(r) = V_{\mathrm{FENE}}(r) + V_{\mathrm{LJ}}(r) - V_{\mathrm{LJ}}(r_c) \ , \tag{3.9}$$

where the FENE potential is

$$V_{\mathrm{FENE}}(r) = -\frac{K}{2}R^2 ln\left[1 - \left(\frac{r - r_0}{R}\right)^2\right] \tag{3.10}$$

where $r_0 = 1$, $K = 98/5$ and $r \in [r_0 - R, r_0 + R]$ [66]. For $r \approx r_0$, Equation 3.10 can be expanded via Taylor expansion[6] and reduced to a harmonic oscillator potential as used in [65]. Figure 3.5b shows a sketch of the FENE potential.

3.2.3 Monomer-Substrate Potential

Recall the sketch of the polymer-substrate system shown in Figure 3.4, the substrate is considered as a homogeneous material with small monomer-type volume elements distributed evenly. Therefore the small volume elements of the substrate can be considered as small united monomers and will have the LJ potential with the monomers of the polymer. Figure

---

[6] $ln\,(1 + x) = x - 1/2x^2 + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}x^n$ when $x \in (-1, 1)$

(a) LJ potential.
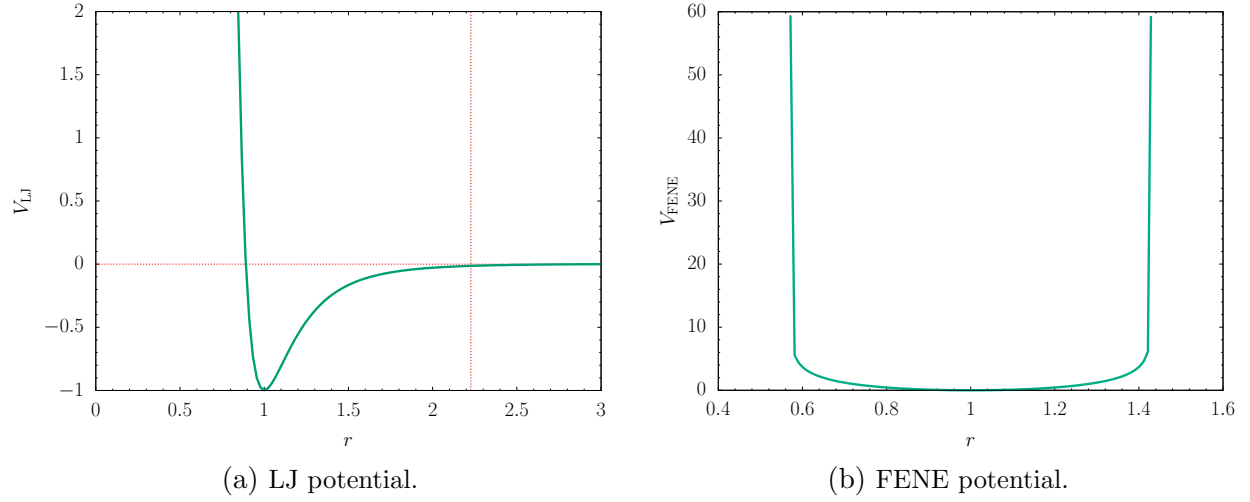
(b) FENE potential.

Figure 3.5: LJ potential and FENE potential are used in the simulations, where r is the monomer-monomer distance in the unit of r_0.
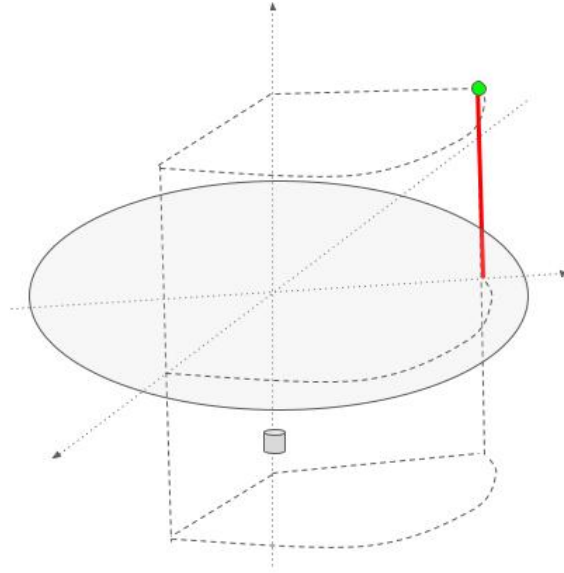
Figure 3.6: Cylindrical coordinate representative of monomer-substrate volume element potential.

3.6 shows this interaction between a volume element (grey cylinder) and a monomer (green bead), where the red line gives the distance to the substrate surface. Therefore in a cylindrical coordinate, which the z direction is perpendicular to the substrate surface and the the

30

substrate-monomer potential is

$$
\begin{aligned}
V_{\mathrm{S}}(z) &= 4n \int_0^{2\pi} d\phi \int_{-\infty}^{z} dz' \int_0^{\infty} \rho d\rho \ \left([(z')^2 + \rho^2]^{-6} - [(z')^2 + \rho^2]^{-3} \right) \\
&= \frac{2\pi n}{3} \left(\frac{2}{15} z^{-9} - z^{-3} \right) \\
&= \epsilon_{\mathrm{s}} \left(\frac{2}{15} z^{-9} - z^{-3} \right)
\end{aligned}
\tag{3.11}
$$

where: $\epsilon_{\mathrm{s}} = \frac{2\pi n}{3}$ is the substrate surface attraction strength and n is the effective "monomer" density [11, 23, 67]. It is important to note that $V_{\mathrm{S}}(z)$ only depends on the z component of the monomer. The polymer-substrate potential can be obtained by summing up the $V_{\mathrm{S}}(z)$ for all monomers.

Finally, using Equation 3.1, the energy of the polymer-substrate system as shown in Figure 3.4 can be calculated.

3.3 Canonical Quantities for Discerning Phase Transitions

For a polymer in a canonical ensemble, its specific heat, end-to-end distance and other thermal quantities are commonly studied for discerning phases and phase transitions [68, 69]. However in this model, since the two ends are anchored onto the substrate, the end-to-end distance will be given as a controllable structural parameter. The radius of gyration representing the size of a polymer will be useful to calculate as a structural quantity. Moreover x, y, z components of the radius of gyration will also be important because the two ends of the polymer are set along the x direction and the z direction is along the vertical direction from the substrate. Therefore, the system will exhibit unsymmetrical (anisotropic) properties along x, y, z directions. Since $V_{\mathrm{S}}(z)$ is only related to z components of the monomer

positions, the z component of the center of mass will be of great interest to use as a possible order parameter.

When the system is in its ordered phases, the distances between monomers will be small and they will be close to each other. If the distance between a pair of monomers is less than a chosen value r_m, the monomers will be referred to as having contact with each other. A quantity named as number of monomer-monomer contacts N_m can be defined and used to describe this monomer-monomer contact. Similarly, the number of monomer-substrate contacts N_s can also be defined as the following: if the z component of the monomer is less than a chosen value r_s, the monomer is considered to be in contact with the substrate.

Then these eight energetic and structural quantities are measured during the simulation process, which will be introduced in Chapter 4. Their derivatives with respect to temperature T will also be calculated using Equation 2.18 during simulations. These quantities and their thermal fluctuation quantities are of great importance to identify the conformational phases and phase transitions. The mathematical formulas for calculating them are listed below:

1. **Energy and specific heat**

   E is calculated using Equation 3.1 and the C_V is calculated using Equation 2.17. For future comparisons between polymers with different length, the specific heat, $c_V = \frac{1}{N}C_V$, is calculated and presented.

2. **The square of radius of gyration R^2 is**

$$R^2 = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_{\text{c.m.}})^2 \ , \tag{3.12}$$

   where

$$\mathbf{r}_{\text{c.m.}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i \ , \tag{3.13}$$

   is the enter of mass of the polymer and N is the number of monomers of the polymer.

3. x components of R^2 is

$$R_x^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{x}_{\text{c.m.}})^2 \ . \tag{3.14}$$

4. y components of R^2 is

$$R_y^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{y}_{\text{c.m.}})^2 \ . \tag{3.15}$$

5. z components of R^2 is

$$R_z^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{z}_i - \mathbf{z}_{\text{c.m.}})^2 \ . \tag{3.16}$$

6. z component of $\mathbf{r}_{\text{c.m.}}$ is

$$z = |\mathbf{z}_{\text{c.m.}}| = |\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i| \ . \tag{3.17}$$

7. **The number of monomer-monomer contacts N_{m} is calculated by**

$$N_m = \begin{cases} N_m + 1 \ , & \text{if } r_{i,j} \leq r_m \\ N_m \ , & \text{if } r_{i,j} > r_m \end{cases} \ , \tag{3.18}$$

where i and j are different indices of the monomers of the polymer and r_m is the monomer-monomer contact distance. When $r_{i,j} \leq r_m$, the i-th and j-th monomers are considered as to be in contact. The number of monomer-monomer contacts is the total number of all the contacted pairs. r_m is selected according to the system. Different values of r_m, e.g., $r_m = 1.05, 1.1, 1.15, 1.2, 1.3$, were tested to obtain a properly selected value which makes sure that only nearest monomers (first layer of the icosahedral core)

are counted. In this way N_m will be sensitive to possible "phase" transitions caused by the monomer-monomer interactions. Finally $r_m = 1.1$ is selected for this simulation.

8. **The number of monomer-surface contacts** N_s is calculated by

$$N_s = \begin{cases} N_s + 1 \ , & \text{if } z_i \leq r_s \\ N_s \ , & \text{if } z_i > r_s \end{cases} \ , \tag{3.19}$$

where z_i is the z component of the i-th monomer and r_s is the monomer-surface contact distance. When $z_i < r_s$, the i-th monomer is considered to be in contacted with the surface; the number of monomer-surface contacts is the total number of all the contacted monomers. Similarly, r_s also needs to be selected accordingly. Different values of r_s, e.g., $r_s = 1.05, 1.1, 1.15, 1.2$, were tested to obtain a proper value which makes sure that only monomers of the nearest layer along the substrate are counted. In this way N_s will be sensitive of possible phase transitions caused by the monomer-substrate interactions. Finally $r_s = 1.05$ is selected.

All these eight quantities and their thermal fluctuations are recorded to a data file every 100 Monte Carlo sweeps. The used Monte Carlo simulation methodologies will be discussed in the next chapter.

Chapter 4

Simulation Methods

As previously discussed in Chapter 2, to discern phase transitions, the macroscopic canonical quantities will be measured in the thermal equilibrium. In a practical physical experiment, an average of multiple measurements of an observable O will be carried out N times. The average, $\langle O \rangle = 1/N \sum_i^N O_i$, will be considered as the expected value of that observable O with an error $\sigma/\sqrt{N}$, where $\sigma = \sqrt{\langle O^2 \rangle - \langle O \rangle^2}$ is the standard derivation and σ^2 is the variance. Because these measurements are independent, the measured values of the observable are distributed according to a normal distribution.

Similar to Equation 2.14, if the probability distribution $p(O)$ of O is known, then $\langle O \rangle = \sum_O Op(O)$. If the probability of a microstate s is known and the value of the observable O_s can be measured, then

$$\langle O \rangle = \sum_O Op(O) = \sum_s O_s p_s \ , \tag{4.1}$$

where p_s is given in Equation 2.10. Therefore, the key to successful simulations is to find the probability distribution of all microstates of an ensemble. Given sufficient time[1], methods that can efficiently sample all the possible microstates, a property known as ergodicity, are

---

[1]The time is referring to real time process.

of importance. In other words, the ergodic hypothesis has to be satisfied [70]. Markov Chain Monte Carlo (MCMC) can provide such methods.

4.1 Markov Chain and Master Equation

During a Markov Chain Monte Carlo simulation, a set of microstates $(s_1, s_2, s_3, \dots)$ can be generated according to the probability distribution in the thermal equilibrium at the stochastic simulation time steps $(t_1, t_2, t_3 \dots)$. Due to ergodicity, the average of the observable $\langle O \rangle$ over a large number of stochastic time steps is

$$\overline{O} = \frac{1}{N} \sum_{i=1}^{N} O(s_i) \approx \langle O \rangle \ , \tag{4.2}$$

where $\langle O \rangle$ is the statistical ensemble average in Equation 4.1 and N, the number of Monte Carlo time steps, needs to be sufficiently large. Define $p_i(t)$ as the probability of the system at the microstate s_i at simulation time step t. After a certain number of time steps Δt, the system could enter a new microstate s_j. The temporal evolution of the probability of microstate s_i to any another new microstate s_j is

$$\frac{\Delta p_i(t)}{\Delta t} = \sum_j p_j(t) T_{s_j \to s_i} - \sum_j p_i(t) T_{s_i \to s_j} \ , \tag{4.3}$$

where $p_j(t)$ is the probability of the system at microstate s_j at time t and $T_{s_j \to s_i}$ is the transition rate from microstate j to i. Equation 4.3 is the master equation that describes the conservation of probability. In equilibrium, the ensemble probability distribution $p_i(t)$ is stationary; in other words, $\Delta p_i(t)/\Delta t = 0$. Therefore Equation 4.3 yields

$$\sum_j p_j T_{s_j \to s_i} = \sum_j p_i T_{s_i \to s_j} \ , \tag{4.4}$$

which is the global balance condition. To assure solutions in equilibrium, a stricter condition can be introduced as

$$p_j T_{s_j \to s_i} = p_i T_{s_i \to s_j} \ , \tag{4.5}$$

which ensures the probabilities of pairwise exchanges are equal.

$T_{s_i \to s_j}$ can be separated into

$$T_{s_i \to s_j} = S_{s_i \to s_j} A_{s_i \to s_j} \ , \tag{4.6}$$

where $S_{s_i \to s_j}$ is the probability of proposing a trial move from s_i to s_j and $A_{s_i \to s_j}$ is the probability of accepting this proposed update. Substituting Equation 4.6 into Equation 4.5 gives

$$\frac{T_{s_i \to s_j}}{T_{s_j \to s_i}} = \frac{S_{s_i \to s_j} A_{s_i \to s_j}}{S_{s_j \to s_i} A_{s_j \to s_i}} = \frac{p_j}{p_i} \ . \tag{4.7}$$

The ratio of the forward and backward mutual selection probabilities $S_{s_i \to s_j}/S_{s_j \to s_i}$ depends on the update schemes of the system in a Monte Carlo simulation. In most local displacement update schemes, the ratio equals unity; for some global update schemes, the ratio may not be unity [71].

The advantage of Markov Chain Monte Carlo is that only the information of the probability distribution of the ensemble in equilibrium and the information of one previous microstate is needed to generate the next microstate. During the temporal evolution, the more data are collected as simulation time increases, the better estimations of the results can be obtained. A long run of the simulation is needed to set the "correct" expected values of the problem in equilibrium to remove the time correlation in Markov Chain data. Therefore, a good balance between the total simulation runtime and data qualities can be well estimated for future simulations.

4.2   Single Displacement Update

The single displacement update scheme is carried out throughout this dissertation. It is simple, satisfies ergodicity and has equal forward and backward mutual selection probabilities between two microstates. This means $S_{s_i \to s_j}/S_{s_j \to s_i} = 1$ and Equation 4.7 reduces to

$$\frac{p_j}{p_i} = \frac{A_{s_i \to s_j}}{A_{s_j \to s_i}} \ .$$

$$(4.8)$$

Given a homopolymer with N monomers, its current conformation is $\mathbf{X} = \{\mathbf{r}_1, \mathbf{r}_2,$ $\dots, \mathbf{r}_i, \dots, \mathbf{r}_N\}$. In a single displacement update, one of the monomers (the i-th monomer, $i \in [0, N]$) is selected randomly using a pseudo random number generator[2]. Then apply a random displacement update to this i-th monomer with

$$\mathbf{r}_i = \mathbf{r}_i + \Delta\mathbf{r} \ ,$$

$$(4.9)$$

where $\Delta\mathbf{r} = \Delta\mathbf{x}, \Delta\mathbf{y}$ and $\Delta\mathbf{z}$. The updating lengths[3] of $\Delta\mathbf{x}, \Delta\mathbf{y}$ and $\Delta\mathbf{z}$ are selected uniformly from $[-0.3, 0.3]$ throughout this dissertation. Recall that r is scaled by $r_o = 1$, which is defined in Equation 3.18. Therefore, the resultant updated conformation is $\mathbf{X}^{\text{updated}} = \{\mathbf{r}_1,$ $\mathbf{r}_2, \dots, \mathbf{r}_i^{\text{updated}}, \dots, \mathbf{r}_N\}$.

4.3   Metropolis Sampling

All sampling schemes that satisfy detailed balance (Equation 4.8, in general Equation 4.7) and the ergodicity are valid in an MCMC simulation. From Chapter 2, in thermal equilib-

---

[2]A nice discussion about the "art" of random number generator (RNG) can be found in [70] and the RNG used in this simulation is a linear congruential generator (LCG) with an enormous multiplier [72].

[3]$[-0.3, 0.3]$ is selected based on the balance between the computational efficiency and the acceptance probability (around 35% at $T = 0.1$) at low and high temperatures. Various selections of $\Delta\mathbf{r}$ for different temperatures of Metropolis or Parallel tempering, and for different energy bins of Wang-Landau can be found in [71].

rium, the probability of a microstate s_i is $p_i = e^{-E_i/k_B T}/Z$ and the probability of a microstate s_j is $p_j = e^{-E_j/k_B T}/Z$. Although the partition function Z is unknown, only the ratio of p_i and p_j matters. Substitute p_i and p_j into Equation 4.9 and set one of the accepted probability at to maximum value of 1 (unity), Equation 4.8 reduces to

$$A_{s_i \to s_j} = min\left(1, e^{-\frac{E_i - E_j}{k_B T}}\right), \qquad (4.10)$$

where E_i and E_j correspond to the conformational potential energies given in Equation 3.1 [73, 7, 70]. The Metropolis implementation can be described in the following scheme:

Metropolis Sampling Simulation Scheme

1. start from a random conformation X of a microstate;

2. measure the desired energetic and structural quantities based on current conformation;

3. randomly choose the i-th monomer and propose a conformational update;

4. calculate the energy difference between the current conformation and the proposed conformation;

5. generate a random number $p \in (0, 1)$;

6. if $p < e^{-\frac{E_i - E_j}{k_B T}}$, accept the update; if not reverse the update;

7. repeat steps $2 - 6$.

The simulation time should be long enough so that $\overline{O} = \frac{1}{N}\sum_{i=1}^{N} O(s_i) \approx \langle O \rangle = \sum_s p_s O_s$ can be satisfied. At the beginning of the time series, the system will evolve towards the thermal equilibrium [74]. Since none of the simulations runs infinitely long, $\overline{O}$ will scatter around

the theoretical value of $\langle O \rangle$ with a standard deviation

$$\sigma_{\overline{O}} = \sqrt{\langle [\overline{O} - \langle \overline{O} \rangle]^2 \rangle} = \sqrt{\langle \overline{O}^2 \rangle - \langle \overline{O} \rangle^2} \; , \tag{4.11}$$

where $\sigma_{\overline{O}}$ can be obtained from the distribution of an individual measurement O. For all uncorrelated measurements,

$$\sigma_{\overline{O}}^2 = \frac{\sigma_{O_j}^2}{N} \; , \tag{4.12}$$

where $\sigma_{O_j}^2 = \langle O_j^2 \rangle - \langle O_j \rangle^2$ is the population variance of the individual data O_j, which is unknown for a practical calculation. Therefore the statistical error decreases by a factor of $1/\sqrt{N}$.

Although the population variance cannot be obtained, the sampling variance can be measured

$$\hat{\sigma}_O^2 = \left\langle \overline{(O - \overline{O})^2} \right\rangle = \sigma_O{}^2 \left(1 - \frac{1}{N} \right) = \sigma_O{}^2 \left(\frac{N-1}{N} \right) \; . \tag{4.13}$$

The sampling variance $\hat{\sigma}_O^2$ is therefore systematically smaller than the real variance $\sigma_O{}^2$, which is the systematic bias for uncorrelated measurements[4] [75]. The bias corrected error estimator for $\overline{O}$ is

$$\mathrm{err}_{\overline{O}}^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} (O_i - \overline{O})^2 \; . \tag{4.14}$$

For a correlated data set, Equation 4.14 becomes

$$\mathrm{err}_{\overline{O}}^2 = \frac{1}{N(N_{\mathrm{eff}}-1)} \sum_{i=1}^{N} (O_i - \overline{O})^2 \; , \tag{4.15}$$

where N_{eff} is the number of uncorrelated measurements and $N_{\mathrm{eff}} < N$ [7].

---

[4]Detailed discussions can be found in [7, 75].

4.4   Jackknife Analysis

To reduce correlation of a Monte Carlo simulation, Jackknife (JK) binning analysis can be used [76, 77]. The idea of Jackknife binning method is to divide the total data set into K larger data bins, named the Jackknife bins. Instead of considering data only in one small data bin, Jackknife method considers all the data from all the bins except for one bin. Therefore, Jackknife method is also referred to as leave-one-out error cross validation in statistics [78]. The average of O of the k-th JK bin is

$$\overline{O}_k^J = \frac{N\overline{O} - N_k\overline{O}_k^{\mathrm{B}}}{N - N_k} \; , \tag{4.16}$$

where $\overline{O}_k^{\mathrm{B}}$ is the average O of the k-th small bin which has N_k data points. The variance of a JK average is

$$\mathrm{err}^2_{\overline{O}_{JK}} = \frac{K-1}{K} \sum_{k=1}^{K} (\overline{O}_k^J - \overline{O})^2 \; . \tag{4.17}$$

In such a process, the variance of a simulating time series data set can be reduced. Other methods such as bootstrap method can also be used to remove the correlation [79].

4.5   Parallel Tempering

In a Metropolis Monte Carlo simulation, microstates of the tails of the probability of a canonical system are rarely approached given a finite amount of simulation time. Due to high rejection probability from Boltzmann factor at low temperatures, the sampling efficiency may also decrease dramatically. At temperatures near a phase transition, on the other hand, the sampling efficiency may be good but the correlations and fluctuations may grow. At temperatures near a first-order phase transition, the samplings could be stuck in one of the free energy local minimums; at temperatures near a second-order phase transition,

Metropolis simulation may undergo a critical slowing down since the systems are globally correlated near a second-order phase transition. These disadvantages require more efficient simulation methods and parallel tempering is one such method.
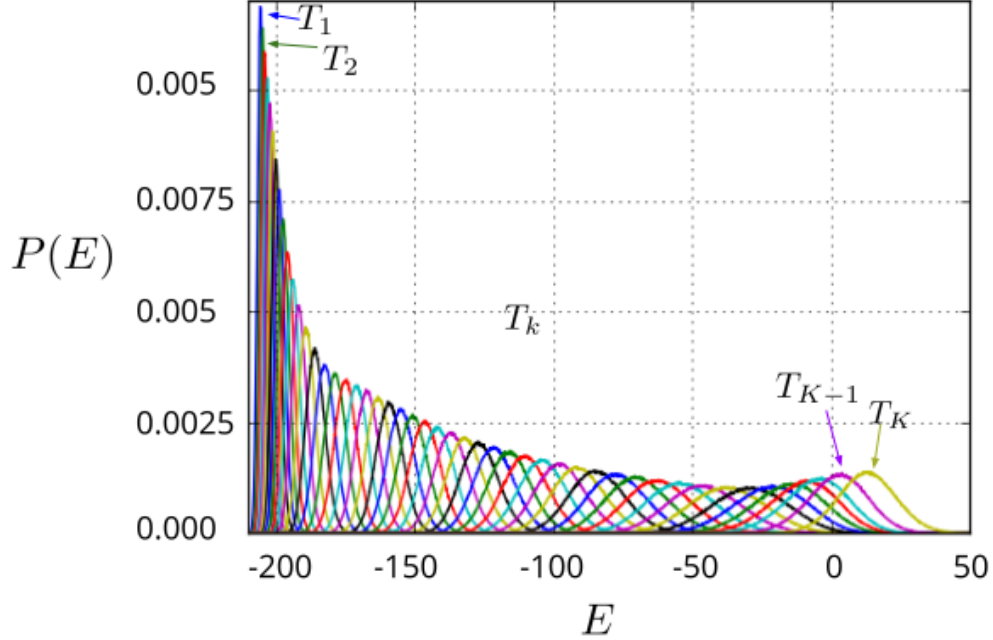

Figure 4.1: The energy distributions $P(E)$ of a set of 48 temperatures in parallel tempering of a 55-mer homopolymer. In this simulation, 48 temperature values are selected in the range of [0.1, 3.1] and their overlaps are sufficiently good to allow efficient configuration exchange attempts. For computational efficiency, the total number of selected temperatures can be reduced, if the overlaps of the energy distribution curves of adjacent temperatures are greater than 40%.

These disadvantages can be partially overcome by applying an artificially generalized ensemble [7]. In parallel tempering, a set of canonical ensembles at various temperatures, $T_1, T_2, \ldots, T_k, \ldots, T_K$, is simulated in parallel with Metropolis sampling. The set of the canonical ensembles constructs an aggregated canonical ensemble. The aggregated canonical ensemble probability is $Z_{\mathrm{PT}}(T_1, \ldots, T_K) = \prod_{k=1}^{K} Z_{can}(T_k)$, where $Z_{can}(T_k)$ is given in Equation 2.11. The probability distributions of all ensembles at various temperatures in a parallel tempering simulation are presented in Figure 4.1. Figure 4.1 is obtained from a free elastic off-lattice polymer of length 55. Overlap of the probability distributions between

adjacent replicas allows the possibility of configuration swaps. The sizes of the overlaps between probability distributions determine the selections of the temperature set. In this simulation 48 temperature values are selected in $[0.1, 3.1]$. When an overlap area is around $40\% - 70\%$, the simulation will be able to sample data efficiently [7, 21].

after a certain number of Monte Carlo time steps, the conformational replicas will attempt to exchange. The acceptance probability for each canonical ensemble is the Metropolis acceptance criterion defined in Equation 4.10. The probability of the exchange between neighboring temperatures T_k and T_{k+1} is

$$P(E_k, T_k; E_{k+1}, T_{k+1}) = \min\left(1, e^{(E_k - E_{k+1})\left(\frac{1}{k_{\mathrm{B}}T_k} - \frac{1}{k_{\mathrm{B}}T_{k+1}}\right)}\right). \tag{4.18}$$

The Parallel Tempering implementation [19, 20, 22] can be described in the following scheme.

Parallel Tempering Simulation Scheme

1. Select a set of temperatures $T_k \in [T_1, T_K]$ where $k = 1, 2, \ldots, K$ and K is the total number of temperatures;

2. Replicas of a polymer-substrate system at different temperatures will be simulated in parallel by the Metropolis method with the acceptance probability defined in Equation 4.10;

3. After a certain number of Monte Carlo sweeps N_{MCS}, replicas would attempt to exchange between neighboring temperatures T_k and T_{k+1} with the PT probability given in Equation 4.18;

4. Repeat steps 2-3.

A flowchart describing the PT simulation scheme is also given in Figure 4.2.
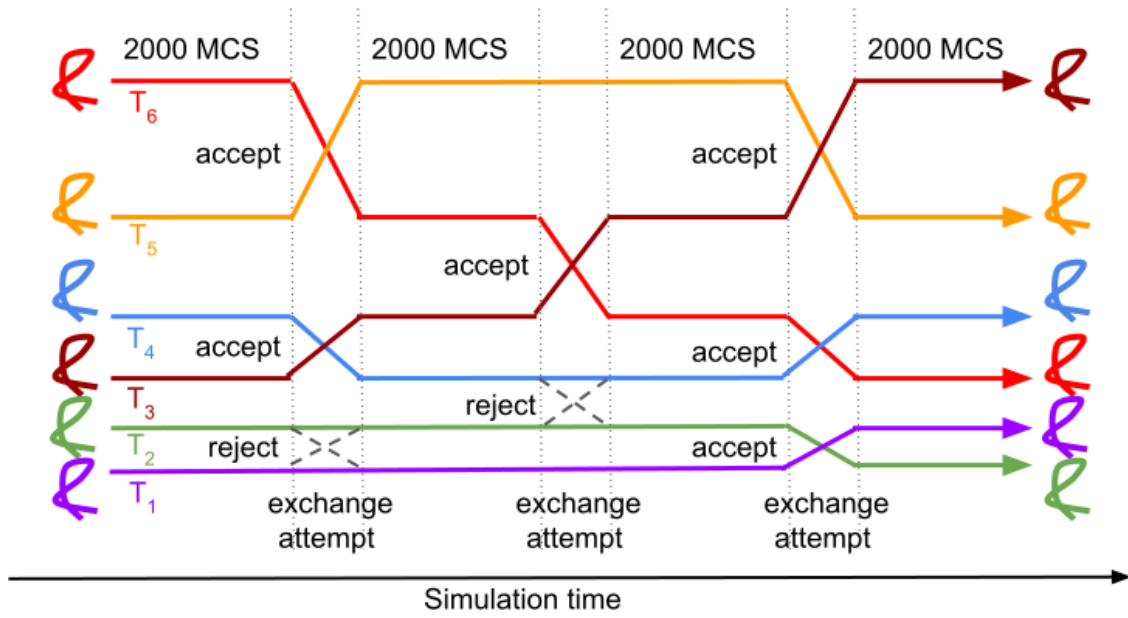
Figure 4.2: Sketch of parallel tempering, where $T_i < T_{i+1}$. The horizontal direction is the simulation time and the vertical direction is the selected temperature.

Chapter 5

Canonical Results for Polymer-Substrate Systems with a Fixed End-to-End Distance

5.1 Introduction

In previous chapters, the related physics background has been introduced, the coarse-grained model is given and the PT simulation methods and the data analysis methods are described. In this chapter, polymer-substrate systems with a fixed end-to-end distance of 25 with various temperatures and various substrate surface attraction strengths are studied. This work is inspired by the previous studies [67, 80, 81, 82] about polymer substrate interactions.

In the beginning of polymer research, the conformations of a polymer chain can be represented as all the possible paths of a random walk with N steps in a lattice with a fixed bond length [83, 84, 85, 86]. Later on, self-avoiding walk (SAW) has been applied to make the polymer model more realistic [87, 88, 89]. Afterwards, a monomer-monomer interaction with a short distance repulsive effect and a long range attractive effect has been used to represent

the non-bonded monomer-monomer interactions [84, 85]. This idea has been applied in this dissertation and the Lennard-Jones potential is used as the non-bonded potential.

Conventionally, the square of the radius of gyration $\langle R^2 \rangle$ has been used to analyze the polymer phases instead of just the radius of gyration $\langle R \rangle$ [84, 85]. This is because $\langle R^2 \rangle \propto N^{2\nu}$, where $\nu = 0.588 \pm 0.001$ for a SAW polymer in 3 dimensions [90].

When a free polymer approaches an energetically attractive substrate, polymer adsorption will happen [23, 67, 91, 92]. Under the competition between the monomer-monomer interactions and the monomer-surface interactions, a coil-to-globular transition in which polymer change from random coil to globular (Θ transition) will occur, as the temperature decreases.

As the temperature decreases further more, a freezing transition occurs, in which the polymer becomes crystalline or whose exact shape is also influenced by the qualities of the substrate e.g., the value of the attractive strength [93].

Polymers with one end grafted to a substrate are also studied [67, 94]. Similarly, one end grafted polymers also could experience a Θ transition, then instead of an adsorption transition, a wetting transition might occur for long chain polymers [24]; then again a freezing transition could happen at low temperatures.

Previous studies of free polymer adsorption and the interactions between the one end grafted polymer and the substrate motivate the study of the phases and phase transitions of a finite size polymer with its two ends grafted into an attractive substrate. A sketch of such a model is given in Figure 3.4. One of the ends is anchored at the origin and the end-to-end distance is along the x axis. The substrate occupies the half space where $z < 0$. The substrate surface attraction strength ϵ_s is defined in Equation 3.11. The energy of such a system is given in Equation 3.1. The simulation scheme will follow the PT simulation procedure given in Section 4.5.

With both of the two ends attached, the conformational entropy will reduce noticeably and the energy of the interactions between the polymer and the substrate/surface will increase dramatically comparing with the polymer adsorption and the one end grafted polymer models. With this new constraint that the two ends of a polymer are attached on the attractive surface, the changes of the conformational phases of such a polymer are of great interest. In this chapter, results of the canonical statistical analysis for such a system at various T and ϵ_{s} will be presented, the conformational phases will be discussed and the T vs. ϵ_{s} phase diagram will be constructed.

5.1.1 Specific Parameters of the Model and Simulations

The model parameters are:

1. $N = 40$ is the length of the homopolymer;

2. $D = 25$ is the distance of the two grafted ends [1];

3. ϵ_s [2] is the attraction strength and $\epsilon_{\mathrm{s}} \in [0, 5]$ for $D = 25$;

4. T [3] is the temperature and $T \in [0.1, 5.6]$.

---

[1]D is a unitless quantity, because it is measured by the unitless length scale r_0, where $r_0 = 1 = \frac{r_{0-\mathrm{actual}}}{r_{0-\mathrm{actual}}}$ and $r_{0-\mathrm{actual}} = \frac{\sigma_{\mathrm{actual}}}{\sqrt[6]{2}}$. σ_{actual} is the actual van del Waals radius, which can be measured for a specific polymer. Therefore, the length scale r_0 is a unitless quantity and the results of this simulation can be applied to different practical systems if their σ_{actual} values are measured.

[2]ϵ_s is also a unitless quantity, since it is measured by the energy scale ϵ_{LJ}, where $\epsilon_{\mathrm{LJ}} = 1 = \frac{\epsilon_{\mathrm{LJ-actual}}}{\epsilon_{\mathrm{LJ-actual}}}$ and $\epsilon_{\mathrm{LJ-actual}}$ is the actual minimum value of the LJ potential, which can be measured for a specific polymer. Therefore, the energy scale ϵ_{LJ} is a unitless quantity and the results of this simulation can be applied to different practical systems if their $\epsilon_{\mathrm{LJ-actual}}$ values are measured.

[3]T is also a unitless quantity, since it is defined as $T = \frac{k_{\mathrm{B}} T_{\mathrm{actual}}}{\epsilon_{\mathrm{LJ-actual}}}$, where k_{B} is the Boltzmann constant, T_{actual} is the actual thermodynamic temperature of the system and $\epsilon_{\mathrm{LJ-actual}}$ is the actual minimum energy of the LJ potential, which can be measured for a specific polymer. Therefore, T is a unitless quantity and the results of this simulation can be applied to different practical systems if their actual thermodynamic temperatures are given and their $\epsilon_{\mathrm{LJ-actual}}$ values are measured.

The PT simulation schemes and parameters are:

1. $K = 32$ is the number of temperatures, where 32 is obtained by the criterion given in Section 4.5;

2. Replicas are assigned to the K different temperatures;

3. In each simulation process, the seeds of the random number generator are assigned differently;

4. Eight energetic and structural quantities are measured for identification of phases transitions;

5. Jackknife analysis is used for error estimation;

6. Various ϵ_s values within the range of $[0, 5]$ are used for discerning phases and phase transitions at different temperatures;

7. A simulated annealing method [95, 96] is used to find the lowest energy structures.

Here are the computational resources:

1. Programs written in C++, MPI Package was used for parallel tempering;

2. Environment: Red Hat Enterprise Linux Server release 6.10 (Santiago);

3. 32 computational cores, Intel(R) Xeon(R) CPU E5-2697 v2 @ 2.70GHz;

4. 716.1 GB total memory;

5. Data analysis: Jupyter notebook via Python3;

6. Plot tool: Gnuplot 5.0;

When conducting simulations using the above tools, the average runtime is around $2 \sim 3$ days and the data analysis procedures could also take days for a set of parameters. The total number of the computational CPU core-years for implementing all the simulations in this project is: $2 \sim 3$ (days) $\times 32$ (cores) $\times 21$ (different ϵ_s values) $\approx 3.7 \sim 5.6$ core-years, in addition to the time for data analysis.

5.2 Results

In this section, a showcase result will be presented, followed by the thermodynamic and structural results for the polymer-substrate systems. Finally a phase diagram parameterized by temperature and the attractive strength is constructed.

5.2.1 Results of Polymer-Substrate Systems with $D = 25$ and $\epsilon_s = 0$ with Various T: a Showcase

A showcase result of thermodynamic and structural properties for a polymer-substrate systems with $D = 25$ and $\epsilon_s = 0$ with various T is represented in Figure 5.1. The top figure shows the results of c_V, the derivatives of the square radius of gyration R^2 and its x component R_x^2. The x, y components of the derivatives of the square radius of gyration R^2 and the z component of the center of mass are presented in the middle figure. The bottom figure illustrates the derivatives of the number of monomer-monomer contacts $d \langle N_m \rangle / dT$ and the number of monomer-monomer surface contacts $d \langle N_s \rangle / dT$.

The peaks of the c_V curve at $T \approx 0.16$ and 0.38 indicate dramatic energetic changes and can be used to discern "phase" transitions. The peak at $T \approx 0.16$ indicates the freezing transition which can be found in both non-grafted and singly grafted polymers [67]. The other peak at $T \approx 0.38$ indicates a Θ transition in which the polymer will collapse from a random coil structure into a globular structure. Such transitions are noticeable for non-
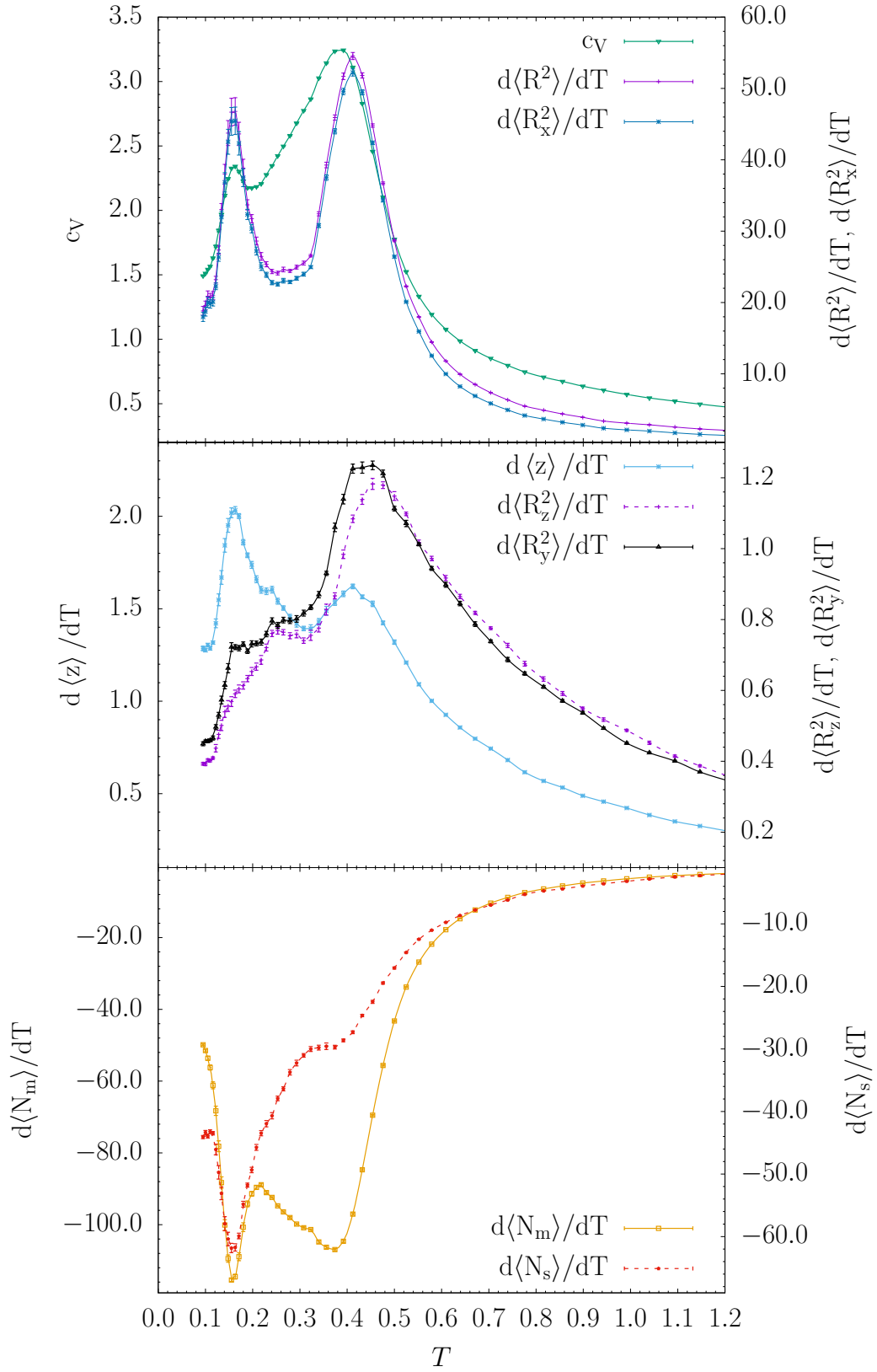
Figure 5.1: Thermodynamic and structural properties for polymer-substrate systems with $D = 25$ and $\epsilon_s = 0$ with various T. Some error bars are within the size of the data points.

grafted polymers but unnoticeable for singly grafted polymers [67]. However, the peaks indicating the Θ transition of c_V curves will reduce to a shoulder or even disappear as ϵ_s increases. The different behaviors of these surface interactions are of interest among free polymers, singly grafted and doubly grafted polymers.

Together with the observations of the peaks of other curves, these two phase transition bands can be confirmed, e.g., the curves of the derivatives of the square radius of gyration $\langle R^2 \rangle$ and its x, y, z components. In this showcase example, since $\epsilon_s = 0$, the surface of the substrate only acts as a hard wall. Since the distance between the two grafted ends of the polymer, D, is along the x axis as shown in Figure 3.4 and D is fairly large compared to the polymer length $N = 40$, the change in $\langle R_x^2 \rangle$ dominates the change in the square radius of gyration $\langle R^2 \rangle$ and leads to the changes along y and z axes. Since the peaks of the c_V curve match the peaks of the curves of $\langle R^2 \rangle$ and its x, y, z components, this indicates that the conformational changes of the polymer leads to energy changes during the Θ and freezing transitions.

There is also a "shoulder" pattern of the c_V curve around $T \approx 0.33$ if one observes carefully. That shoulder-like signal matches a "corner" pattern of the fluctuation of the $\langle N_m \rangle$ curve. This means the topology of the monomers is restructured without changing the energy dramatically. These signals indicate a conformational entropic transition around $T \approx 0.33$ without causing a dramatic energetic change. This transition is similar to the adsorption transition of a free polymer and substrate system [11, 67]. It is also similar to a wetting transition of a singly grafted polymer and substrate system [67, 93]. However, for a doubly graft substrate system, the melting transition may be a better term to describe this transition, because during this transition as T increases, the polymer transitions from a globular chain with an spherical droplet-like structure to a pure 3D globular structure without the droplet. In such a transition, the average distances between various monomer pairs increase. It looks like the droplet is melted during this transition.

51

In summary, with information of these peak-like, valley-like, shoulder-like and corner-like signals of the curves of energetic and structural quantities (and the curves of the fluctuations of these quantities), structural transitions of the polymer-substrate systems at temperatures around the interesting ranges are researched, three types of structural transitions are identified and four types of structures are found and attached in Figure 5.2, which shows three transitions and four conformational phases. Note that, for simplicity, the prime notation is used for the temperature derivative of an observable, e.g., $d\langle R^2\rangle/dT = \langle R^2\rangle'$.
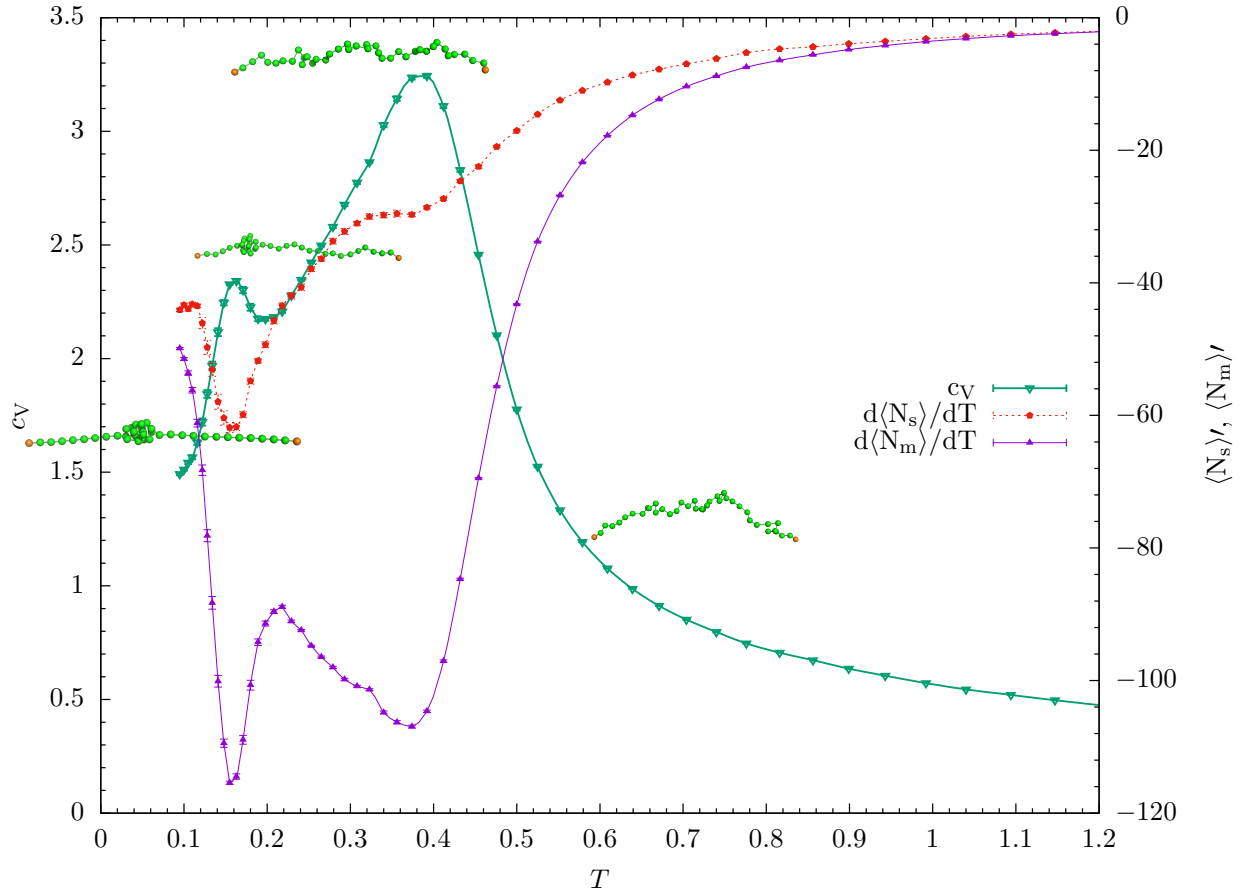
Figure 5.2: c_{V}, $d\langle N_s\rangle/dT$ and $d\langle N_m\rangle/dT$ curves of polymer-substrate systems with $D = 25$ and $\epsilon_{\mathrm{s}} = 0$ with four different conformational structures. See the text for an explanation.

In Figure 5.2, when $T < 0.16$, the dominating structures are compact structures of an icosahedral core with two straight strands attached. Since $\epsilon_{\mathrm{s}} = 0$, this icosahedral core

minimizes the energy of this system [97, 98, 99]. Such structures can be identified in many viral coat protein structures [100]. On the other hand, if $\epsilon_s \neq 0$, the minimum energy of the polymer-substrate system could form an fcc crystalline structure [101]. At around $T \approx 0.16$, the polymer goes through the freezing transition. When $0.16 < T < 0.33$, as temperature increases, the energy also increases in the system; therefore, the dominating structures will start to melt and the monomers will have more freedom to move in space. The polymer is in its globular phase with an spherical droplet structure. The icosahedral core formed at low temperatures is still present in this temperature range but becomes a droplet. The droplet will oscillate forward and backward along the strands between the two anchored points. This phase only occurs when ϵ_s is small and as ϵ_s increases, this phase becomes a globular polymer chain with a semi-spherical droplet phase. At around $T \approx 0.33$, the polymer goes through the melting transition. As the temperature range slightly increases to $0.33 < T < 0.38$, a small amount energy increase can melt the spherical droplet. The energy increase will not change the c_V curve dramatically but will result in a strong signal of the fluctuation of $\langle N_m \rangle$ curve. This indicates that the conformational phase change is led by re-arrangement of the monomer-monomer topology within the droplet. At around $T \approx 0.38$, the polymer goes through the Θ transition. When $T > 0.38$, the structure becomes a random coil gas-like structure, which can move in space more actively than the globular liquid-like chain.

In summary, the signals at $T \approx 0.16$ indicate the freezing transition separating the compact crystalline phase and the globular phase; the signals at $T \approx 0.33$ indicate the melting transition separating the globular phase with a droplet and the globular phase without a droplet; the signals at $T \approx 0.38$ indicate the Θ transition separating the globular phase and the random coil phase.

Applying a similar data analysis procedure, the conformational phases and phase transitions of one simulation of a given set of D and ϵ_s values can be found. In the following sections, the results of various ϵ_s values with different T values will be provided. Only inter-

esting signals lead to critical changes in the phase diagram will be explained briefly in the following section.

5.2.2 Canonical Results

In this section, canonical results of polymer-substrate systems with $D = 25$ and various ϵ_s values with different T values will be given. A total of 21 different $\epsilon_\text{s} \in [0.0, 5.0]$ values are used in simulations and their results are plotted in 8 figures.

Initially the ϵ_s values are chosen with an interval of 0.5, e.g., $\epsilon_\text{s} = 0.0, 0.5, 1.0, \ldots, 5.0$. With those data, applying the methods given in Section 5.2.1, a primary phase diagram can be constructed. According to the primary phase diagram, more ϵ_s values will need to be selected to describe the phase diagram in more detail. With all the canonical results, the primary phase diagram can be improved, and finally the conformational phase diagram of a polymer-substrate system with $D = 25$ and various ϵ_s with different T can be constructed in the end of the chapter.

The c_V vs. T curves of the 21 various ϵ_s values are plotted in Figure 5.3. Two peaks around $T \approx 0.16$ and $T \approx 0.38$ are notable, which identify the freezing and Θ transitions accompanying with dramatic energetic changes. There are also less noticeable shoulders, which could imply melting transitions with non-dramatic energetic changes. Those shoulder patterns need to be compared with patterns of curves of other canonical quantities to obtain credible information about possible phase transitions.

The transition temperature of the freezing phase transition (T_fr) is at around $T \approx 0.16$ when $\epsilon_\text{s} = 0.0$. It decreases slightly, as ϵ_s increases. This is understandable that the energy of monomer-substrate interaction increases as ϵ_s increases.

The transition temperature of the melting transition (T_m) is at $T \approx 0.33$ when $\epsilon_\text{s} = 0.0$. It decreases slightly, as ϵ_s increase; eventually, it disappears when $\epsilon_\text{s} \approx 2.4$. The reason is that the energy of monomer-substrate is larger than the energy of the monomer-monomer
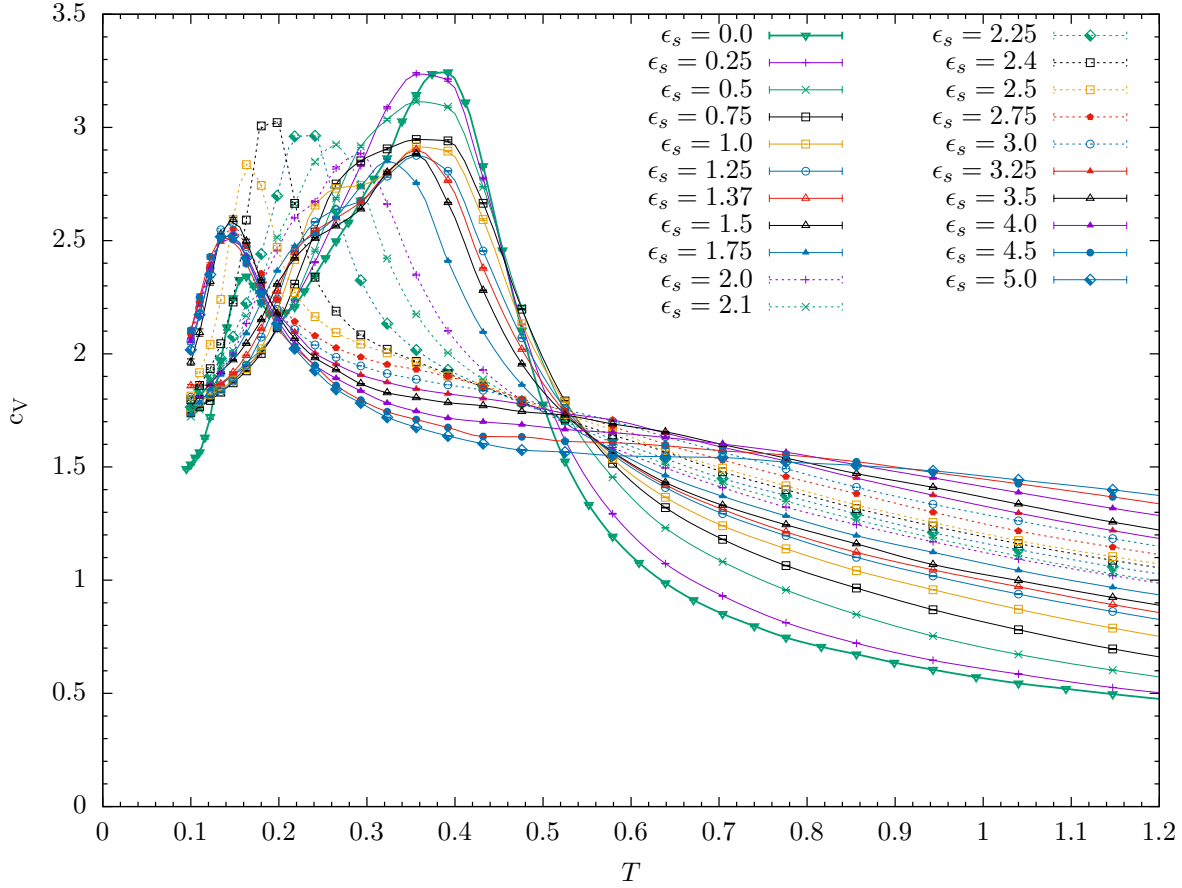
54

Figure 5.3: c_V curves of polymer-substrate systems with $D = 25$ and various ϵ_s and T values.

interaction and the formations of an icosahedral core structure will be suppressed when $\epsilon_s > 2.4$. The $\langle N_m \rangle'$ and $\langle N_s \rangle'$ curves provide the competition between the tend of forming an icosahedral core structure and adsorption along the substrate surface.

In Figure 5.3, when $0.25 \leq \epsilon_s \leq 1.75$, the location of peak, initially at $T = 0.38$, will keep moving to the left as ϵ_s increases. There are also shoulders on these curves. The shoulders will become less noticeable as ϵ_s increases. As ϵ_s is increased further, the shoulders disappear and the peak moves towards $T = 0.15$, which indicates the phase transition between compact crystalline phase and liquid-like globular phase.

The $d\langle R^2\rangle/dT$ vs. T curves of the 21 various ϵ_s values are plotted in Figure 5.4. Two peaks around $T = 0.15$ and $T = 0.42$ of the curves are notable, which support the transition temperatures found in the c_V curves.
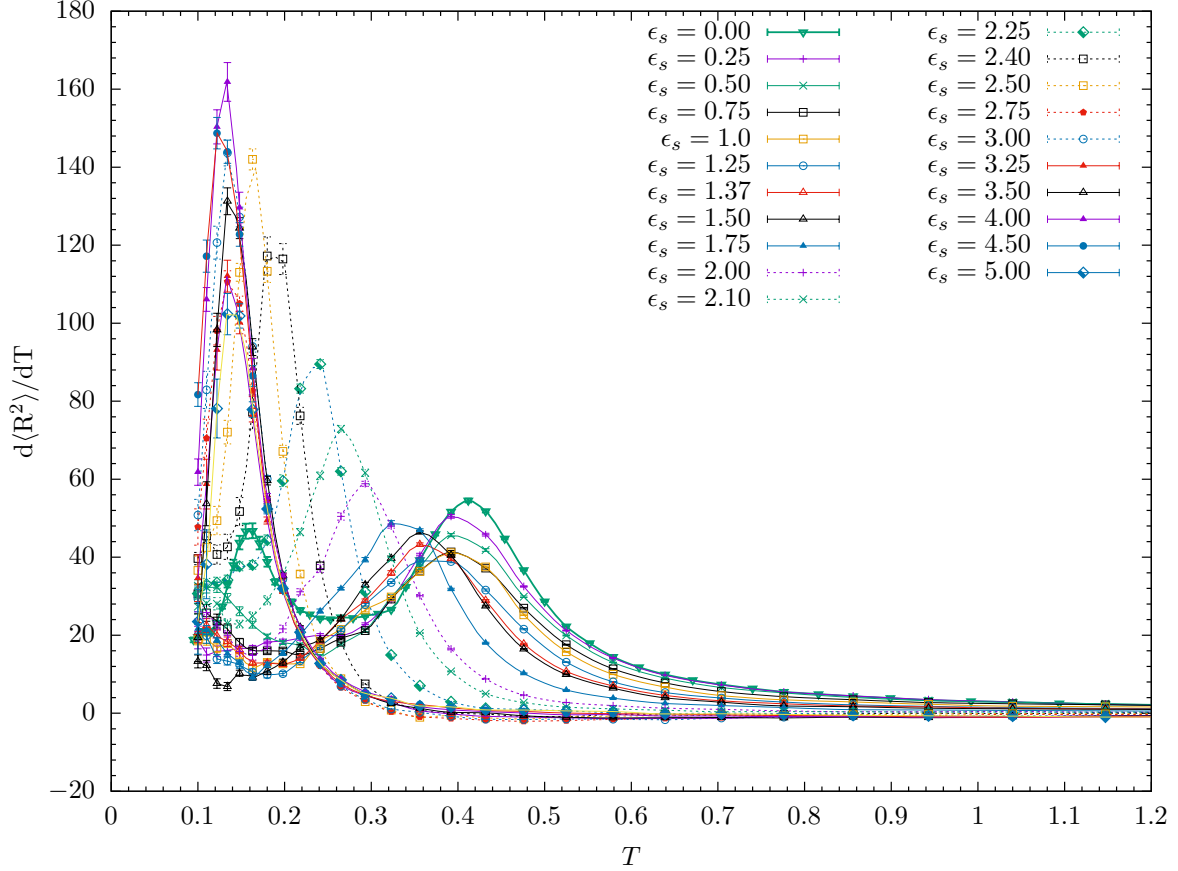
Figure 5.4: $d\langle R^2\rangle/dT$ vs. T curves of 21 various ϵ_s values.

Note that when $\epsilon_s \neq 0$, the double peak patterns do not disappear. This gives more noticeable signals for the two conformational phase transitions, Θ and freezing transitions, comparing with the c_V curves. However, as ϵ_s increases, the required energy needed for a phase transition will decrease and there is no dramatic energy change. This causes the c_V curve to reduce to one peak as ϵ_s increases.

The $d \langle R_{\mathrm{x}}^2 \rangle /dT$ vs. T curves of 21 various ϵ_{s} values are plotted in Figure 5.5. Two peaks around $T = 0.15$ and $T = 0.42$ of the curves are also noticeable, which support the phase transitions found in the c_{V} and $d \langle R^2 \rangle /dT$ curves.
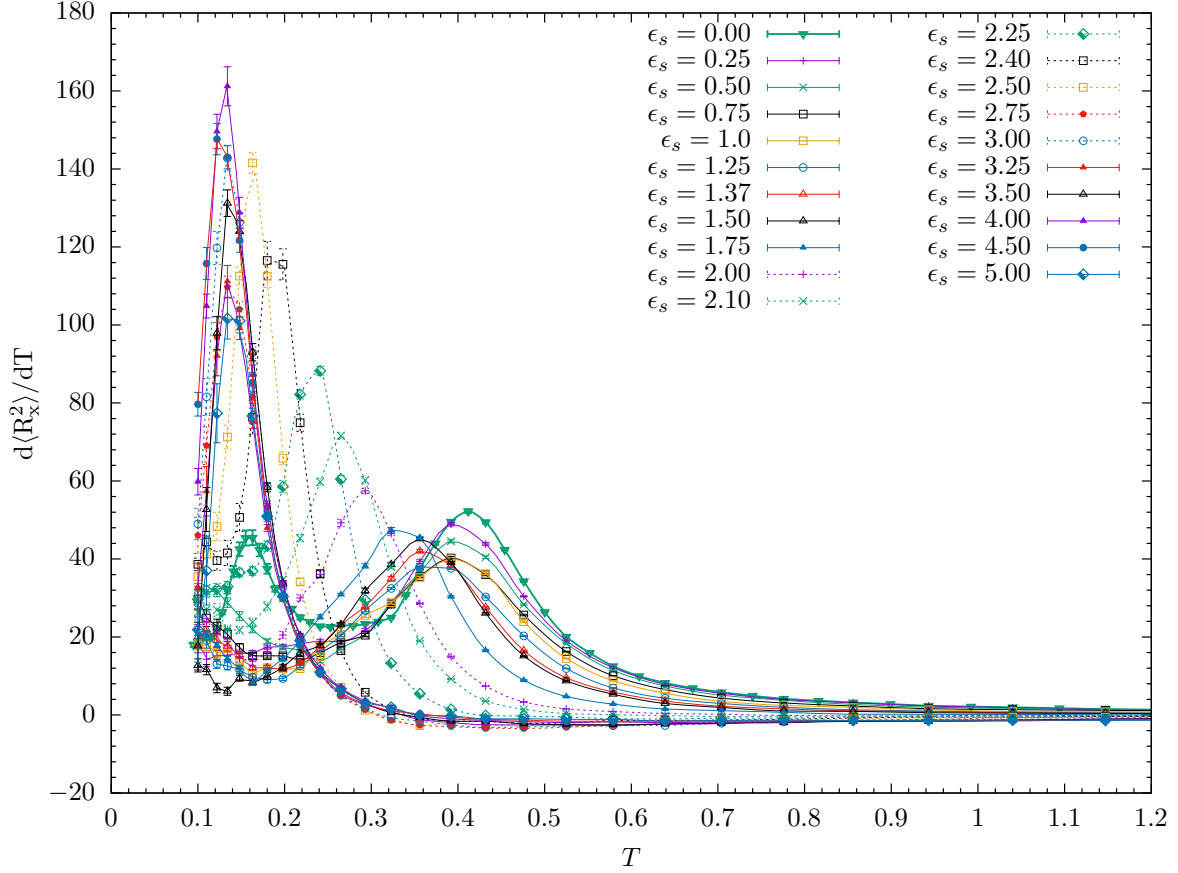
Figure 5.5: $d \langle R_{\mathrm{x}}^2 \rangle /dT$ vs. T curves of 21 various ϵ_{s} values.

Note that the transition temperatures of the curves of $d \langle R_{\mathrm{x}}^2 \rangle /dT$ correspond to the transition temperatures in Figure 5.4, because $D = 25$ is a fairly large value compared to the polymer length of 40.

The $d\left\langle R_\mathrm{y}^2\right\rangle/dT$ vs. T curves of 21 various ϵ_s values are plotted in Figure 5.6.
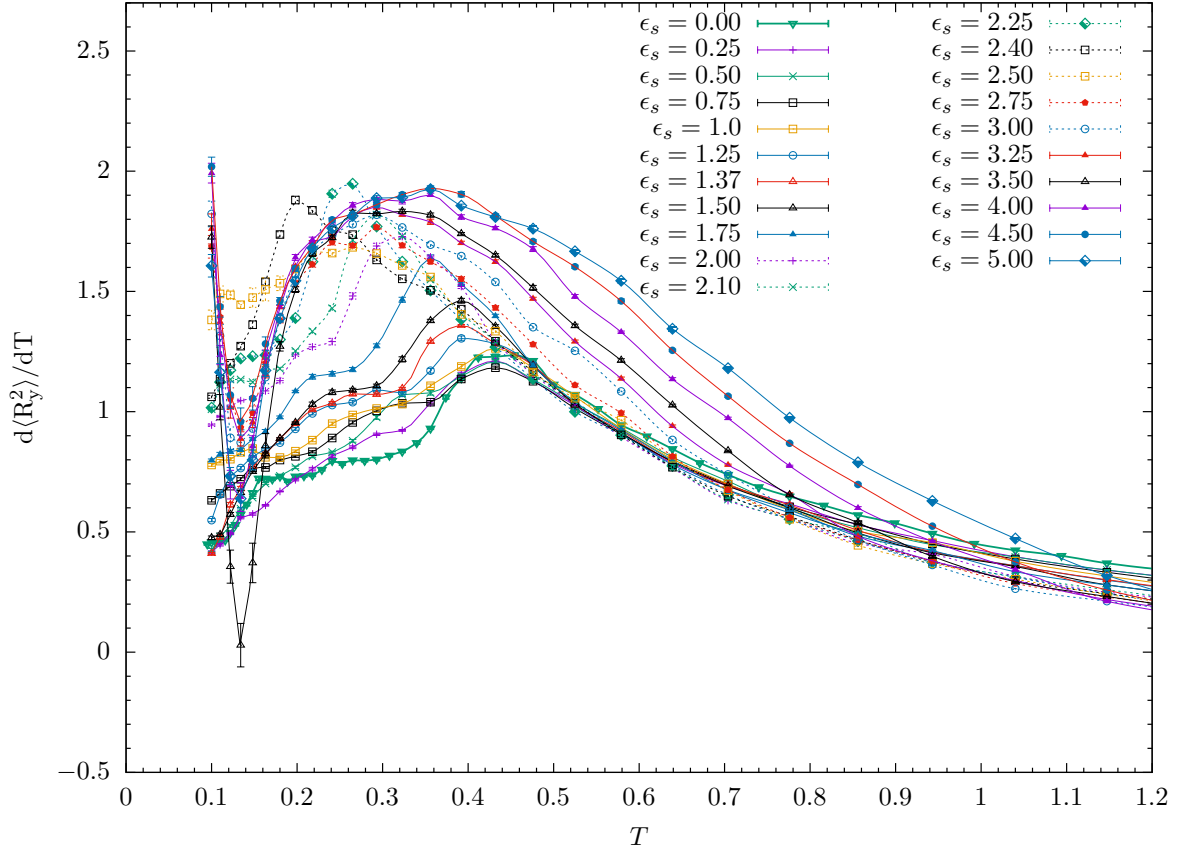
Figure 5.6: The $d\left\langle R_\mathrm{y}^2\right\rangle/dT$ vs. T curves of 21 various ϵ_s values.

Note that there are valleys instead of peaks at around $T \approx 0.15$. This is because when the compact crystalline phase with an icosahedral core transitions to the liquid-like globular conformational phase, the radius of gyration along the y direction will decrease. This will lead to the increase of $\left\langle R^2 \right\rangle$ along the x direction.

The $d\left\langle R_{\mathrm{z}}^2\right\rangle/dT$ vs. T curves of 21 various ϵ_{s} values values are plotted in Figure 5.7. The $d\left\langle R_{\mathrm{z}}^2\right\rangle/dT$ curves show more interesting changes than $d\left\langle R^2\right\rangle/dT$, $d\left\langle R_{\mathrm{x}}^2\right\rangle/dT$ and $d\left\langle R_{\mathrm{y}}^2\right\rangle/dT$ curves as ϵ_{s} increases.
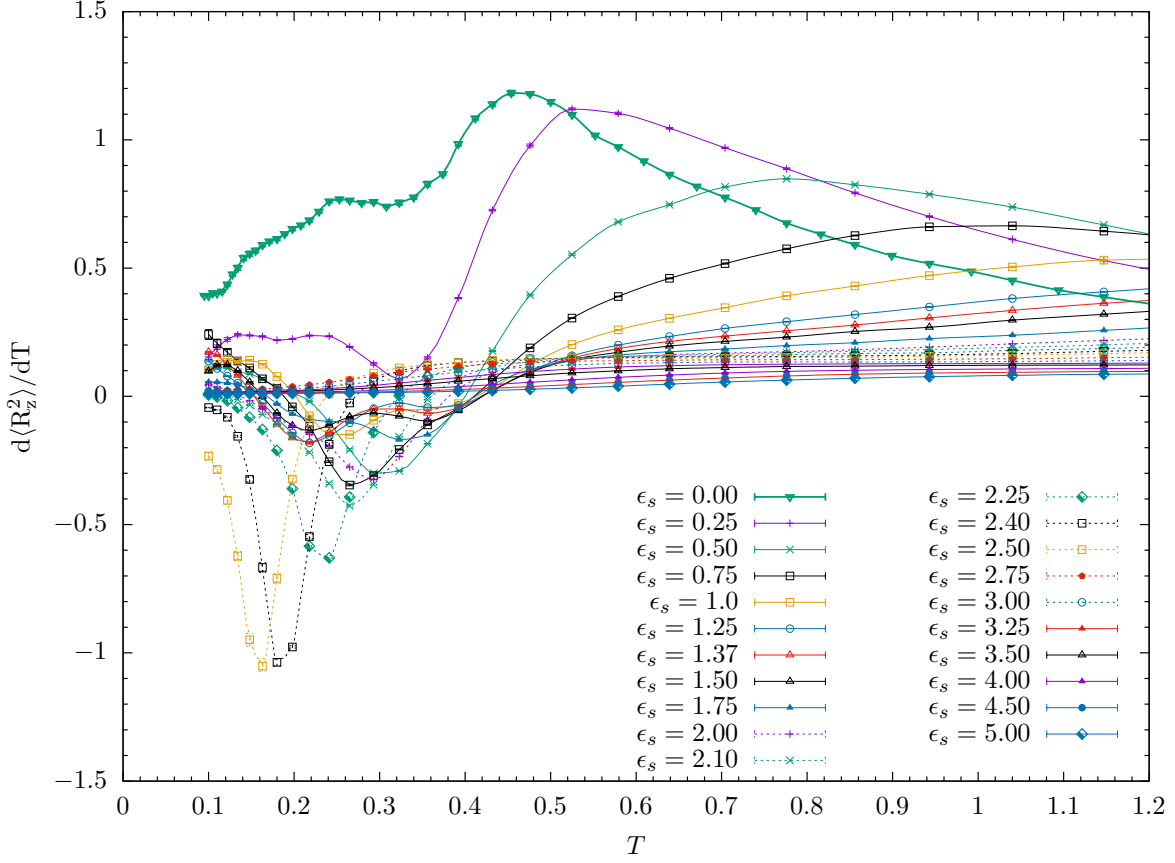
Figure 5.7: $d\left\langle R_{\mathrm{z}}^2\right\rangle/dT$ vs. T curves of 21 various ϵ_{s} values.

Note that when $\epsilon_{\mathrm{s}} = 0$, a shoulder occurs at $T \approx 0.16$, a valley occurs at $T \approx 0.32$ and a peak occurs at $T \approx 0.44$. These signals are corresponding to the freezing transition, the melting transitions and the adsorption transition. Note that also when the monomer-substrate potential switches on, the $d\left\langle R_{\mathrm{z}}^2\right\rangle/dT$ curves change dramatically. The $d\left\langle R_{\mathrm{z}}^2\right\rangle/dT$ curves are also very sensitive to the change of ϵ_{s}. When $\epsilon_{\mathrm{s}} = 0.25$, the shoulder of the curve of $\epsilon_{\mathrm{s}} = 0$ at $T \approx 0.16$ becomes a peak. This is because when the $\epsilon_{\mathrm{s}} \neq 0.0$, there will be a minimum distance between the monomers and the substrate since the essential nature of the

59

monomer-substrate interaction is a Lennard-Jones potential. When $0.25 < \epsilon_s < 2$, ground structures are not single layer structures and the curves behave similarly (double peaks) since the energy between the monomer-monomer interactions and energy between the monomer-substrate interactions are comparable. When $2 \leq \epsilon_s \leq 2.5$, the curves show noticeable dips at the freezing transition. This implies that the monomer-monomer interaction is stronger than the monomer-substrate interaction. When a little energy is added to the ground state structures, then the structures will "melt" on the substrate. Fewer signals of the curves mean fewer changes on structures.

The $d\langle z \rangle / dT$ vs. T curves of the 21 various ϵ_s values are plotted in Figure 5.8 and the minimum energy structures at temperature $T = 0.1$ for various ϵ_s are presented in Figure 5.9. The signals of the $d\langle z \rangle / dT$ curves are very similar to those of the $d\langle R_z^2 \rangle / dT$ curves shown in Figure 5.7, The signals of the $d\langle z \rangle / dT$ curves are very similar to those of the
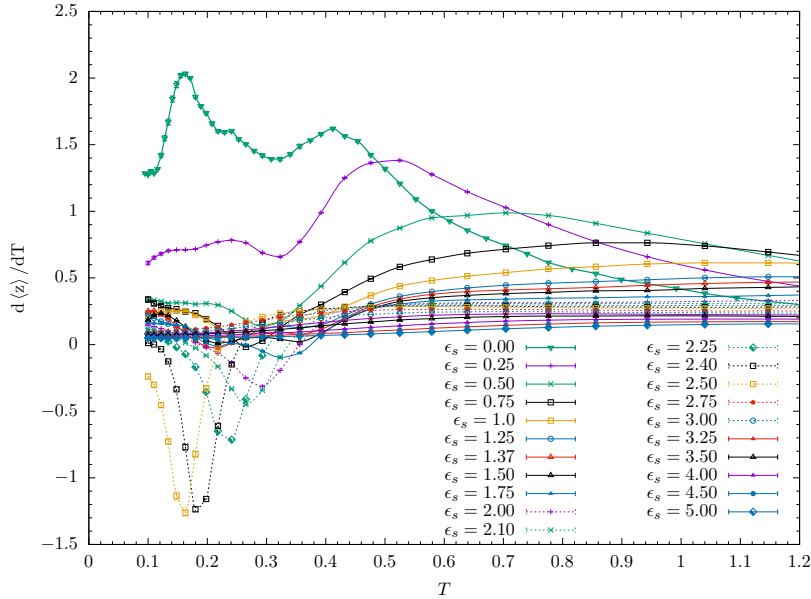
Figure 5.8: $d\langle z \rangle / dT$ vs. T curves of 21 various ϵ_s values.
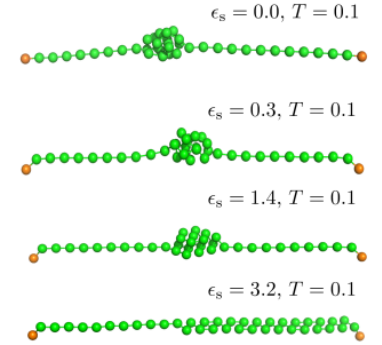
Figure 5.9: Structures at low temperature with various ϵ_s.

$d\langle R_z^2 \rangle / dT$ curves shown in Figure 5.7, because they both reflect the same conformational changes along the z direction. Therefore, $d\langle R_z^2 \rangle / dT$ and $d\langle z \rangle / dT$ could be used as order

parameters for the freezing and Θ transitions. With the information provided in Figure 5.8, a deeper understanding for the minimum energy structures in Figure 5.9 for various ϵ_s can be gained. When $\epsilon_s = 0$, the signal (peak at $T \approx 0.16$) indicated the freezing transition of the $d\langle z \rangle / dT$ curve is noticeable. This is because there is not monomer-substrate interaction, a little energy input could lead to a huge increase of the center of mass along the z direction. When $\epsilon_s > 0$, the freezing transition peak signal of the $d\langle z \rangle / dT$ curve becomes flatten and much less noticeable. This is because even a small amount of monomer-substrate interaction will prevent $\langle z \rangle$ increases as T increases. When ϵ_s increases further and $1.25 < \epsilon_s < 2.5$, the minimum energy structure becomes a two layered crystalline structure due to the increase of ϵ_s value. This results in that the freezing transition peak signal of the $d\langle z \rangle / dT$ curve, when $\epsilon_s = 0$, becomes valleys and the valley signal becomes more noticeable as ϵ_s increases. This is because the monomer-substrate interaction gets stronger. When ϵ_s increases further more and $\epsilon_s > 2.5$, the minimum energy structure becomes a single layer crystalline structure due to strong monomer-substrate interactions. This results in that the freezing transition signal of the $d\langle z \rangle / dT$ curve becomes unnoticeable since the monomer-substrate interactions are so strength then the polymer will hardly extend in the z direction and will mainly extend in the x, y directions. In summary, $d\langle z \rangle / dT$ curve can help to obtain better understanding of the minimum energy structures at low temperatures as ϵ_s varies. This section gives an idea about how to understand the conformational changes by the interpretations the signals of the energetic and structural quantities (and their thermal fluctuations). For other quantities and their thermal calculations, similar approaches can be applied to gain understanding of the conformational structures in Thermodynamics.

The $d\langle N_m \rangle / dT$ vs. T curves of the 21 various ϵ_s values are plotted in Figure 5.10. The valleys of these curves indicate "phase" transitions, because $\langle N_m \rangle$ will decrease with the increase of temperature and energy. These signals are corresponding to those of the $d\langle R_z^2 \rangle / dT$ and $d\langle z \rangle / dT$ curves.
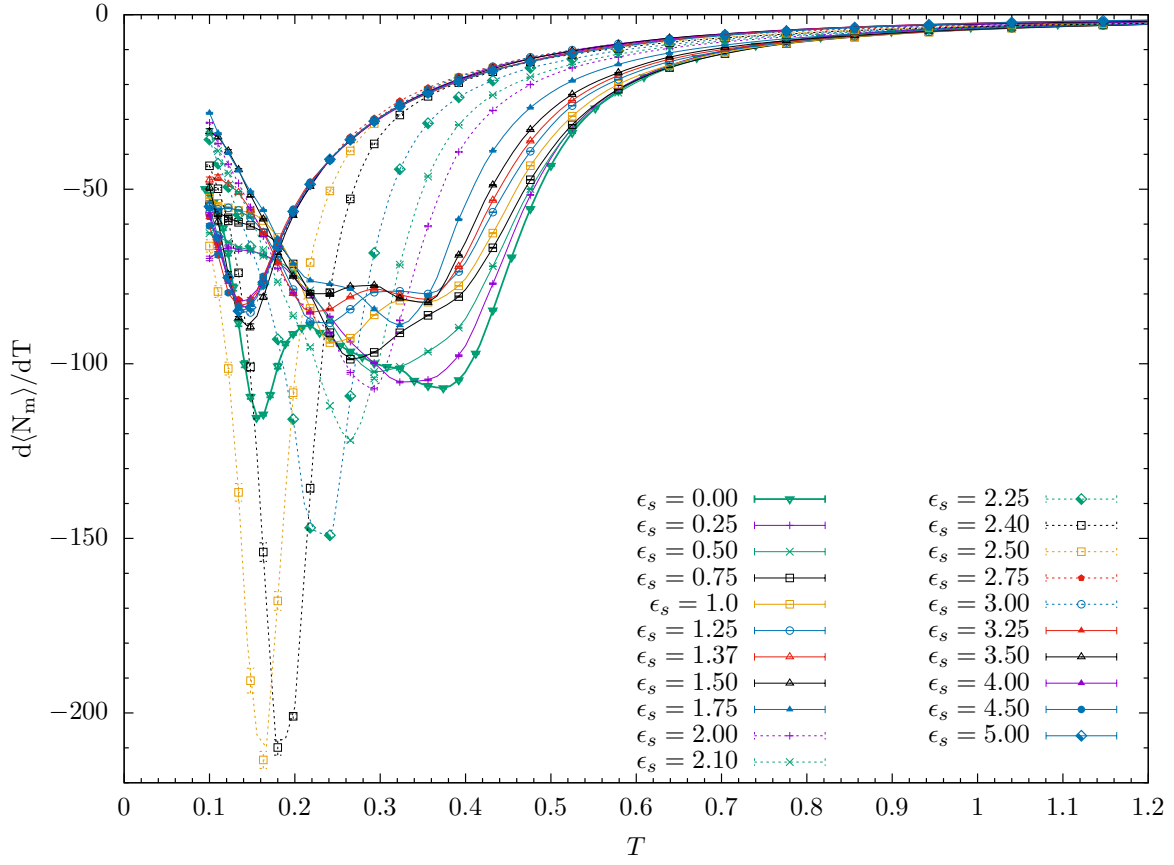
Figure 5.10: $d\langle N_{\mathrm{m}}\rangle/dT$ vs. T curves of 21 various ϵ_{s} values.

The $d\langle N_{\mathrm{s}}\rangle/dT$ vs. T curves of the 21 various ϵ_{s} values are plotted in Figure 5.11. When ϵ_{s} is small, the valleys of these curves indicate freezing transitions and the values of the transitions temperatures are comparable to those of the $d\langle N_{\mathrm{m}}\rangle/dT$ curves in Figure 5.10. However, when ϵ_{s} is larger, the transitions from the crystalline phase to the globular phase are indicated by peaks of the $d\langle N_{\mathrm{s}}\rangle/dT$ curves. This is because the change of the contact number of monomer-surface leads to dramatic energy changes.
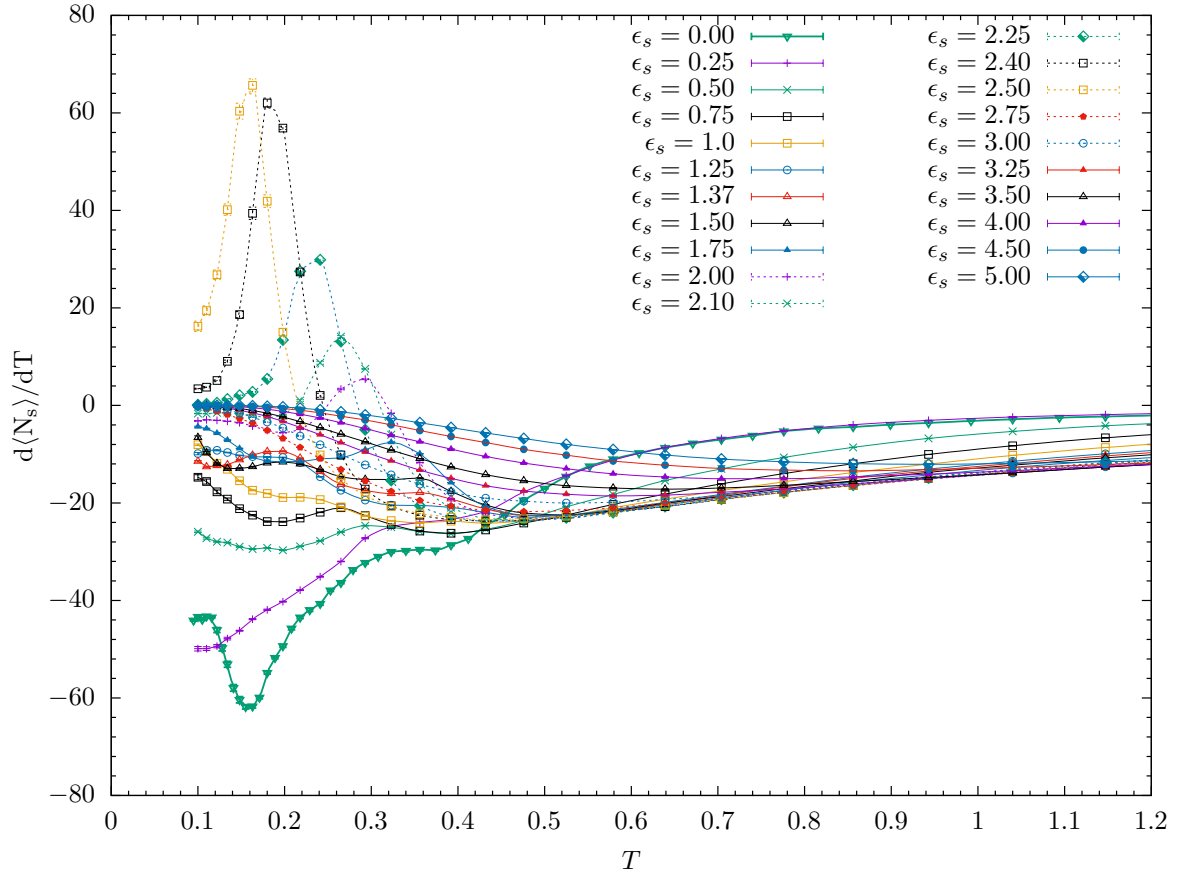
Figure 5.11: $d \langle N_s \rangle /dT$ vs. T curves of 21 various ϵ_s values.

Rearrange Data for Discerning Phase Transitions Vertically

With the canonical results from the previous section, horizontal transition bands (phase transitions caused by the changes of T) can be obtained and are shown in Figure 5.16. To obtain the phase transitions vertically (phase transitions caused by the changes of ϵ_s), the data can be rearranged according to the same T values with different ϵ_s values and the c_V vs. ϵ_s curves at various temperatures can be plotted. These curves are presented in Figures 5.12 - 5.15.

63
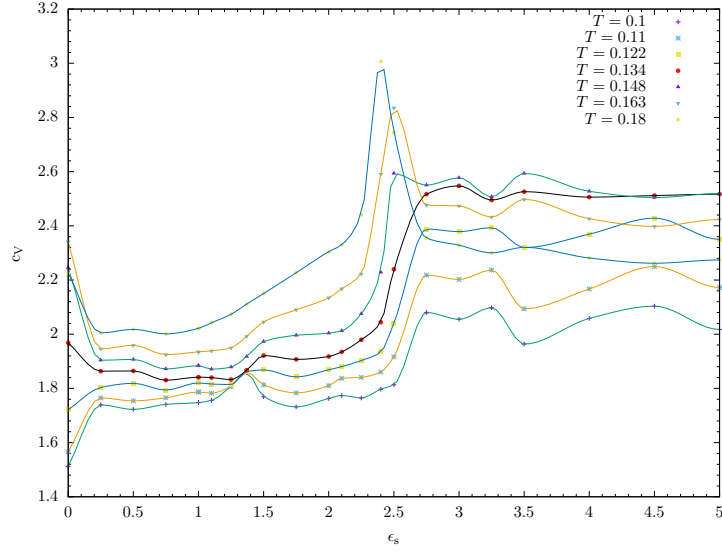
Figure 5.12: c_V vs. ϵ_s curves at various temperatures $T \in [0.1, 0.18]$.
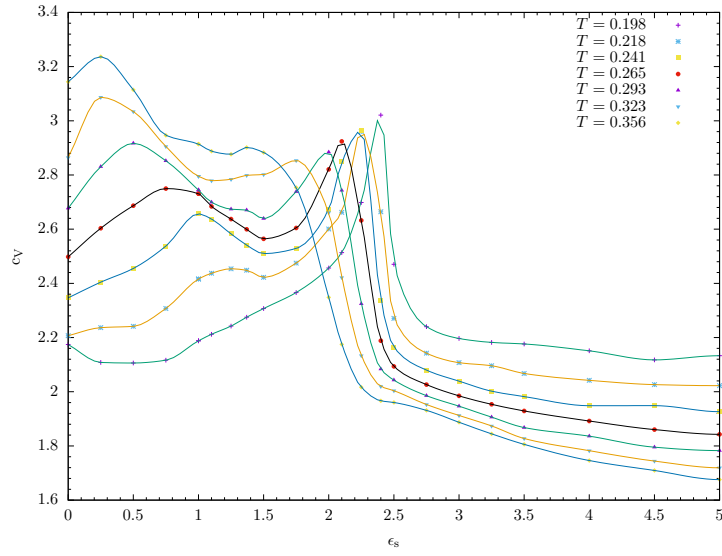
Figure 5.13: c_V vs. ϵ_s curves at various temperatures $T \in [0.198, 0.356]$.

Figure 5.12 shows that there are three possible transitions at $\epsilon_s \approx 0.2$, 1.2 and 2.4 at low temperatures. Figure 5.12, 5.13 and 5.14 show three possible transition strips which are changing with the increase of temperature and ϵ_s. From the observations from these figures,
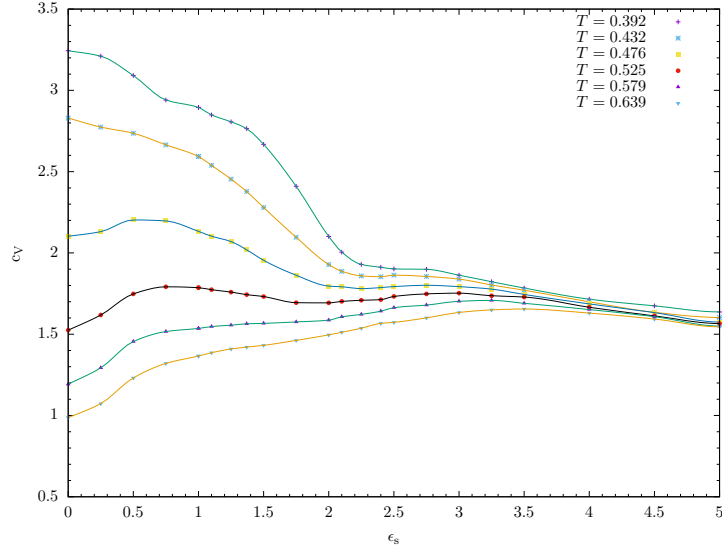
Figure 5.14: c_V vs. ϵ_s curves at various temperatures $T \in [0.392, 0.639]$.
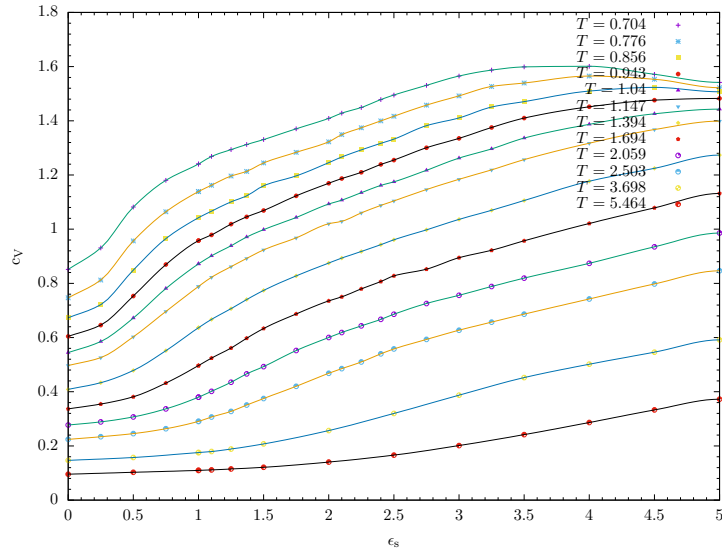
Figure 5.15: c_V vs. ϵ_s curves at various temperatures $T \in [0.704, 5.464]$.

three transitions can be discovered, the three transitions are corresponding to the transition between the globular phase with an icosahedral core to the semi-spherical droplet globular phase; the transition between the semi-spherical droplet globular phase to a 2D globular

phase and the transition between the 3D globular phase with local adsorption clusters and the 3D globular phase.

5.2.3 Phase Diagrams and Conformational Structures

Considering all the results provided in Section 5.2.2 and verifying the structures of all possible corresponding phases, a T vs. ϵ_s conformational phase diagram can finally be constructed and presented in Figure 5.16.
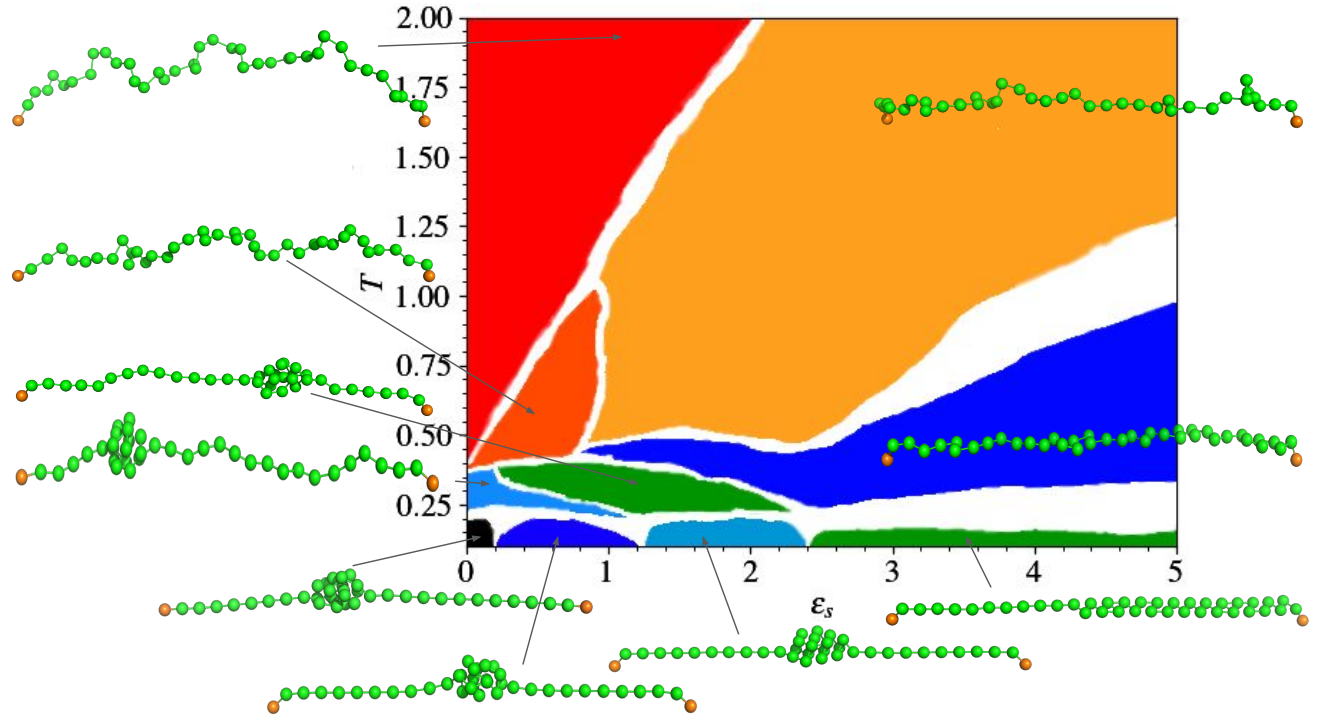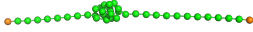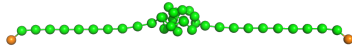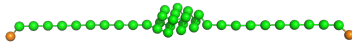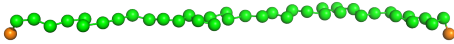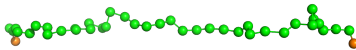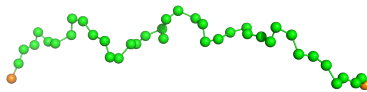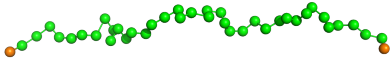

Figure 5.16: T vs. ϵ_s conformational pseudophase diagram of polymer-substrate systems with $D = 25$. See the text for an explanation.
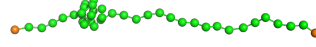
The different phases are distinguished by their colors. The representative structures are also presented. The information about the phases and the corresponding structures are listed in Table 5.1, in a counter-clockwise order starting at the structure at the lower left hand corner of Figure 5.16.

66

Table 5.1: Different phases and their representative structures.

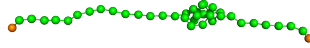| Phases	Structures	ϵ_s	T
1) compact icosahedral structure with straight strands		0.0	0.1
2) compact icosahedral structure with parallel strands		0.3	0.1
3) two-layer compact crystalline structure		1.4	0.1
4) single layer compact crystalline structure		3.2	0.1

5) 2D globule		3.0	0.432
6) 3D globule with local adsorption clusters		3.0	1.694
7) 3D random coil		1.1	2.05
8) 3D globule		0.5	0.579

9) 3D globule with a spherical droplet		0.0	0.293
10) 2D globule with a semi-spherical droplet		1.1	0.323

When constructing the phase diagram, the c_V curves play important roles, can be a potential order parameter. Other order parameter candidates could be the thermal fluctuations of $\langle N_m \rangle$ and $\langle N_s \rangle$. However none of them can be used as the only order parameter and a comprehensive understanding of all of them has to be used for the construction of the phase diagram.

5.3   Summary

In this chapter, canonical quantities of doubly grafted polymer-and-substrate systems with a fixed end-to-end distance $D = 25$ are studied. Previously studies of free polymer-and-substrate systems and singly grafted polymer-and-substrate systems are summarized in Section 5.1. A showcase data analysis process is introduced in detail and the conformational phases and phase transitions of this showcase data set are also discussed in detail presented. Ideas about how to understand the conformational changes by the interpretations the signals of the energetic and structural quantities (and their thermal fluctuations) is given by using

the examples of analyzing the $d\langle z\rangle/dT$ curves with different ϵ_s values. The conformational phases transition as ϵ_s varies are also studied by rearranging the data obtained from the PT simulations according to the same T values with different ϵ_s values.

Finally, the T vs. ϵ_s phase diagram is constructed and the representative structures for different phases are listed.

Due to the nature of the canonical analysis, for finite systems, only transition strips can be obtained. To reduce the size of the transition bands, the microcanonical analysis method introduced in Chapter 2 can be used. The conformational analysis is time consuming and labor intensive. Machine learning methods for classifying the conformations automatically are preferred for efficiency in the future [102, 103, 104].

Chapter 6

Computer Simulation Results for Polymer-Substrate Systems with a Fixed Substrate Surface Attraction Strength

6.1 Introduction

In the previous chapter, the results for doubly grafted polymer-substrate systems with a fixed end-to-end distance D at various temperatures and various substrate surface attraction strengths were discussed. With both ends grafted, the polymers have stronger monomer-substrate interactions compared to the free and singly grafted polymers when they have the same ϵ_s values. For the doubly grafted polymer-substrate systems, their monomer-substrate interactions become stronger as the D value increases. For free and singly grafted polymers, since no constraints are along the x, y directions, they are isotropic in the xy-plane that is parallel to the substrate surface. Therefore, when they interact with the substrate, structures

with an axial symmetry tend to be formed, e.g., compact icosahedral structures, compact dish-like layered crystalline structures at low temperatures with various ϵ_s values [67] and "mushroom-like" structures [105]. However, for doubly grafted polymers, since both ends of the polymers are grafted and the two ends are located along the x direction, they are not isotropic in the xy-plane that is parallel to the substrate surface. Therefore, the constraint of the two grafted ends breaks the isotropic properties of the free and singly grafted polymers and the conformational structures will be affected by the end-to-end distance D values. The conformational structures depend on the potentials in the energy model and the D value affects the number of the monomers involved in the monomer-substrate interactions and the monomer-monomer interactions. Hence, the conformational phases and phase transitions of doubly grafted polymers with different D values are of interest.

In this chapter, the conformational phase and conformational transitions of the doubly grafted polymer-and-substrate systems with a fixed substrate surface attraction strength $\epsilon_s = 1.0$ at various temperatures and various end-to-end distances D are studied. The D values vary from 1 to 38 with an interval of 1. The physics, the model and the simulation methods are similar to those used in Chapter 5. However, some specific parameters of the model and the simulation procedures will be different and given below.

6.1.1 Specific Parameters of the Model and Simulations

The modeling parameters for the doubly grafted polymer-and-substrate systems with a fixed substrate surface attraction strength $\epsilon_s = 1.0$ at various temperatures and various end-to-end distances D are:

1. $N = 40$ is the length of the homopolymer;

2. $\epsilon_s = 1.0$ is the attraction strength;

3. D is the distance between the two grafted ends and $D \in [1, 38]$;

4. T is the temperature and $T \in [0.1, 5.6]$.

The PT simulation schemes and parameters are:

1. $K = 32$ is the number of temperatures, where 32 is obtained by the criterion given in Section 4.5;

2. Replicas are assigned to the K different temperatures;

3. In each simulation process, the seeds of the random number generator are assigned differently;

4. Eight energetic and structural quantities are measured for identification of phases transitions;

5. Jackknife analysis is used for error estimation;

6. Various D values within the range of $[1, 38]$ are used for discerning phases and "phase" transitions at different temperatures;

7. Simulated annealing method is used to find the lowest energy structures.

The computational resources used are the same as mentioned in Section 5.1.1. The average runtime for most D values is again around $2 \sim 3$ days, but there could be longer runtimes for certain D values, e.g., for $D = 11$, the replica exchange rate of the parallel tempering simulation is lower compared to the rates of other D values. Therefore, the runtime for $D = 11$ is 5 times longer than the average runtime. The proper modeling parameters of simulations with different D values have to be tried in small-scale trial runs before the full-scale simulations are conducted.

If only considering the average runtime of simulations with different D values, the approximate total number of the computational CPU core-years for implementing all the simulations

in this chapter is $2 \sim 3$ (days) $\times$ 32 (cores) $\times$ 38 (different D values) $\approx 6.7 \sim 10.0$ core-years, in addition to the time for data analysis.

6.2 Results

In this section, the results of the doubly grafted polymer-and-substrate systems with a fixed substrate surface attraction strength $\epsilon_{\mathrm{s}} = 1.0$ at various temperatures and various end-to-end distances D are presented. The result analysis methods are similar to those used in Chapter 5, therefore the detailed explanation of the analysis methods will not be repeated. The results will be explained by comparing with the results presented in Chapter 5. This section begins with three sets of example results of polymer-substrate systems with $\epsilon_{\mathrm{s}} = 1.0$ and $D = 5,\ 15$ and 30. They will be presented by comparing with the results of systems with $\epsilon_{\mathrm{s}} = 1.0$ and $D = 25$ given in Chapter 5. Then all of the canonical quantities of different D values will be presented in eight groups of figures.

6.2.1 Results for Polymer-Substrate Systems with $\epsilon_{\mathrm{s}} = 1.0$ and $D = 5, 15, 30$

Examples of thermodynamic and structural properties for doubly grafted polymer-substrate systems with $\epsilon_{\mathrm{s}} = 1.0$, $D = 5,\ 15$ and 30 and various T values are presented in Figure 6.1. Since $D = 30$ is close to $D = 25$, the results of $D = 30$ will be discussed first, by comparing with the results of the previous chapter.

When $D = 30$, in the last column of Figure 6.1, one peak indicates the "melting" transition at $T \approx 0.275$ in the sub-figure (g), and a shoulder indicates the "freezing" transition at $T \approx 0.2$. This similar "one shoulder and one peak" pattern can also be spotted in the c_{V} vs. T curve of polymer-substrate systems for $D = 25$ in Figure 5.3. Similarly, in the high temperature range, no obvious signal for the Θ transition can be spotted at the c_{V} curve

for $D = 30$. However, the signal at $T \approx 1.2$ for the Θ transition can be discerned in the $d \langle R_z^2 \rangle /dT$ and $d \langle z \rangle /dT$ curves shown in the sub-figure (h). Note that the wide "bumps" at $T \approx 1.2$ of the $d \langle R_z^2 \rangle /dT$ and $d \langle z \rangle /dT$ curves are wide "peaks" actually; if the temperature range increases, the "bumps" will become peaks and they indicate that the polymers extend slowly along the z direction as T increases. Notice that the peak signals of the $d \langle R_x^2 \rangle /dT$ and $d \langle R_y^2 \rangle /dT$ curves always appear before the peak signals of the $d \langle R_z^2 \rangle /dT$ and $d \langle z \rangle /dT$ curves. This is because when the attraction strength $\epsilon_s = 1.0$, the monomer-substrate interactions are only related to the z component of a monomer. Therefore, the polymer's extension along the z direction needs more energy than the extensions in the x, y directions. This is also true for $D = 5$ and 15 by comparing the $d \langle R_x^2 \rangle /dT$ and $d \langle R_y^2 \rangle /dT$ curves with the $d \langle R_z^2 \rangle /dT$ and $d \langle z \rangle /dT$ curves. It seems that $d \langle R^2 \rangle /dT$ and $d \langle R_x^2 \rangle /dT$ curves show strong correlations for larger D values, e.g., $D = 25, 30$. Note that the shoulder signal indicating the freezing transition of the c_V curve matches the valley signal of the $d \langle N_s \rangle /dT$ curve, which is also similar to what has been observed for $D = 25$ in the previous chapter. Therefore, $d \langle N_s \rangle /dT$ curves are important to study for identifying various conformational structures at low temperatures because they are more sensitive to the changes of the compact structures compared to other quantities. Together with the contact map method [41, 34] and z component distribution method [106], these signals discussed above could be used comprehensively to detect the icosahedral core and the fcc crystalline conformations at low temperatures.

For a stronger contrast, results for $D = 5$ will be discussed. When $D = 5$, in the first column of Figure 6.1, two peaks can be found at $T \approx 0.2$ (freezing transition) and 0.4 ("melting" transition), and a shoulder at $T \approx 1.35$ (Θ transition) for the c_V curve. Compared to the results for $D = 30$, it seems that systems with smaller D values are more sensitive to the energy changes and behave similarly to free and singly grafted polymer-and-substrate systems. These peaks signals match the valleys signals on the $d \langle N_m \rangle /dT$ curves. Therefore,
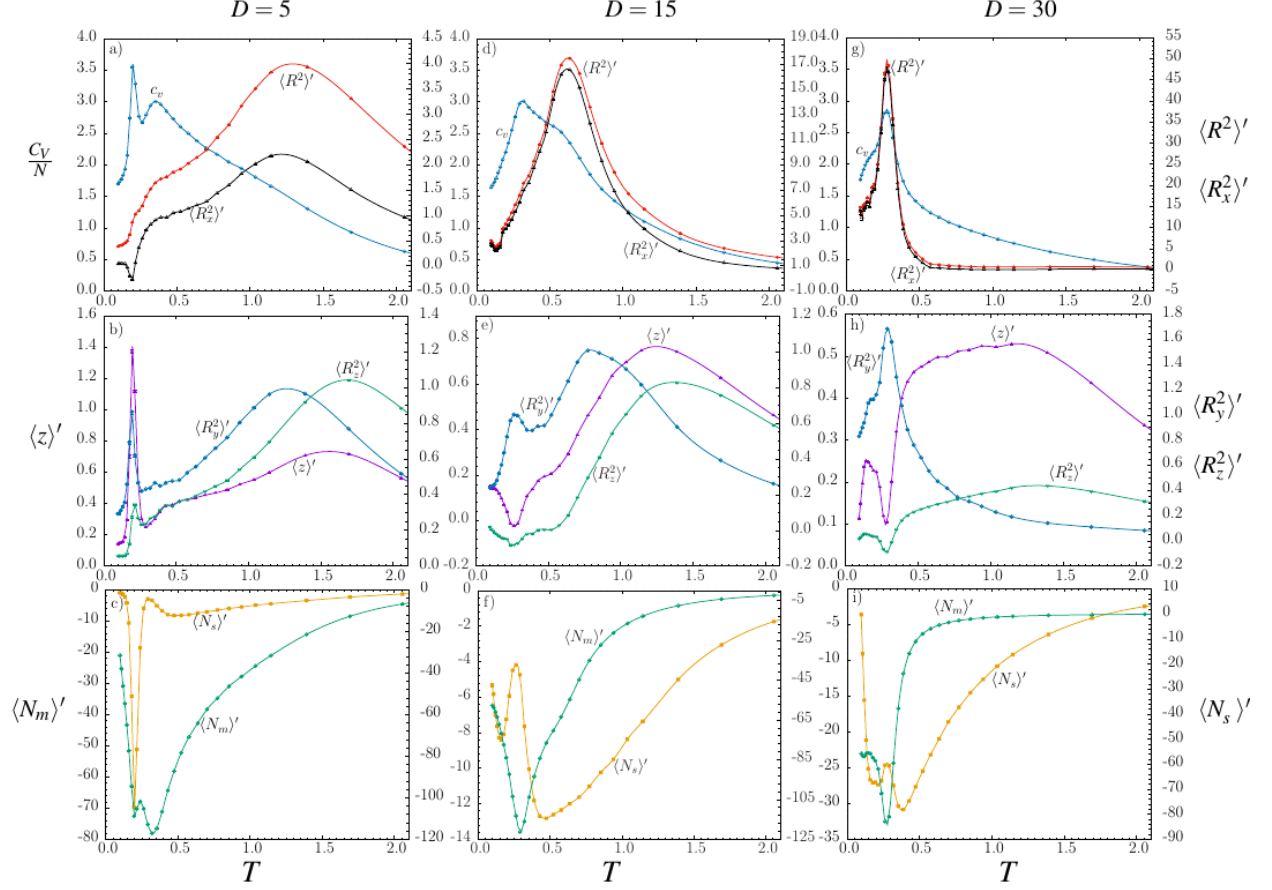
Figure 6.1: Thermodynamic and structural quantities for polymer-substrate systems with $\epsilon_\mathrm{s} = 1.0$, $D = 5, 15$ and 30 and various T. Some error bars are within the size of the data points.

$d \langle N_m \rangle /dT$ could be used to indicate various transitions for smaller D values. When $D = 5$, note that the Θ transition signal ($T \approx 1.4$) of the $d \langle R^2 \rangle /dT$ curve is located in between those of $d \langle R_x^2 \rangle /dT$, $d \langle R_y^2 \rangle /dT$ and $d \langle R_z^2 \rangle /dT$ curves. The Θ transition signal of the $d \langle R_x^2 \rangle /dT$ curve matches that of the $d \langle R_y^2 \rangle /dT$ curve but not that of the $d \langle R_z^2 \rangle /dT$ curve. This is a good example of the finite size effect. One possible reason can be that when energy is added to the system, the polymer is expected to extend in the x and y directions first, since the polymer-substrate potential only exists in the z direction. As the energy increases further,

the polymer will finally extend in z direction. For an infinitely long polymer, these different peak signal locations should reduce to a single value.

When $D = 15$, the patterns indicating various transitions of the quantities fall between the patterns with $D = 5$ and $D = 30$. For simplicity, detailed explanations will be omitted for $D = 15$. To understand how the minimum energy structures may change as D changes, Figure 6.2 presents the minimum energy structures for $D = 8$ and $D = 12$. When $D < 11$, the minimum energy structures are fcc crystalline structures and when $11 < D < 21$, the minimum energy structures become compact imperfect icosahedral core structures. This reflects the competition between the monomer-monomer interactions and the monomer-substrate interactions as D changes.
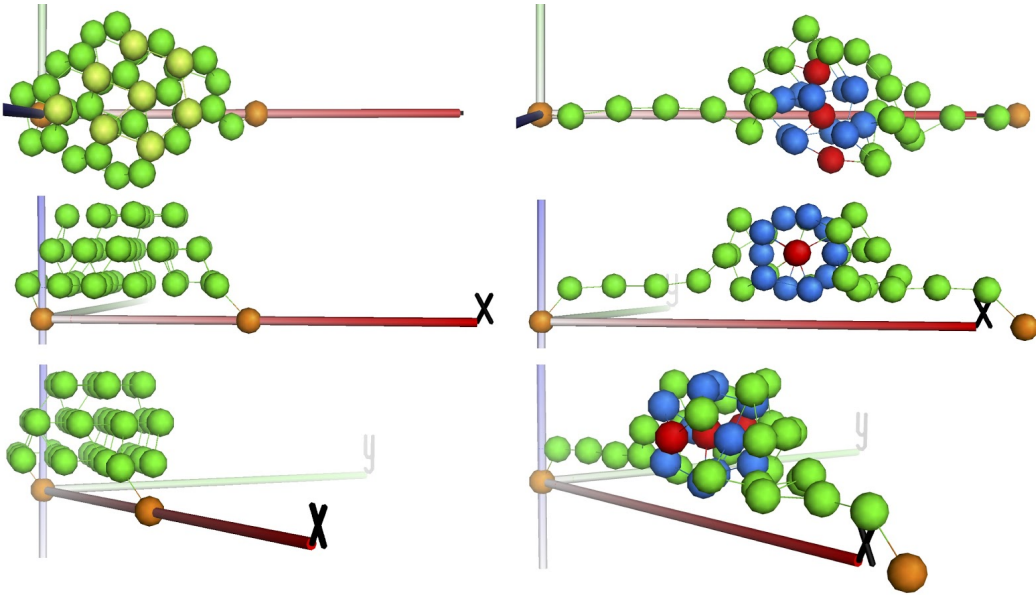
Figure 6.2: Top view, side view and 45° top view of the minimum energy structures for $D = 8$ (left) and $D = 12$ (right). Detailed explanations are in the text.

6.2.2 Canonical Results

In this section, all results of the canonical quantities for different D values will be presented. Since the amount of data is larger than those of the previous chapter, the data are split into

four intervals, which are $D \in [1, 10]$, $D \in [11, 20]$, $D \in [21, 30]$ and $D \in [31, 38]$. When D is large, there are fewer conformational varieties. Therefore, some data with larger D values are omitted. Since all of the data are analyzed in a similar manner as those in the previous chapter, for simplicity, not all of the results will be discussed in detail; only unusual results that may trigger some interesting findings in physics will be discussed.

The c_V vs. T curves of various D values are plotted in Figures 6.3 - 6.6. Note that when $D \in [1, 11]$, the transition temperatures of the "melting" transitions are almost unchanged at around 0.36. The peak signal indicating the freezing transition, initially at $T = 0.23$ when $D = 1$, will keep moving to the left as D increases and will disappear when $D > 11$. At low temperatures, an fcc compact crystalline phase can be discerned when $D \in [1, 11)$ as shown in Figure 6.35. There are also shoulder signals indicating the Θ transitions on these curves. When D is small, the Θ transition signals of the curves are unnoticeable. However, the same transition can be hinted by the signals of the $d\langle R^2 \rangle / dT$ curves. The Θ transition signals will move to the left and become more notable as D increases. The c_V vs. T curve will eventually reduce to a simple one peak curve as D increases further, because the number of monomers that can be used to construct various conformational structures decreases as D increases. Notice the unusual signals for $D = 11$ and recall that the replica exchange rate is also small for $D = 11$. The possible reason could be, when $D = 11$, the energy of the fcc crystalline structures and the energy of the imperfect icosahedral core structures are very similar to each other and the energy distributions around the transition temperatures are unimodal, which could indicate that this may be a second-order transition. This could be interesting to compare with previous studies [98, 107].
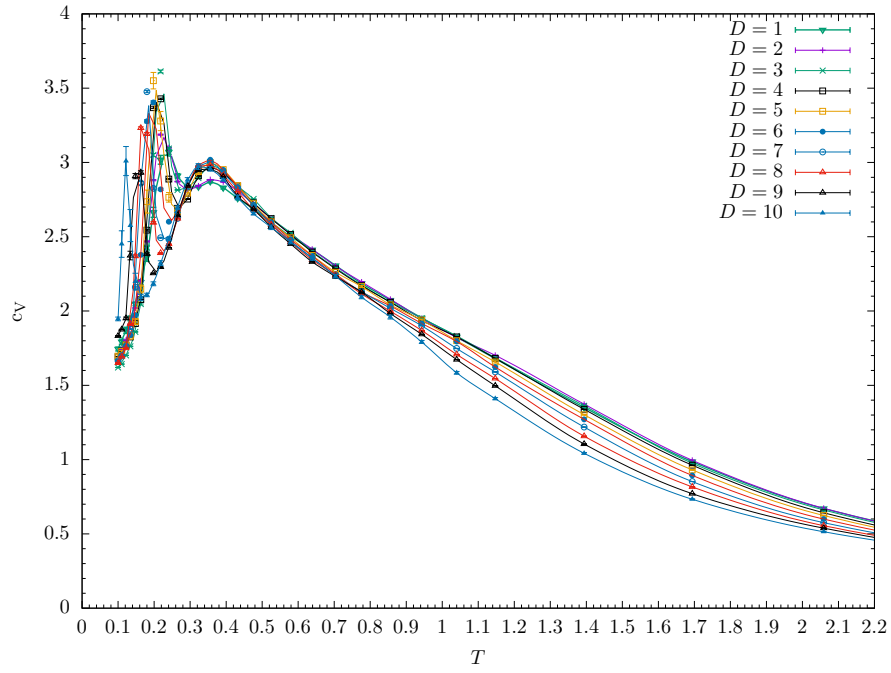
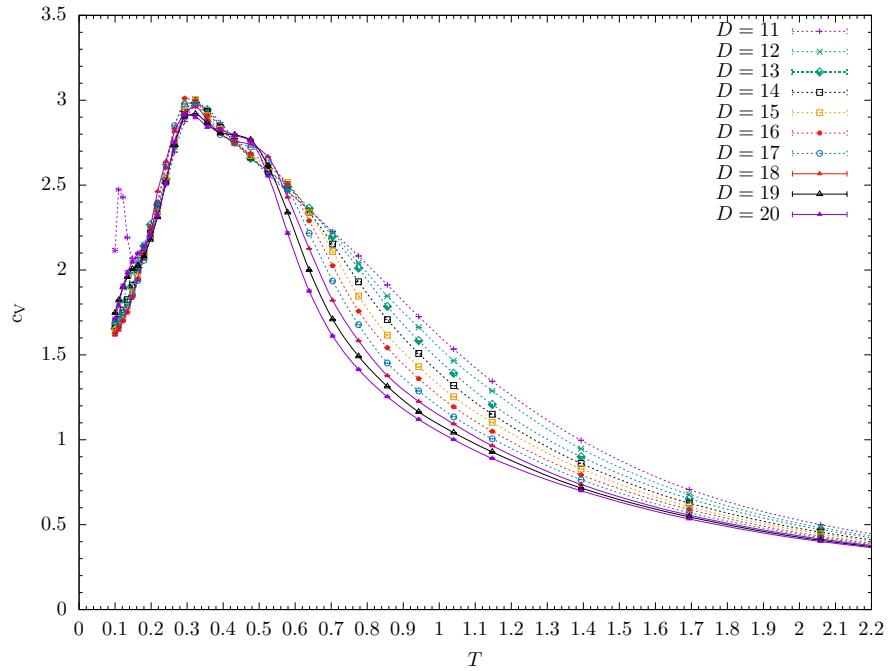Figure 6.3: c_V vs. T curves for $D \in [1, 10]$.

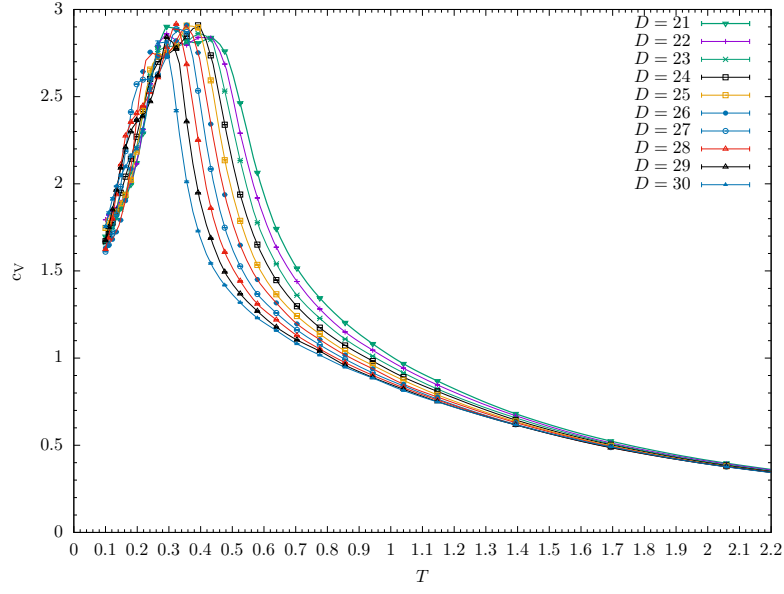Figure 6.4: c_V vs. T curves for $D \in [11, 20]$.

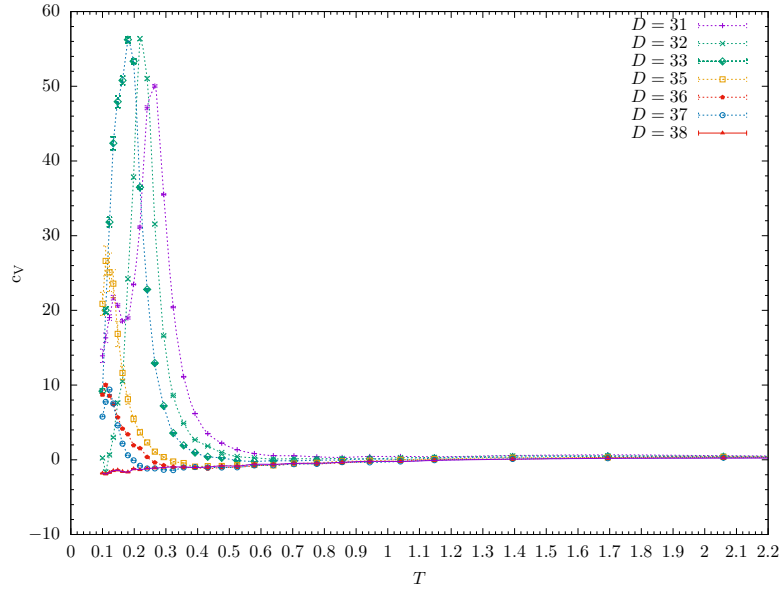Figure 6.5: c_V vs. T curves for $D \in [21, 30]$.

Figure 6.6: c_V vs. T curves for $D \in [31, 38]$.

The $d\langle R^2\rangle/dT$ vs. T curves at various D values are presented in Figures 6.7 - 6.10. These signals of the curves are comparable to those of the c_V curves.
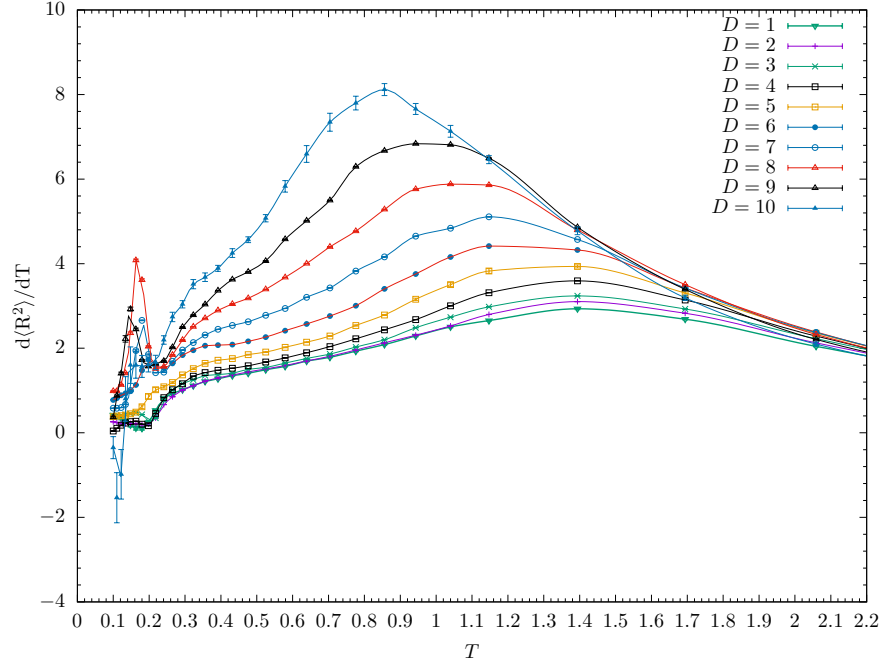

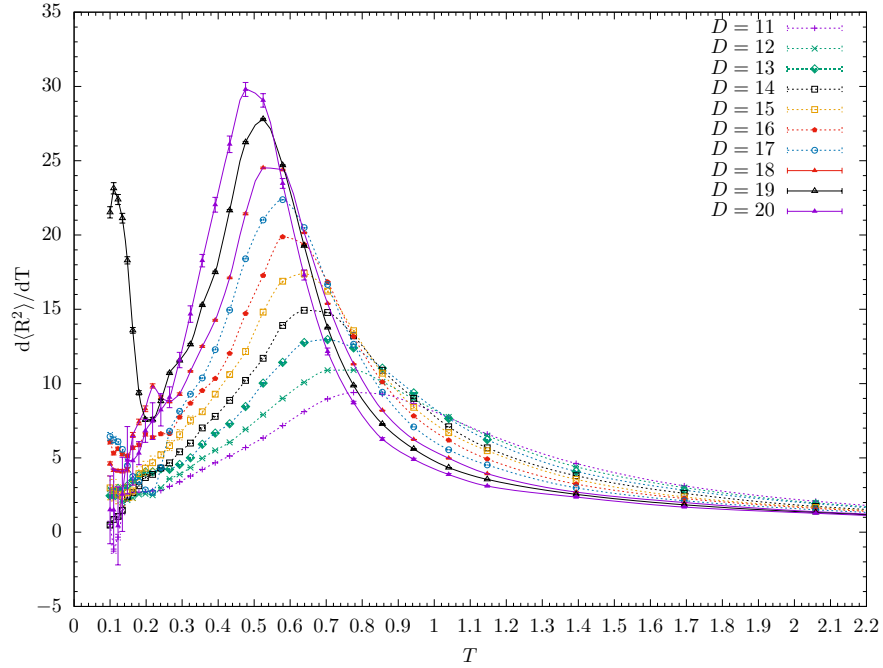Figure 6.7: $d\langle R^2\rangle/dT$ vs. T curves for $D \in [1, 10]$.

Figure 6.8: $d\langle R^2\rangle/dT$ vs. T curves for $D \in [11, 20]$.

81
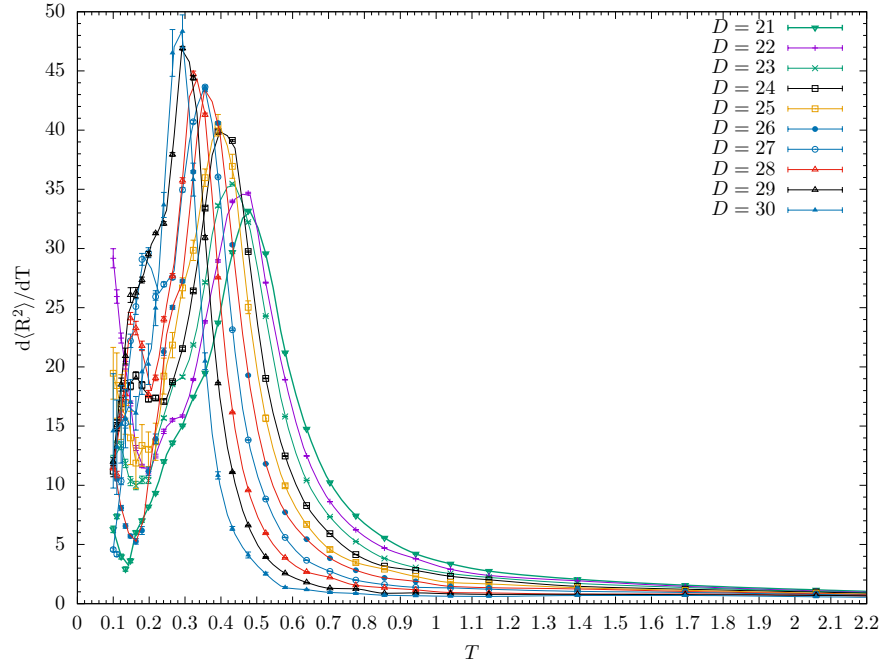
Figure 6.9: $d \langle R^2 \rangle /dT$ vs. T curves for $D \in [21, 30]$.
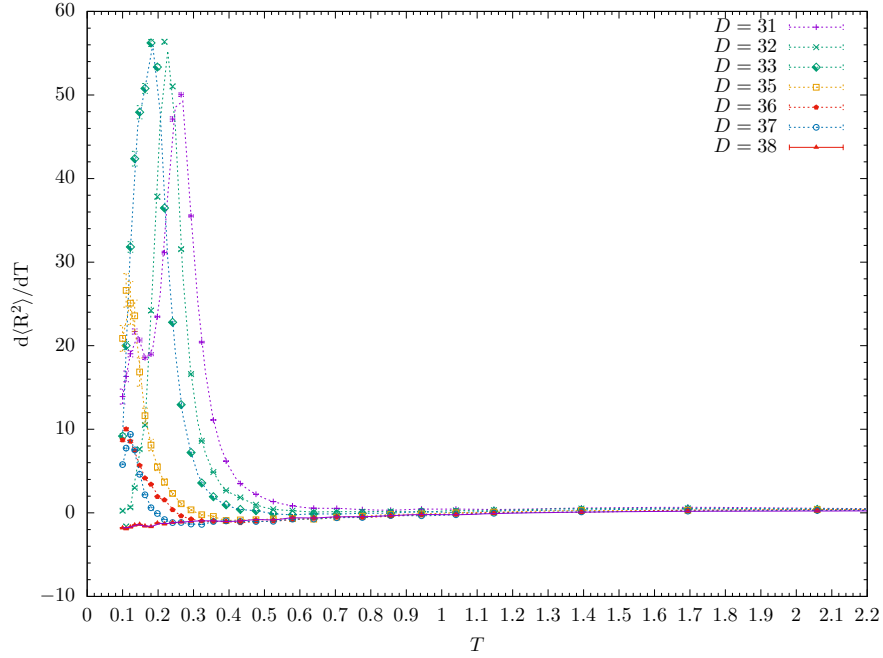
Figure 6.10: $d \langle R^2 \rangle /dT$ vs. T curves for $D \in [31, 38]$.

The $d\langle R_x^2\rangle/dT$ vs. T curves at various D values are given in Figures 6.11 - 6.14. Note that the signals of these curves are comparable to those of the c_V and $d\langle R^2\rangle/dT$ curves.
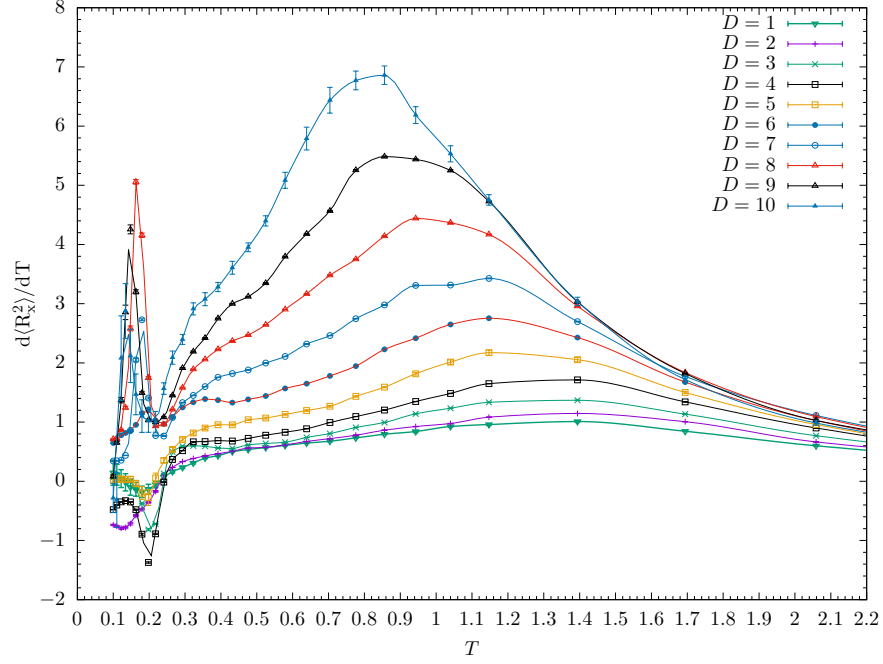

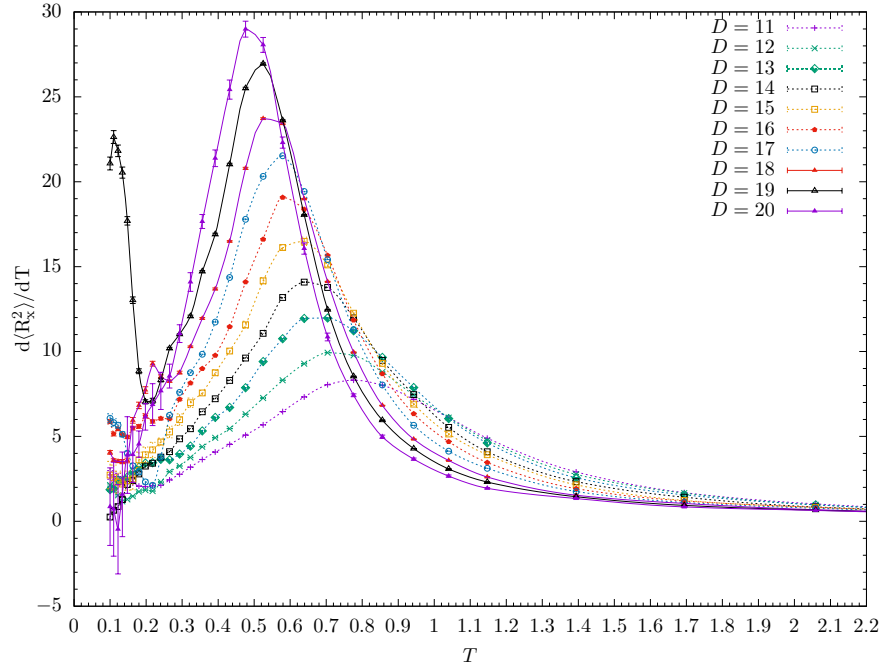Figure 6.11: $d\langle R_x^2\rangle/dT$ vs. T curves for $D \in [1, 10]$.

Figure 6.12: $d\langle R_x^2\rangle/dT$ vs. T curves for $D \in [11, 20]$.
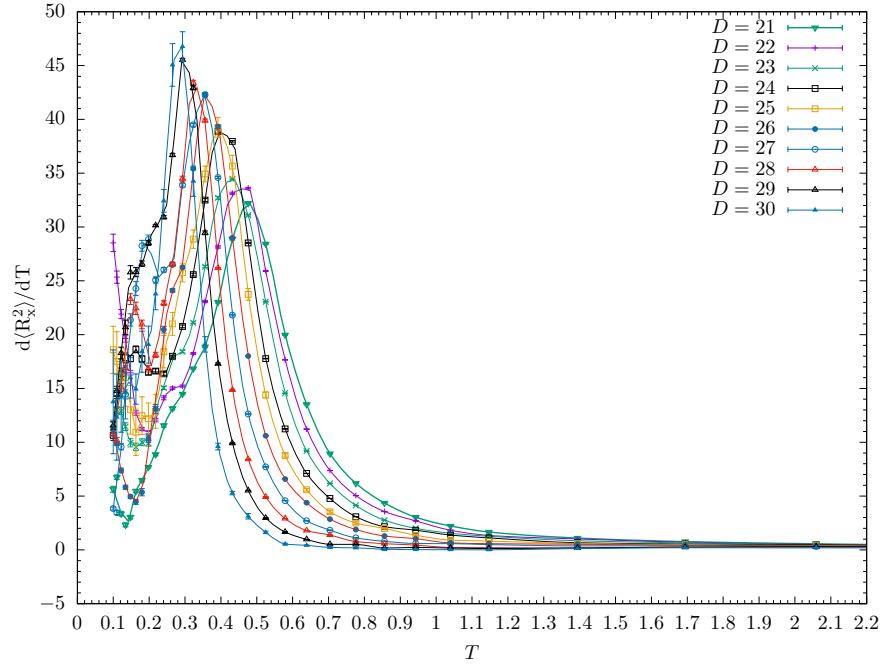
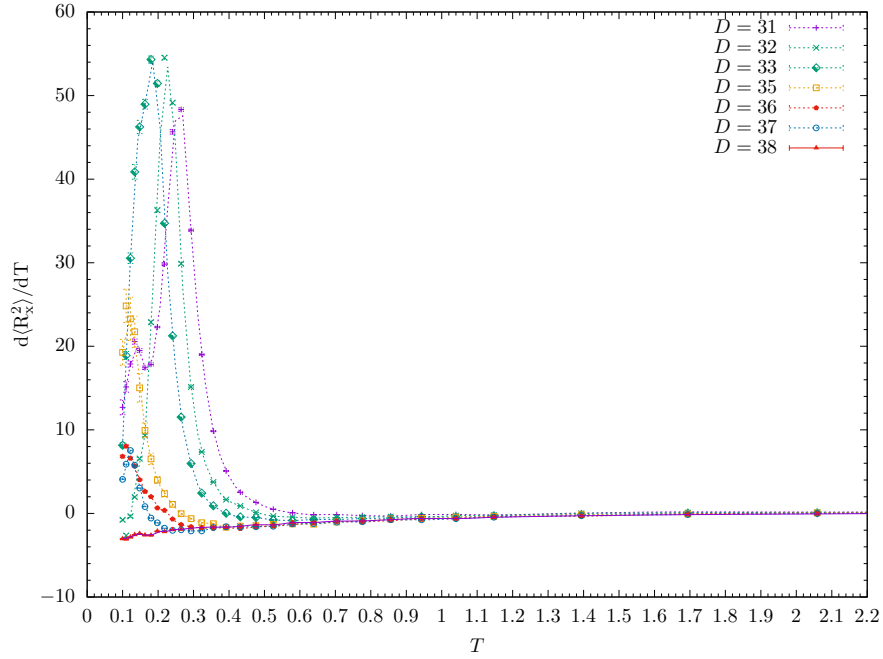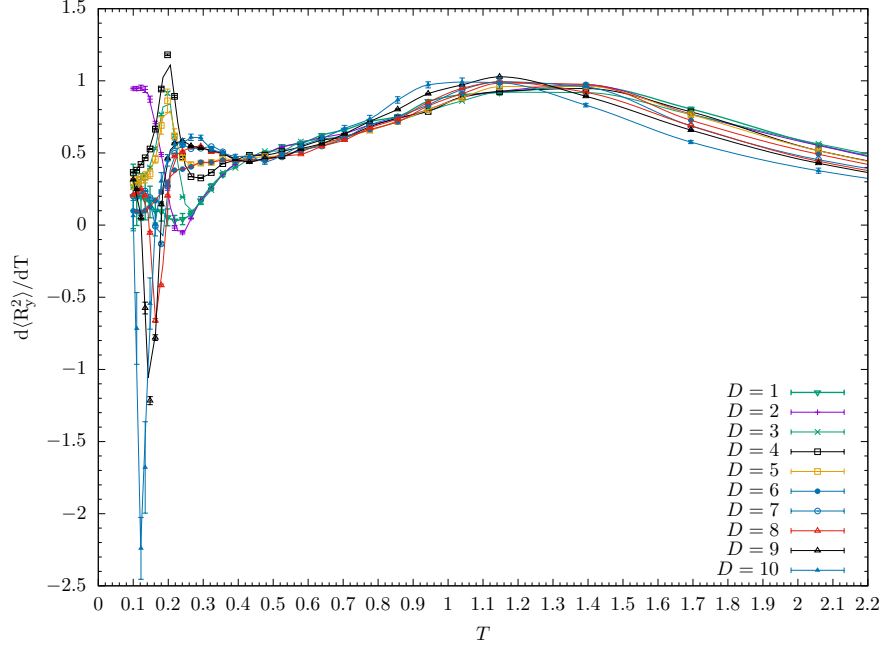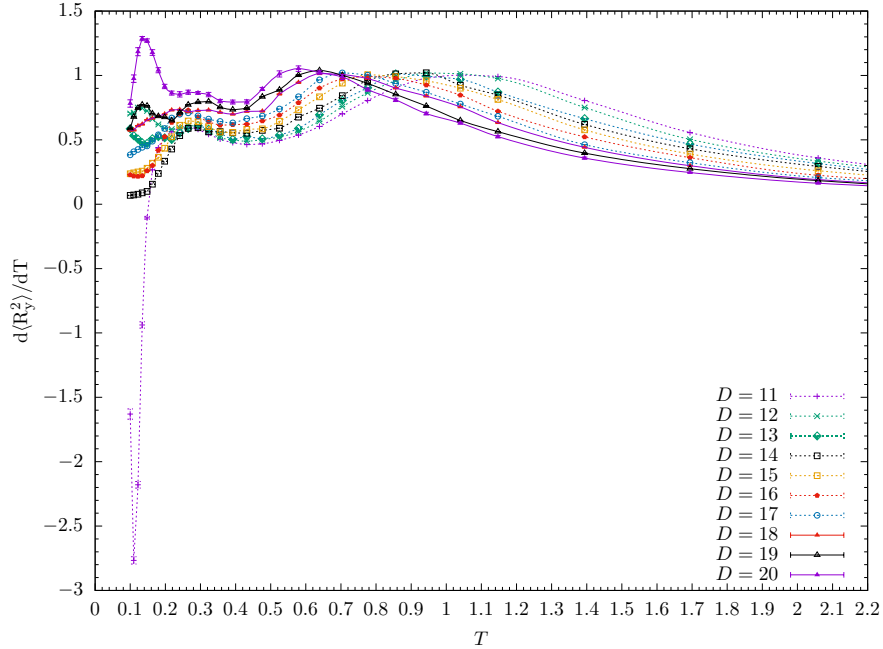Figure 6.13: $d\langle R_x^2\rangle/dT$ vs. T curves for $D \in [21, 30]$.

Figure 6.14: $d\langle R_x^2\rangle/dT$ vs. T curves for $D \in [31, 38]$.

The $d\left\langle R_y^2\right\rangle/dT$ vs. T curves at various D values are given in Figures 6.15 - 6.18. Note that the signals of these curves are comparable to those of the c_{V}, $d\left\langle R^2\right\rangle/dT$ and $d\left\langle R_x^2\right\rangle/dT$ curves. Note the unusual signals when $D = 10, 11$.

Figure 6.15: $d\left\langle R_y^2\right\rangle/dT$ vs. T curves for $D \in [1, 10]$.

Figure 6.16: $d\left\langle R_y^2\right\rangle/dT$ vs. T curves for $D \in [11, 20]$.

Figure 6.17: $d\left\langle R_y^2\right\rangle/dT$ vs. T curves for $D \in [21, 30]$.

Figure 6.18: $d\left\langle R_y^2\right\rangle/dT$ vs. T curves for $D \in [31, 38]$.

The $d \langle R_z^2 \rangle /dT$ vs. T curves at various D values are given in Figures 6.19 - 6.22. Note that the signals of these curves are comparable to those of the c_V, $d \langle R^2 \rangle /dT$, $d \langle R_x^2 \rangle /dT$ and $\langle R_y^2 \rangle /dT$ curves.

Figure 6.19: $d \langle R_z^2 \rangle /dT$ vs. T curves for $D \in [1, 10]$.

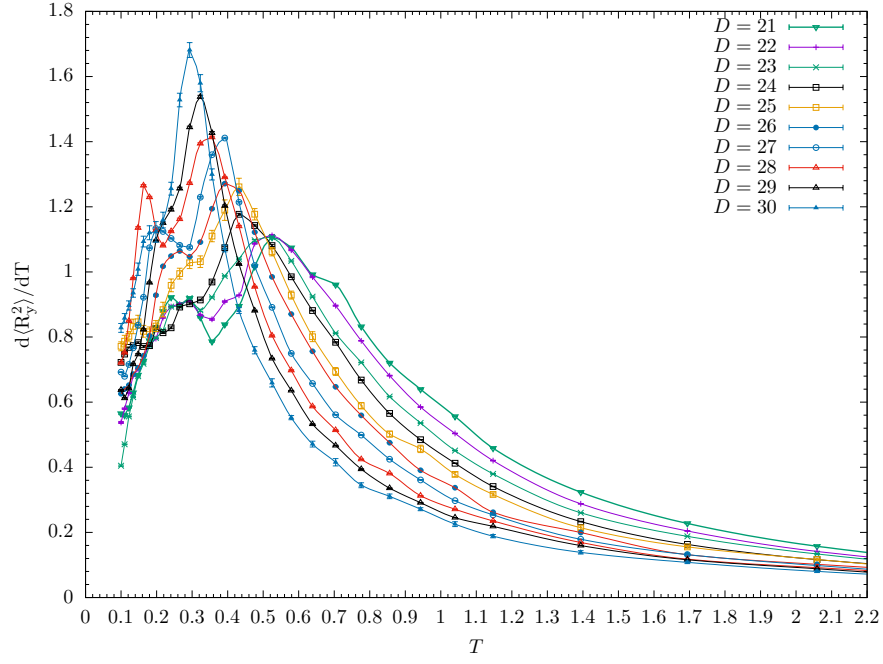Figure 6.20: $d \langle R_z^2 \rangle /dT$ vs. T curves for $D \in [11, 20]$.

87

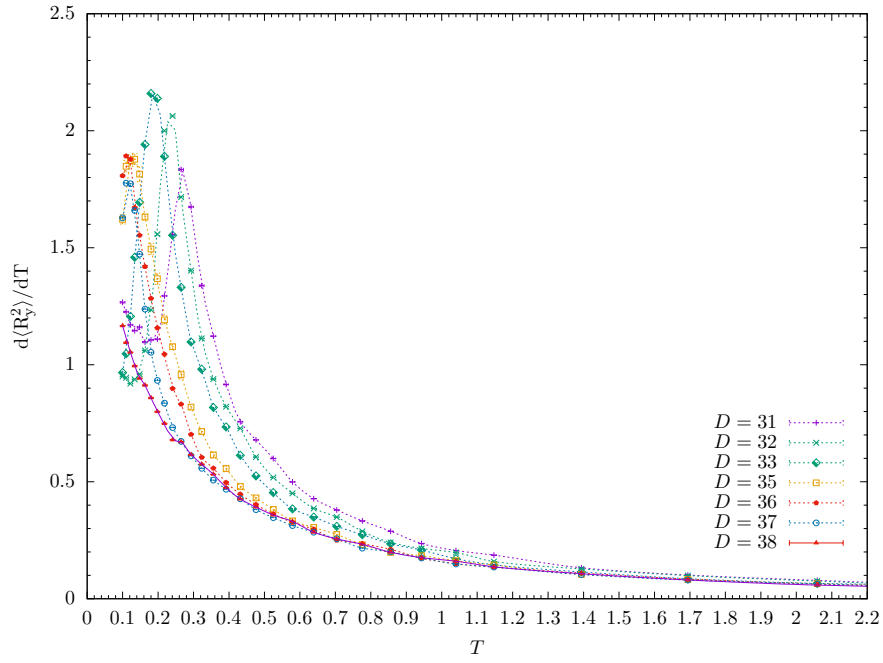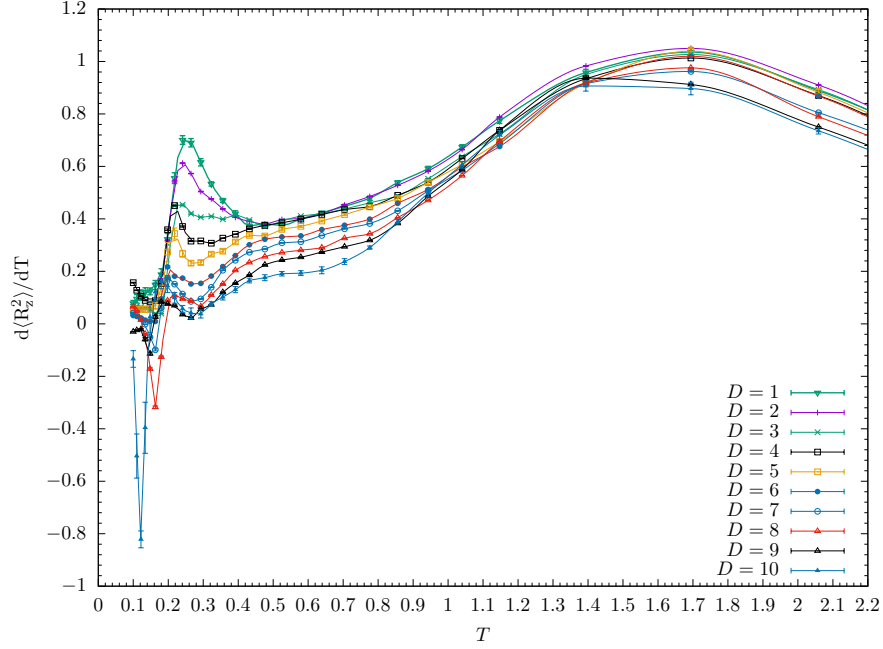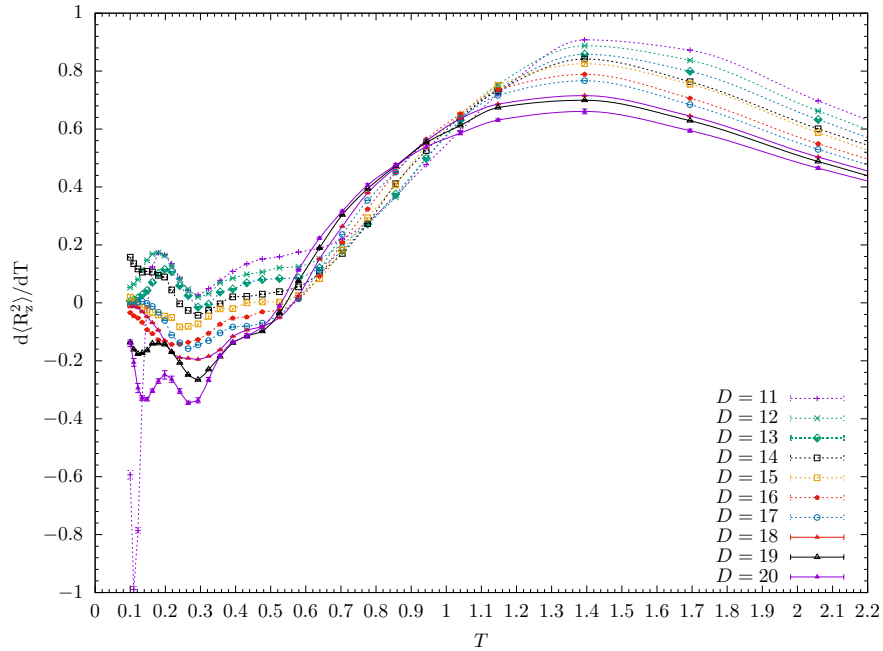Figure 6.21: $d\langle R_z^2\rangle/dT$ vs. T curves for $D \in [21, 30]$.

Figure 6.22: $d\langle R_z^2\rangle/dT$ vs. T curves for $D \in [31, 38]$.

88

The $d\langle z\rangle/dT$ vs. T curves at various D values are given in Figures 6.23 - 6.26. The patterns of these curves should match those of the $d\langle R_z^2\rangle/dT$ curves. The differences between them are the systematic error.

Figure 6.23: $d\langle z\rangle/dT$ vs. T curves for $D \in [1, 10]$.

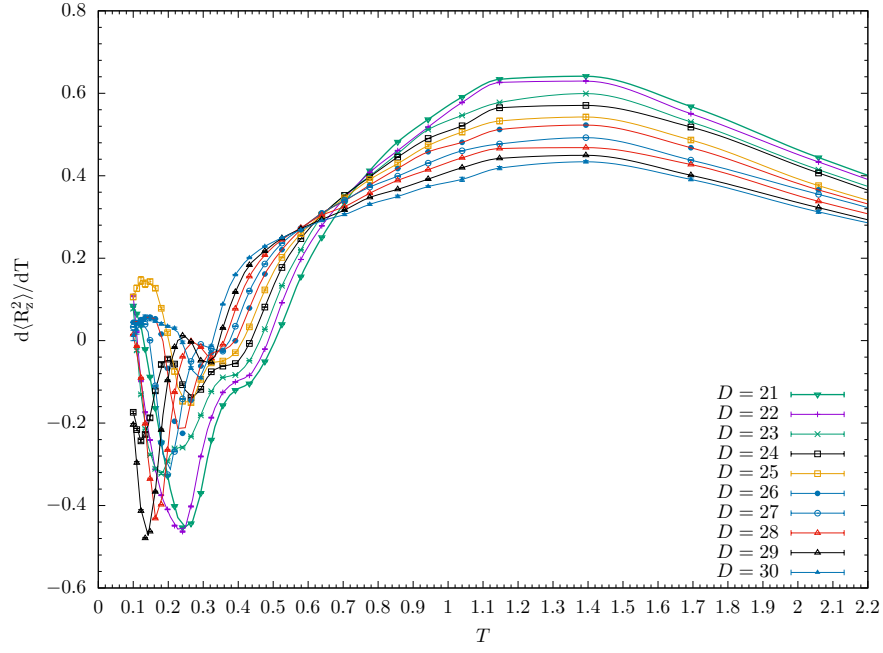Figure 6.24: $d\langle z\rangle/dT$ vs. T curves for $D \in [11, 20]$.
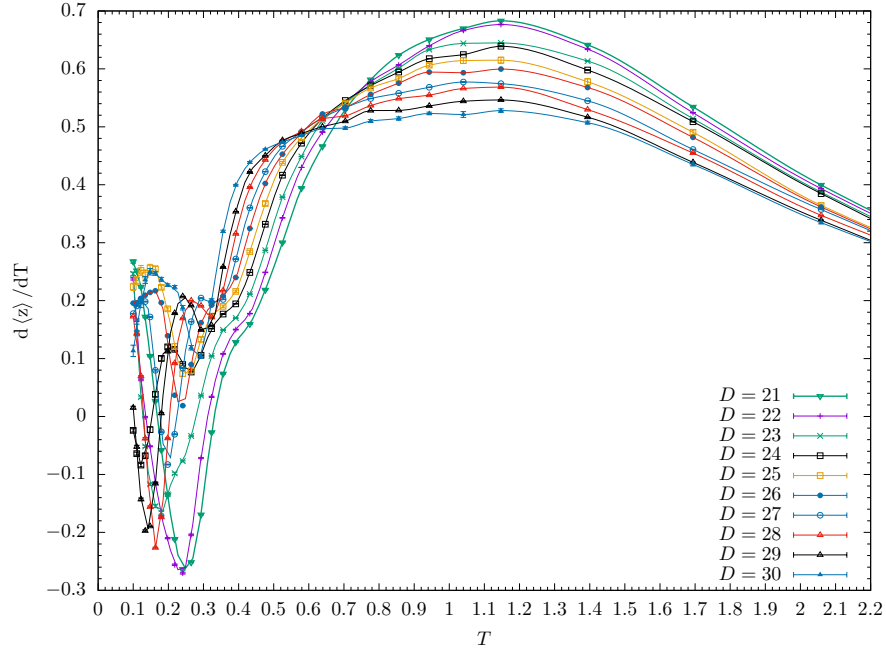
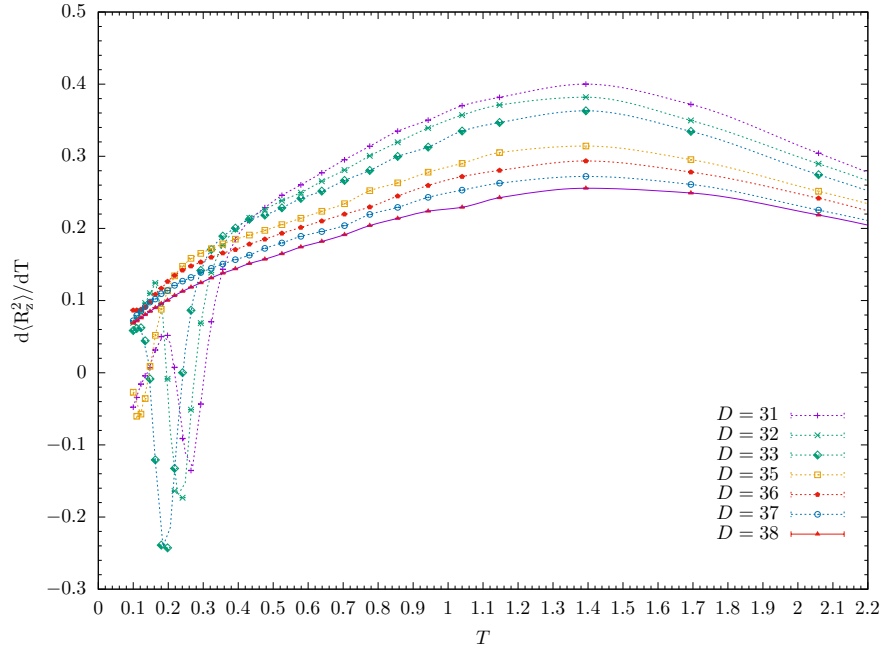Figure 6.25: $d\langle z\rangle/dT$ vs. T curves for $D \in [21, 30]$.
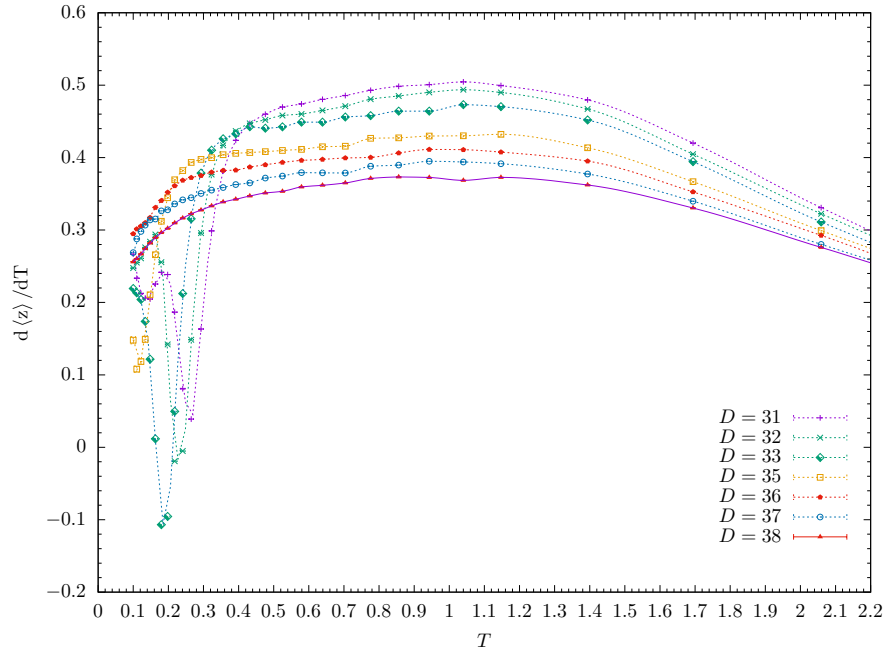
Figure 6.26: $d\langle z\rangle/dT$ vs. T curves for $D \in [31, 38]$.

The $d\langle N_m \rangle /dT$ vs. T curves at various D values are given in Figures 6.27 - 6.30. The patterns of these curves should match those of previous curves. The signals of the $d\langle N_m \rangle /dT$ curves can be used to identify the "melting" transitions.
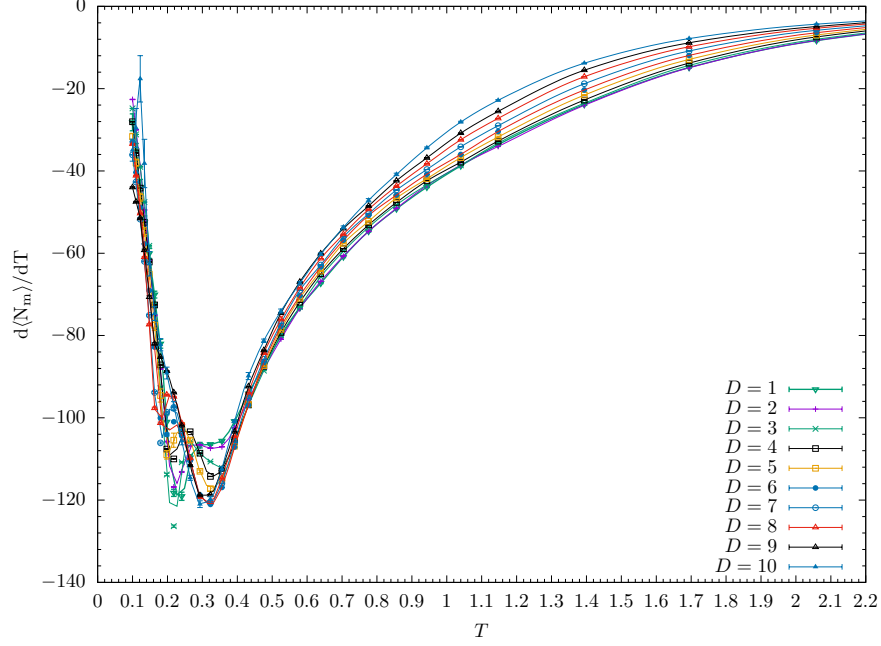
Figure 6.27: $d\langle N_m \rangle /dT$ vs. T curves for $D \in [1, 10]$.
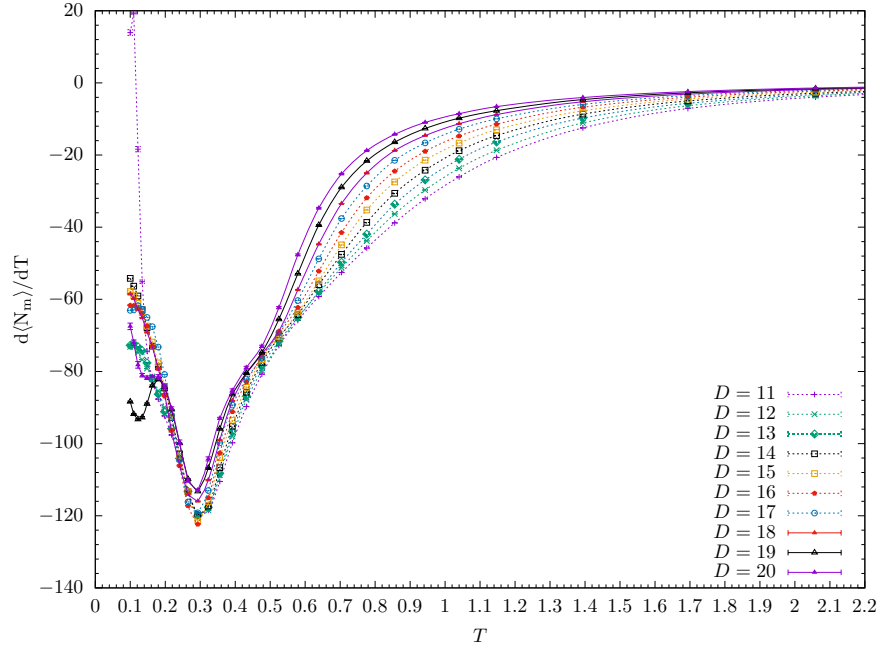
Figure 6.28: $d\langle N_m \rangle /dT$ vs. T curves for $D \in [11, 20]$.
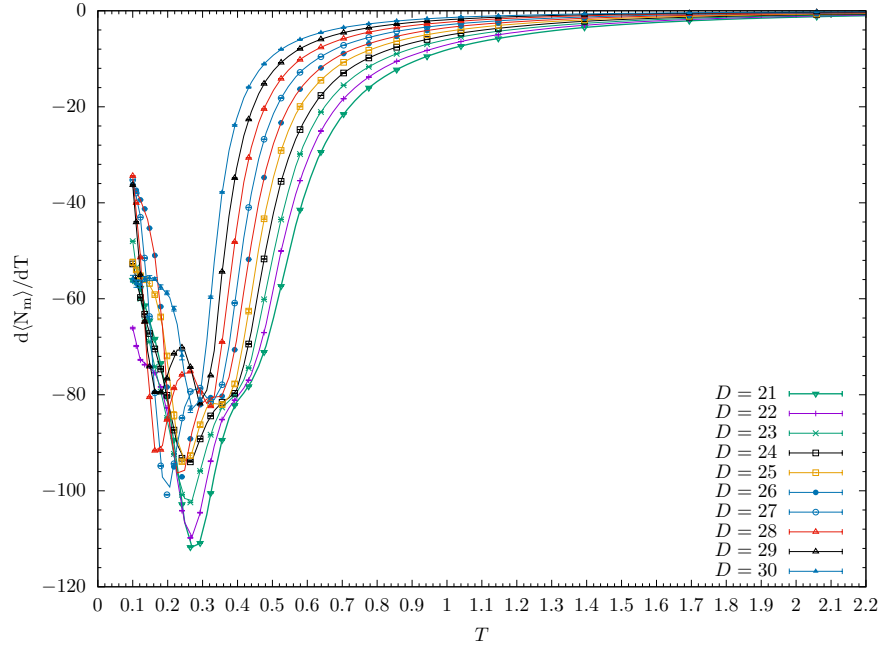
Figure 6.29: $d\langle N_m\rangle/dT$ vs. T curves for $D \in [21, 30]$.
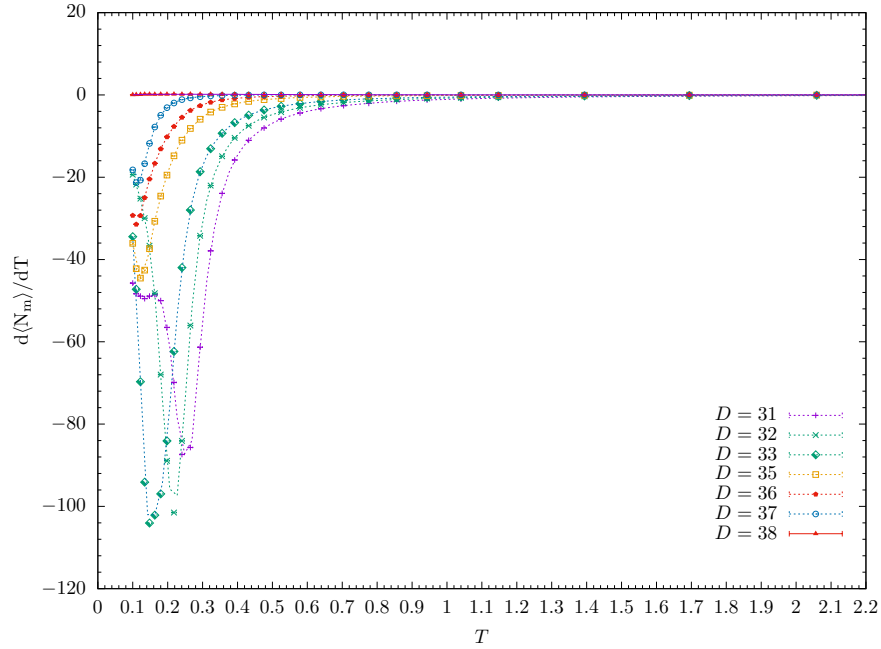
Figure 6.30: $d\langle N_m\rangle/dT$ vs. T curves for $D \in [31, 38]$.

The $d \langle N_s \rangle / dT$ vs. T curves at various D values are given in Figures 6.31 - 6.34. The signals of the $d \langle N_s \rangle / dT$ curves can provide important supportive information for the previous curves.
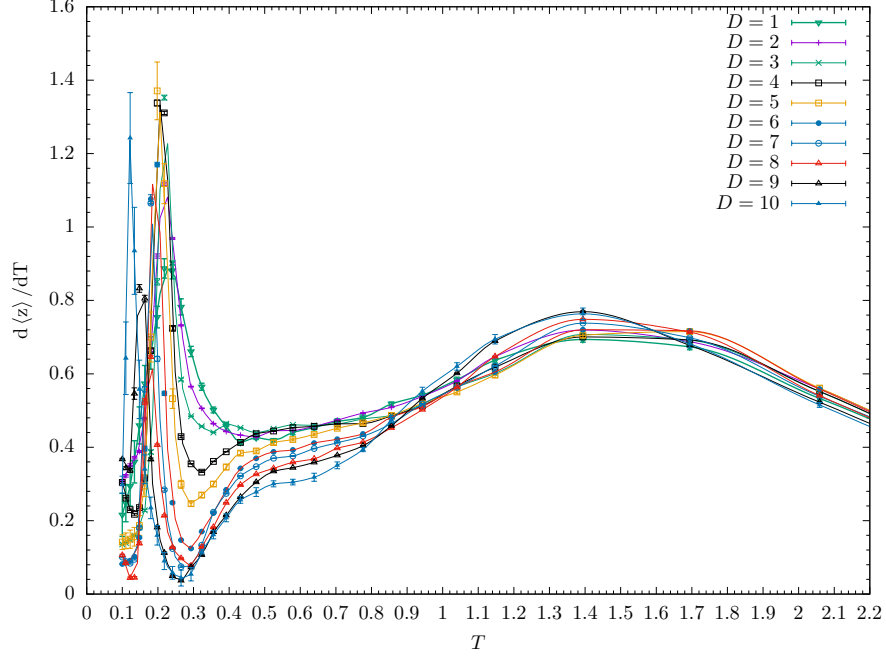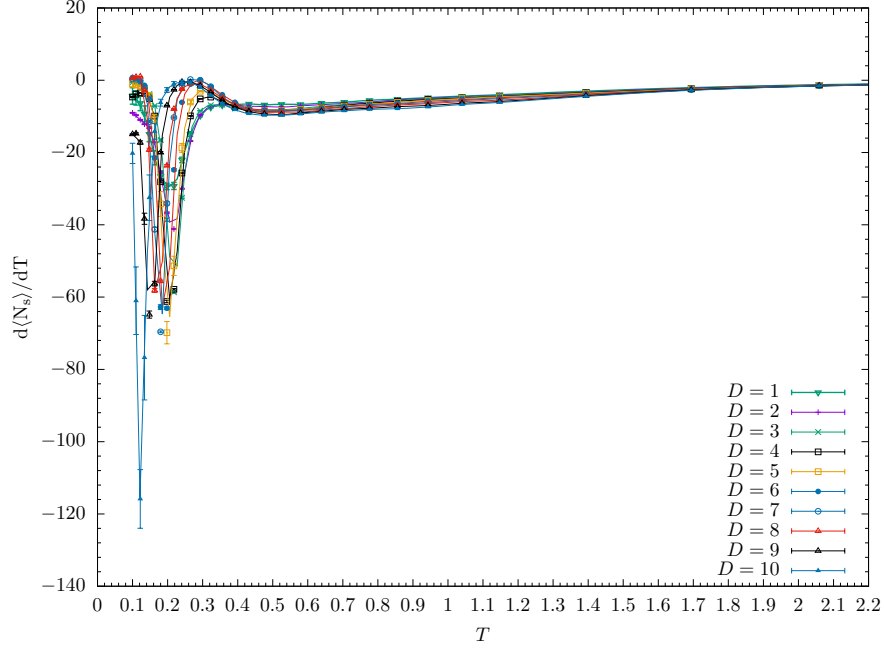

Figure 6.31: $d \langle N_s \rangle / dT$ vs. T curves for $D \in [1, 10]$.
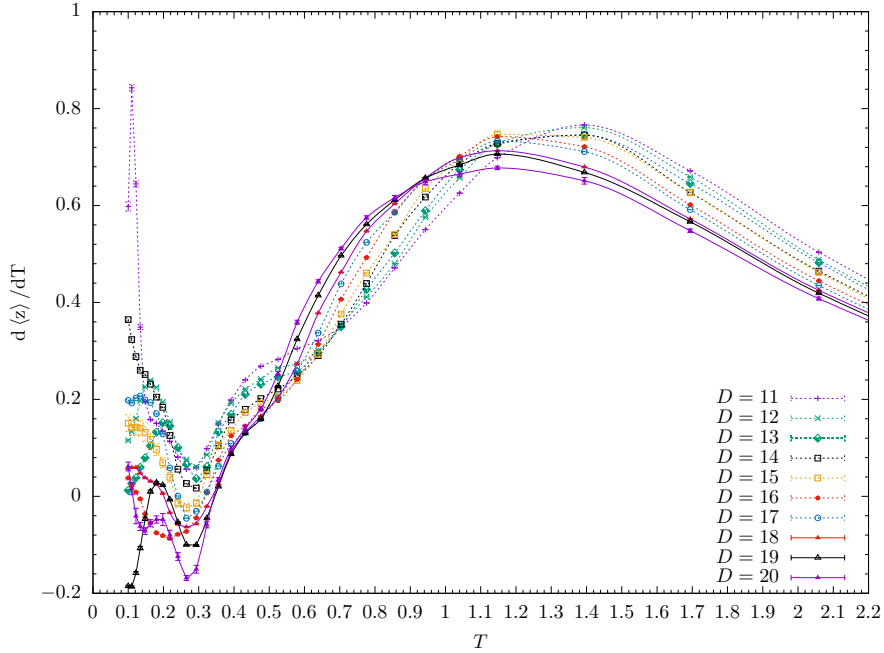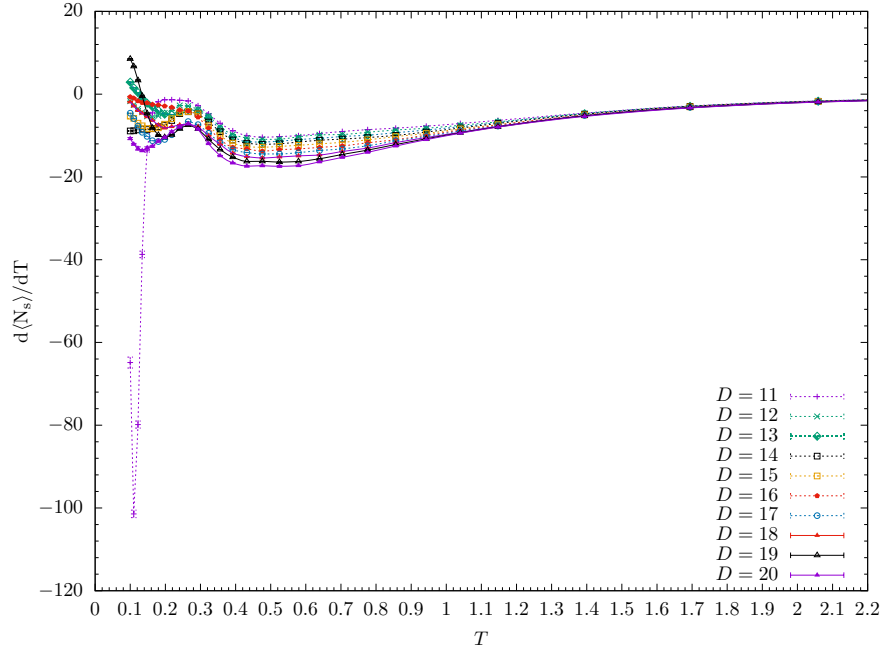

Figure 6.32: $d \langle N_s \rangle / dT$ vs. T curves for $D \in [11, 20]$.
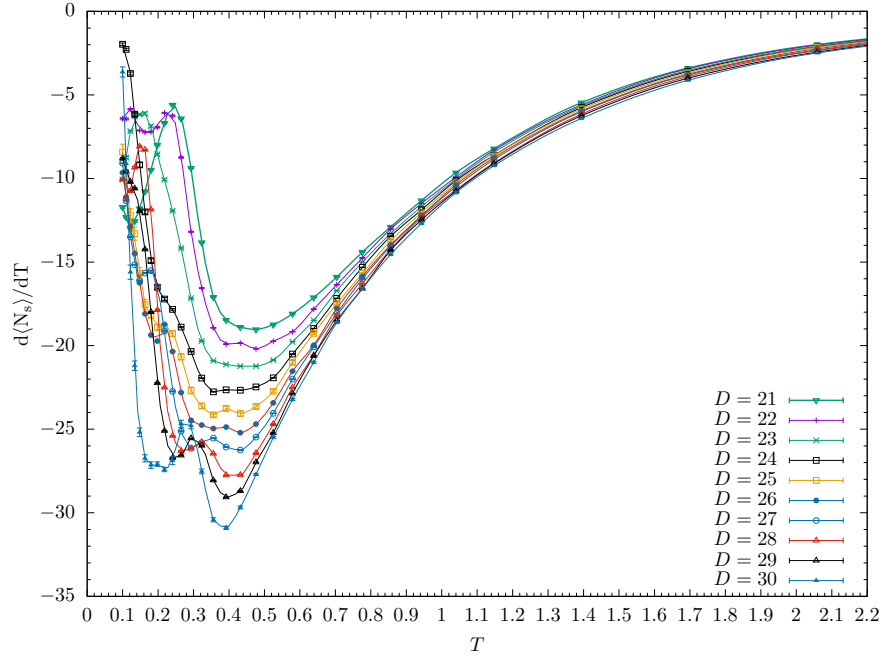
Figure 6.33: $d \langle N_s \rangle /dT$ vs. T curves for $D \in [21, 30]$.
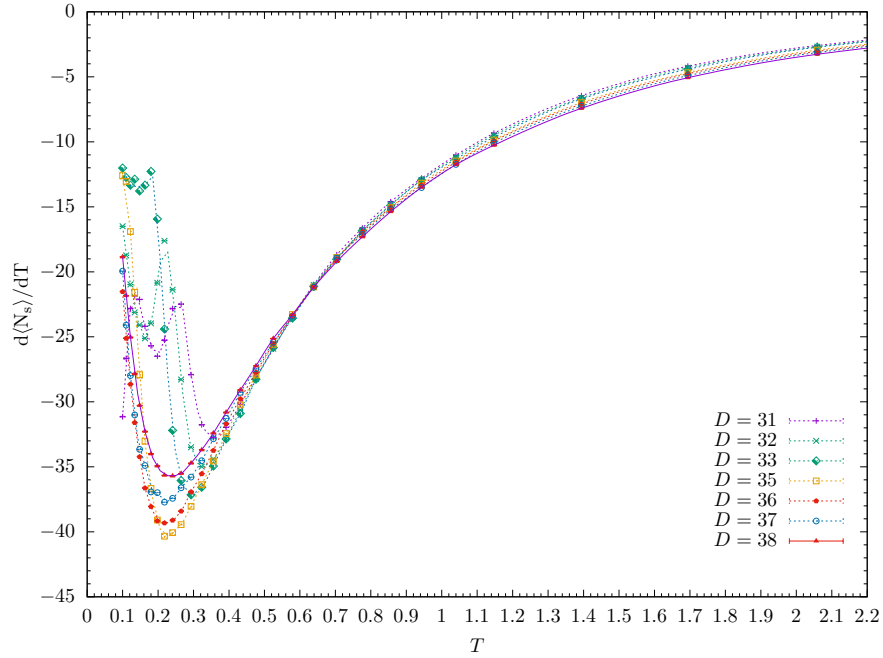
Figure 6.34: $d \langle N_s \rangle /dT$ vs. T curves for $D \in [31, 38]$.

6.2.3 Phase Diagrams and Conformational Structures

Comprehensively considering all the results provided and verifying the structures of all possible corresponding phases, a T vs. D conformational phase diagram can finally be created and presented in Figure 6.35.
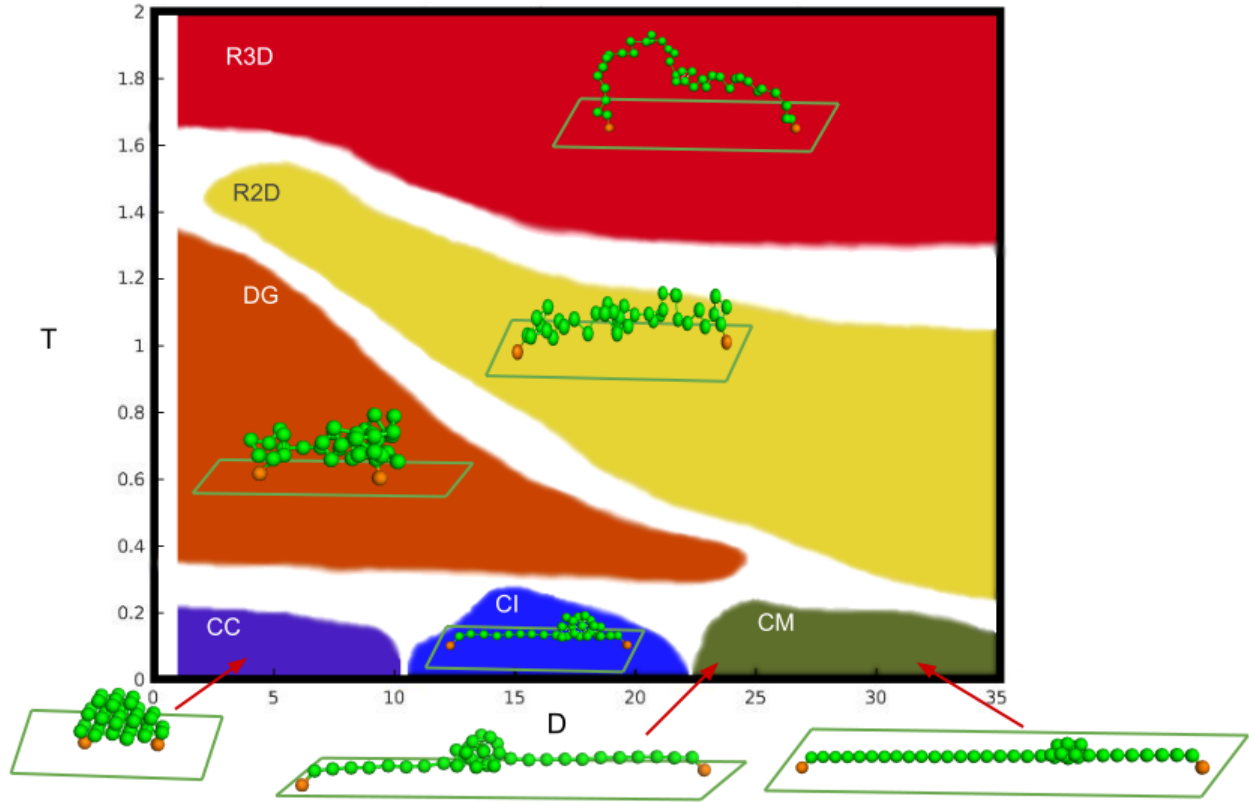
Figure 6.35: T vs. D conformational pseudophase diagram of polymer-substrate systems when $\epsilon_s = 1.0$.

Different phases are distinguished by different colors in Figure 6.35. The representative structures are also presented. The phases are:

1. CC: Compact Crystalline phase (layered structures);

2. CI: Compact Icosahedral phase (structures with an icosahedral core);

3. CM: Compact Mixed (mixed phase of layered and icosahedral structures);

4. R2D: 2D Random coil phase (structures of the polymer fluctuate mainly in a 2D plane paralleling to the substrate surface);

5. R3D: 3D Random coil phase;

6. DG: Droplet-like Globule phase (note that monomers of the polymer in this phase move less actively than those in the R2D phase);

The explanation of the phase diagram Figure 6.35 starts from the lower left corner and continues counterclockwise. When T is low ($0 < T < 0.2$) and D is small ($D < 11$), the doubly grafted polymer-and-substrate system is in its frozen compact crystalline phase, where the monomer-substrate interactions are dominant. Keep the temperature low and increase the end-to-end distance D, the minimum energy crystalline structures become compact structures with an icosahedral core. As D increases further ($22 < D < 39$), more monomers form the strand which is parallel to the x axis. Therefore, the number of the monomers forming complex structures reduces; so that competition between the monomer-monomer interactions and the monomer-substrate interactions becomes more sensible to the number of monomers forming complex structures. As D changes, the crystalline structures and structures with an imperfect icosahedral core occur alternately. Hence, this phase is named as compact mixed phase. When D is large ($22 < D < 39$), as temperature increases, the frozen compact structure "sublimes"; the polymer enters its 2D random coil phase, where the structures of the polymer fluctuate actively in a 2D plane paralleling to the substrate surface. This is due to the monomer-substrate interactions, since $\epsilon_{\mathrm{s}} = 1$. As temperature increases, the polymer enters its 3D random coil phase. Then, keeping the temperature fairly high and decreasing the end-to-end distance D significantly, the polymer goes through the Θ transition and the gas-like random coil structures become droplet-like globular structures. Finally, as the temperature decreases further, the polymer returns to its frozen compact conformational phase.

6.3 Summary

In this chapter, canonical quantities of polymer-substrate systems with attraction strength $\epsilon_s = 1.0$ are studied. The conformational phases and phase transitions are presented. The T vs. D phase diagram is constructed. Canonical analysis provides a simple and efficient approach to study the doubly grafted polymer-and-substrate systems. To reduce the size of the transition bands caused by finite size effect, microcanonical analysis method can be applied.

The D values are selected by an interval of 1 in this simulation. In the future, for a specific interested area on the phase diagram, higher resolutions of the D values can be selected to obtain more detailed information.

Chapter 7

Summary

The work presented in this dissertation has focused on the research of the conformational and thermodynamic properties of a finite-size doubly grafted polymer-substrate system using advanced parallel replica-exchange Monte Carlo method together with sophisticated canonical analysis methods. This doubly grafted polymer-substrate system is mimicking a short polymer with its both ends attached to an attractive substrate in surrounding solvents implicitly, e.g., a motor protein (Kinesin) walking on a microtubule, such a system could be spotted commonly in nanotechnology and biological systems. The polymer chain is simulated by a coarse-grained model such as a chain of monomers connected by unbreakable elastic bonds (FENE potential), with monomer-monomer interactions (Lennard-Jones potential) and monomer-substrate interactions (Lennard-Jones based interaction between monomers and substrate). The conditions of the surrounding solvents are simulated by the temperature values T. The strength of the interactions between monomers and substrate of different materials is simulated by the attraction strength ϵ_{s}. The coarse-grained models are not only able to simulate such systems efficiently but also can provide generic and statistical information about them. The results obtained for the coarse-grained models can be applied

to a broad range of physical systems if the energy and length scales of a Lennard-Jones interactions of such a system are measured.

In Chapter 5, canonical quantities of doubly grafted polymer-and-substrate systems with a fixed end-to-end distance $D = 25$ are studied and presented. Previously, others have studied the conformational structures for the free polymer-and-substrate systems and singly grafted polymer-and-substrate systems. Their results are summarized and compared with this work. The T vs. ϵ_s phase diagram is constructed and the representative structures for different phases are presented. The results obtained for the doubly grafted polymer-and-substrate systems in this work, in combination with previous studies for free and singly grafted polymer-and-systems, could provide an overall picture of the conformational studies of polymer substrate interactions. Based on the analysis of the energetic and structure quantities (and their thermal fluctuations), various conformational phases can be discerned, such as compact icosahedral structures (with different strand directions), compact crystalline structures (with different layer numbers), globular structures (2D and 3D), droplet structures (semispherical and spherical) and random coil structures.

Throughout this work, the geometric constraint of a polymer with its both ends fixed along the x direction breaks the isotropic properties of the free and singly grafted polymers and leads to different conformational structures. The potentials in the energy model and the D values affect the number of the monomers involved in the monomer-substrate interactions and the monomer-monomer interactions. Hence, the conformational phases and phase transitions of doubly grafted polymers with different D values are of interest.

In Chapter 6, canonical quantities of polymer-substrate systems with attraction strength $\epsilon_s = 1.0$ are studied. The conformational phases and phase transitions are presented. The T vs. D phase diagram is constructed. Based on the analysis of the energetic and structural quantities (and their thermal fluctuations), various conformational phases can be discerned for various D values. When T is low, the doubly grafted polymer-and-substrate system is

in its frozen compact phases, e.g., compact crystalline phase ($D < 11$), compact icosahedral phase ($11 < D < 22$) and compact mixed phase ($22 < D < 39$). As temperature increases, its frozen compact structures could become droplet-like globular structures for smaller D values; or become 2D random coil structures for larger D values. As temperature increases further, the polymer enters its 3D Random coil phase. The conformational transitions of the minimum energy structures reflect the competition between the monomer-monomer interactions and the monomer-substrate interactions as D changes. The results of this work could provide a thermodynamic perspective for understanding the native end-to-end distances of certain doubly grafted polymers in nanotechnology and biological systems, e.g., the step length of a protein walker.

Based on the procedures constructed in this dissertation, conformational phases and phase transitions of doubly grafted polymer-and-substrate systems with various D, ϵ_s and T values can be obtained. For related future work, the research could be further improved: 1) Microcanonical analysis method can be used to reduce the size of the transition bands. Therefore, more accurate phase diagrams can be obtained and the understanding of the phases and the phase transitions will be more clear. 2) The conformational analysis is time consuming and labor intensive. Automatic computerized methods for classifying the conformations are highly desired, e.g., machine learning classification methods. 3) If applying some modifications to such a model, more real world applications can be simulated. For example, if applying random walk processes to the positions of the polymer ends which are anchored onto the substrate and allowing the ends to move in the substrate surface, a walking protein motor can be simulated.

Bibliography

[1] T. T. Goodrich, A. W. Wark, R. M. Corn, and H. J. Lee, "Surface plasmon resonance imaging measurements of protein interactions with biopolymer microarrays," in *New and Emerging Proteomic Techniques*, pp. 113–130, Physical Review Letters, 2006.

[2] K. Werner, "Cryo-em enters a new era," *eLife*, vol. 3, 2014.

[3] J. Marko, E. Siggia, S. Smith, and C. Bustamante, "Entropic elasticity of λ-phage dna," *Science*, vol. 265, pp. 1599–1600, 1994.

[4] A. Ashkin, "Optical trapping and manipulation of neutral particles using lasers," *Proceedings of the National Academy of Sciences*, vol. 94, no. 10, pp. 4853–4860, 1997.

[5] K. Whitley, M. Comstock, and Y. Chemla, "High-resolution optical tweezers combined with single-molecule confocal microscopy," *Methods in Enzymology*, vol. 582, pp. 137–169, 2017.

[6] B. A. Berghuis, M. Köber, T. van Laar, and N. H. Dekker, "High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers," *Methods*, vol. 105, pp. 90–98, 2016.

[7] M. Bachmann, *Thermodynamics and statistical mechanics of macromolecular systems.* Cambridge University Press, 2014.

[8] T. E. Gartner III and A. Jayaraman, "Modeling and simulations of polymers: a roadmap," *Macromolecules*, vol. 52, no. 3, pp. 755–786, 2019.

[9] J. R. Ray, "Microcanonical ensemble monte carlo method," *Physical Review A*, vol. 44, no. 6, p. 4061, 1991.

[10] C. Junghans, M. Bachmann, and W. Janke, "Microcanonical analyses of peptide aggregation processes," *Physical Review Letters*, vol. 97, no. 21, p. 218103, 2006.

[11] M. Möddel, W. Janke, and M. Bachmann, "Systematic microcanonical analyses of polymer adsorption transitions," *Physical Chemistry Chemical Physics*, vol. 12, no. 37, pp. 11548–11554, 2010.

[12] W. Janke, "Multicanonical monte carlo simulations," *Physica A: Statistical Mechanics and its Applications*, vol. 254, no. 1-2, pp. 164–178, 1998.

[13] B. A. Berg, "Introduction to multicanonical monte carlo simulations," *Fields Inst. Commun*, vol. 26, no. 1, pp. 1–24, 2000.

[14] U. H. Hansmann, Y. Okamoto, and F. Eisenmenger, "Molecular dynamics, langevin and hydrid monte carlo simulations in a multicanonical ensemble," *Chemical Physics Letters*, vol. 259, no. 3-4, pp. 321–330, 1996.

[15] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 86, no. 10, p. 2050, 2001.

[16] D. Seaton, T. Wüst, and D. Landau, "A wang–landau study of the phase transitions in a flexible homopolymer," *Computer Physics Communications*, vol. 180, no. 4, pp. 587–589, 2009.

[17] D. Landau, S.-H. Tsai, and M. Exler, "A new approach to monte carlo simulations in statistical physics: Wang-landau sampling," *American Journal of Physics*, vol. 72, no. 10, pp. 1294–1302, 2004.

[18] A. C. Farris, D. T. Seaton, and D. P. Landau, "Effects of lattice constraints in coarse-grained protein models," *The Journal of Chemical Physics*, vol. 154, no. 8, p. 084903, 2021.

[19] R. H. Swendsen and J.-S. Wang, "Replica monte carlo simulation of spin-glasses," *Physical Review Letters*, vol. 57, no. 21, p. 2607, 1986.

[20] C. J. Geyer *et al.*, "Computing science and statistics: Proceedings of the 23rd symposium on the interface," *American Statistical Association, New York*, vol. 156, 1991.

[21] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Physical Chemistry Chemical Physics*, vol. 7, no. 23, pp. 3910–3916, 2005.

[22] K. Hukushima and K. Nemoto, "Exchange monte carlo method and application to spin glass simulations," *Journal of the Physical Society of Japan*, vol. 65, no. 6, pp. 1604–1608, 1996.

[23] M. Möddel, M. Bachmann, and W. Janke, "Conformational mechanics of polymer adsorption transitions at attractive substrates," *The Journal of Physical Chemistry B*, vol. 113, no. 11, pp. 3314–3323, 2009.

[24] K. Binder, J. Baschnagel, M. Müller, W. Paul, and F. Rampf, "Simulation of phase transitions of single polymer chains: Recent advances," in *Macromolecular Symposia*, vol. 237, pp. 128–138, Wiley Online Library, 2006.

[25] G. Jaeger, "The ehrenfest classification of phase transitions: introduction and evolution," *Archive for History of Exact Sciences*, vol. 53, no. 1, pp. 51–81, 1998.

[26] A. Tiessen, P. Pérez-Rodríguez, and L. J. Delaye-Arredondo, "Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes," *BMC Research Notes*, vol. 5, no. 1, pp. 1–23, 2012.

[27] H. Gould and J. Tobochnik, *Statistical and thermal physics: with computer applications*. Princeton University Press, 2010.

[28] W. Janke, "Canonical versus microcanonical analysis of first-order phase transitions," *Nuclear Physics B-Proceedings Supplements*, vol. 63, no. 1-3, pp. 631–633, 1998.

[29] T. Koci, *Structure forming processes in mesoscopic polymer systems*. PhD thesis, University of Georgia, 2016.

[30] C. Junghans, M. Bachmann, and W. Janke, "Thermodynamics of peptide aggregation processes: An analysis from perspectives of three statistical ensembles," *The Journal of Chemical Physics*, vol. 128, no. 8, p. 02B628, 2008.

[31] S. Schnabel, D. T. Seaton, D. P. Landau, and M. Bachmann, "Microcanonical entropy inflection points: Key to systematic understanding of transitions in finite systems," *Physical Review E*, vol. 84, no. 1, p. 011127, 2011.

[32] K. Qi and M. Bachmann, "Classification of phase transitions by microcanonical inflection-point analysis," *Physical Review Letters*, vol. 120, no. 18, p. 180601, 2018.

[33] K. Sitarachu, R. Zia, and M. Bachmann, "Exact microcanonical statistical analysis of transition behavior in ising chains and strips," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2020, no. 7, p. 073204, 2020.

[34] D. Aierken and M. Bachmann, "Comparison of conformational phase behavior for flexible and semiflexible polymers," *Polymers*, vol. 12, no. 12, p. 3013, 2020.

[35] Y.-Z. Xu, C. H. Yeung, H.-J. Zhou, and D. Saad, "Entropy inflection and invisible low-energy states: Defensive alliance example," *Physical Review Letters*, vol. 121, no. 21, p. 210602, 2018.

[36] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, *et al.*, "Atomic-level characterization of the structural dynamics of proteins," *Science*, vol. 330, no. 6002, pp. 341–346, 2010.

[37] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, "Atomic-level characterization of protein–protein association," *Proceedings of the National Academy of Sciences*, vol. 116, no. 10, pp. 4244–4249, 2019.

[38] C. A. Hunter and H. L. Anderson, "What is cooperativity?," *Angewandte Chemie International Edition*, vol. 48, no. 41, pp. 7488–7499, 2009.

[39] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[40] Y. Fang, "Gibbs free energy formula for protein folding," *Thermodynamics—Fundamentals and Its Application in Science, Intech, Rijeka*, pp. 47–82, 2012.

[41] A. C. K. Farris, *Extended Ensemble Monte Carlo Simulations of Coarse-grained Protein Models*. PhD thesis, University of Georgia, 2019.

[42] K. A. Dill, "Theory for the folding and stability of globular proteins," *Biochemistry*, vol. 24, no. 6, pp. 1501–1509, 1985.

[43] K. A. Dill, S. Bromberg, K. Yue, H. S. Chan, K. M. Ftebig, D. P. Yee, and P. D. Thomas, "Principles of protein folding—a perspective from simple exact models," *Protein Science*, vol. 4, no. 4, pp. 561–602, 1995.

[44] R. Agarwala, S. Batzoglou, V. Dančik, S. E. Decatur, S. Hannenhalli, M. Farach, S. Muthukrishnan, and S. Skiena, "Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the hp model," *Journal of Computational Biology*, vol. 4, no. 3, pp. 275–296, 1997.

[45] H. Nymeyer, A. E. García, and J. N. Onuchic, "Folding funnels and frustration in off-lattice minimalist protein landscapes," *Proceedings of the National Academy of Sciences*, vol. 95, no. 11, pp. 5921–5928, 1998.

[46] A. Irbäck, C. Peterson, F. Potthast, and O. Sommelius, "Local interactions and protein folding: A three-dimensional off-lattice approach," *The Journal of Chemical Physics*, vol. 107, no. 1, pp. 273–282, 1997.

[47] K. F. Lau and K. A. Dill, "A lattice statistical mechanics model of the conformational and sequence spaces of proteins," *Macromolecules*, vol. 22, no. 10, pp. 3986–3997, 1989.

[48] G. Shi, T. Wuest, Y. W. Li, and D. P. Landau, "Protein folding of the hop model: a parallel wang—landau study," in *Journal of Physics: Conference Series*, vol. 640, p. 012017, IOP Publishing, 2015.

[49] F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, "Toy model for protein folding," *Physical Review E*, vol. 48, no. 2, p. 1469, 1993.

[50] A. Irbäck and F. Potthast, "Studies of an off-lattice model for protein folding: Sequence dependence and improved sampling at finite temperature," *The Journal of Chemical Physics*, vol. 103, no. 23, pp. 10298–10305, 1995.

[51] M. W. Matsen, "Phase behavior of block copolymer/homopolymer blends," *Macromolecules*, vol. 28, no. 17, pp. 5765–5773, 1995.

[52] R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, and J. M. Thornton, "Pdbsum: Structural summaries of pdb entries," *Protein Science*, vol. 27, no. 1, pp. 129–134, 2018.

[53] W. L. DeLano *et al.*, "Pymol: An open-source molecular graphics tool," *CCP4 Newsletter on Protein Crystallography*, vol. 40, no. 1, pp. 82–92, 2002.

[54] H. Matsumura, K. Shiomi, A. Yamamoto, Y. Taketani, N. Kobayashi, T. Yoshizawa, S.-i. Tanaka, H. Yoshikawa, M. Endo, and H. Fukayama, "Hybrid rubisco with complete replacement of rice rubisco small subunits by sorghum counterparts confers c4 plant-like high catalytic activity," *Molecular Plant*, vol. 13, no. 11, pp. 1570–1581, 2020.

[55] R. B. Martin, "Free energies and equilibria of peptide bond hydrolysis and formation," *Biopolymers: Original Research on Biomolecules*, vol. 45, no. 5, pp. 351–353, 1998.

[56] J. E. Lennard-Jones, "Cohesion," *Proceedings of the Physical Society (1926-1948)*, vol. 43, no. 5, p. 461, 1931.

[57] M. J. Williams and M. Bachmann, "Stabilization of helical macromolecular phases by confined bending," *Physical Review Letters*, vol. 115, no. 4, p. 048301, 2015.

[58] K. Kremer and G. S. Grest, "Dynamics of entangled linear polymer melts: A molecular-dynamics simulation," *The Journal of Chemical Physics*, vol. 92, no. 8, pp. 5057–5086, 1990.

[59] Y. Dai and M. Bachmann, "Bending-stiffness dependent generic structural transitions of helical polymers," in *Journal of Physics: Conference Series*, vol. 1012, p. 012007, IOP Publishing, 2018.

[60] J. J. Sakurai and E. D. Commins, "*Modern quantum mechanics, revised edition*," 1995.

[61] L. D. Landau, *The classical theory of fields*, vol. 2. Elsevier, 2013.

[62] E. E. Anderson, " *Modern physics and quantum mechanics*," 1971.

[63] S. T. Reeve and A. Strachan, "Error correction in multi-fidelity molecular dynamics simulations using functional uncertainty quantification," *Journal of Computational Physics*, vol. 334, pp. 207–220, 2017.

[64] L. Kulakova, G. Arampatzis, P. Angelikopoulos, P. Hadjidoukas, C. Papadimitriou, and P. Koumoutsakos, "Data driven inference for the repulsive exponent of the lennard-jones potential in molecular dynamics simulations," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[65] D. Ceperley, M. Kalos, and J. L. Lebowitz, "Computer simulation of the dynamics of a single polymer chain," *Physical Review Letters*, vol. 41, no. 5, p. 313, 1978.

[66] K. Qi and M. Bachmann, "Autocorrelation study of the θ transition for a coarse-grained polymer model," *The Journal of Chemical Physics*, vol. 141, no. 7, p. 074101, 2014.

[67] M. Möddel, W. Janke, and M. Bachmann, "Comparison of the adsorption transition for grafted and nongrafted polymers," *Macromolecules*, vol. 44, no. 22, pp. 9013–9019, 2011.

[68] D. Ceperley, M. Kalos, and J. L. Lebowitz, "Computer simulation of the static and dynamic properties of a polymer chain," *Macromolecules*, vol. 14, no. 5, pp. 1472–1479, 1981.

[69] R. C. Armstrong, S. K. Gupta, and O. Basaran, "Conformational changes of macromolecules in transient elongational flow," *Polymer Engineering & Science*, vol. 20, no. 7, pp. 466–472, 1980.

[70] D. P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics*. Cambridge University Press, 2014.

[71] S. Schnabel, W. Janke, and M. Bachmann, "Advanced multicanonical monte carlo methods for efficient simulations of nucleation processes of polymers," *Journal of Computational Physics*, vol. 230, no. 12, pp. 4454–4465, 2011.

[72] G. Marsaglia and A. Zaman, "A new class of random number generators," *The Annals of Applied Probability*, pp. 462–480, 1991.

[73] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[74] M. Newman and G. Barkema, *Monte carlo methods in statistical physics chapter 1-4*, vol. 24. Oxford University Press: New York, USA, 1999.

[75] W. Janke, "Statistical analysis of simulations: Data correlations and error estimation," *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, vol. 10, pp. 423–445, 2002.

[76] B. Efron, *The jackknife, the bootstrap and other resampling plans*. SIAM, 1982.

[77] R. G. Miller, "The jackknife-a review," *Biometrika*, vol. 61, no. 1, pp. 1–15, 1974.

[78] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[79] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical Science*, pp. 54–75, 1986.

[80] A. Milchev and K. Binder, "Static and dynamic properties of adsorbed chains at surfaces: Monte carlo simulation of a bead-spring model," *Macromolecules*, vol. 29, no. 1, pp. 343–354, 1996.

[81] K. Binder and A. Milchev, "Polymer brushes on flat and curved surfaces: How computer simulations can help to test theories and to interpret experiments," *Journal of Polymer Science Part B: Polymer Physics*, vol. 50, no. 22, pp. 1515–1555, 2012.

[82] M. J. Williams and M. Bachmann, "The effect of surface adsorption on tertiary structure formation in helical polymers," *The Journal of Chemical Physics*, vol. 147, no. 2, p. 024902, 2017.

[83] P.-G. De Gennes and P.-G. Gennes, *Scaling concepts in polymer physics*. Cornell University Press, 1979.

[84] M. Doi, *Introduction to polymer physics*. Oxford University Press, 1996.

[85] G. R. Strobl and G. R. Strobl, *The physics of polymers*, vol. 2. Springer, 1997.

[86] M. Rubinstein, R. H. Colby, *et al.*, *Polymer physics*, vol. 23. Oxford University Press New York, 2003.

[87] N. Madras and G. Slade, *The self-avoiding walk*. Springer Science & Business Media, 2013.

[88] C. Vanderzande, *Lattice models of polymers*. No. 11, Cambridge University Press, 1998.

[89] G. Rychlewski and S. G. Whittington, "Self-avoiding walks and polymer adsorption: low temperature behaviour," *Journal of Statistical Physics*, vol. 145, no. 3, pp. 661–668, 2011.

[90] A. Sokal, "Monte carlo methods in statistical mechanics: foundations and new algorithms," in *Functional integration*, pp. 131–192, Springer, 1997.

[91] E. Eisenriegler, K. Kremer, and K. Binder, "Adsorption of polymer chains at surfaces: Scaling and monte carlo analyses," *The Journal of Chemical Physics*, vol. 77, no. 12, pp. 6296–6320, 1982.

[92] A. N. Rissanou, S. H. Anastasiadis, and I. A. Bitsanis, "A monte carlo study of the coil-to-globule transition of model polymer chains near an attractive surface," *Journal of Polymer Science Part B: Polymer Physics*, vol. 47, no. 24, pp. 2462–2476, 2009.

[93] S. Metzger, M. Müller, K. Binder, and J. Baschnagel, "Surface excess in dilute polymer solutions and the adsorption transition versus wetting phenomena," *The Journal of Chemical Physics*, vol. 118, no. 18, pp. 8489–8499, 2003.

[94] S. Metzger, M. Müller, K. Binder, and J. Baschnagel, "Adsorption transition of a polymer chain at a weakly attractive surface: Monte carlo simulation of off-lattice models," *Macromolecular Theory and Simulations*, vol. 11, no. 9, pp. 985–995, 2002.

[95] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[96] D. J. Wales and J. P. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," *The Journal of Physical Chemistry A*, vol. 101, no. 28, pp. 5111–5116, 1997.

[97] S. Schnabel, M. Bachmann, and W. Janke, "Elastic lennard-jones polymers meet clusters: Differences and similarities," *The Journal of Chemical Physics*, vol. 131, no. 12, p. 124904, 2009.

[98] K. Qi, B. Liewehr, T. Koci, B. Pattanasiri, M. J. Williams, and M. Bachmann, "Influence of bonded interactions on structural phases of flexible polymers," *The Journal of Chemical Physics*, vol. 150, no. 5, p. 054904, 2019.

[99] S. Schnabel, T. Vogel, M. Bachmann, and W. Janke, "Surface effects in the crystallization process of elastic flexible polymers," *Chemical Physics Letters*, vol. 476, no. 4-6, pp. 201–204, 2009.

[100] O. M. Elrad and M. F. Hagan, "Encapsulation of a polymer by an icosahedral virus," *Physical Biology*, vol. 7, no. 4, p. 045003, 2010.

[101] J. Gross, T. Neuhaus, T. Vogel, and M. Bachmann, "Effects of the interaction range on structural phases of flexible polymers," *The Journal of Chemical Physics*, vol. 138, no. 7, p. 074905, 2013.

[102] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nature Physics*, vol. 13, no. 5, pp. 431–434, 2017.

[103] T. Vogel and M. Bachmann, "Conformational phase diagram for polymers adsorbed on ultrathin nanowires," *Physical Review Letters*, vol. 104, no. 19, p. 198302, 2010.

[104] A. M. Samarakoon, K. Barros, Y. W. Li, M. Eisenbach, Q. Zhang, F. Ye, V. Sharma, Z. Dun, H. Zhou, S. A. Grigera, *et al.*, "Machine-learning-assisted insight into spin ice Dy(2)Ti(2)O(7)," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020.

[105] T. Kreer, S. Metzger, M. Müller, K. Binder, and J. Baschnagel, "Static properties of end-tethered polymers in good solution: A comparison between different models," *The Journal of Chemical Physics*, vol. 120, no. 8, pp. 4012–4023, 2004.

[106] B. Liewehr and M. Bachmann, "Smart polymeric recognition of a hexagonal monolayer," *EPL (Europhysics Letters)*, vol. 127, no. 6, p. 68003, 2019.

[107] A. Khokhlov, "Theory of the polymer chain collapse for the d-dimensional case," *Physica A: Statistical Mechanics and its Applications*, vol. 105, no. 1-2, pp. 357–362, 1981.