### GENERIC STRUCTURAL TRANSITIONS OF HELICAL POLYMERS WITH CONFINED TORSION STRENGTH

by

#### Yifan Dai

(Under the Direction of Michael Bachmann)

#### Abstract

 $\alpha$ -helices are secondary structures existing in most macromolecular systems such as proteins, DNA and RNA. The reasons behind the formation of structural phases are essential for understanding their biological functions. For this purpose, we performed parallel-tempering replica-exchange simulations of a coarse-grained polymer model. The tertiary folds composed of secondary structure segments are stabilized by the effects of bending and torsion. We systematically investigate the structural transitions by varying bending constraints at confined torsion strengths and construct the hyperphase diagrams. This study lends insight into the impact of competing bending and torsion effects and explains why polymers often exhibit helical structures.

INDEX WORDS: polymer, parallel-tempering, coarse-grained, helical structures, bending, torsion.

### Generic Structural Transitions of Helical Polymers with Confined Torsion Strength

by

#### Yifan Dai

B.S., Physics, Nanjing University of Information Science and Technology, 2009

M.S., Physics, Fudan University, 2012

M.S., Statistics, University of Georgia, 2015

M.S., Computer Science, 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

C2021

Yifan Dai

All Rights Reserved

### Generic Structural Transitions of Helical Polymers with Confined Torsion Strength

by

YIFAN DAI

Approved:

Major Professor: Michael Bachmann

Committee: David Landau Heinz-Bernd Schuttler

Electronic Version Approved:

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

## Acknowledgments

First, I would like to thank Budda, You warm my heart and show me the way and provide me with strength to overcome difficulties. With your blessings, I achieve tremedous heights in both spiritual and academic.

Second, I would like to thank my parents. Thank you for your unconditional love and supports. Your teachings make me strong.

Third, I would like to thank Dr.Bachmann, for your patience, motivation and immense knowledge. I could not have imagined how difficult would it be without your directions and continuous supports. I would also thank Dr.Landau and Dr.Schuttler for being my committee members. I had very useful discussion with Dr.Williams.

Last but not least, I would like to thank my wife Yao and my sons Leon and Maxwell. Thank you for bringing me everthing.

# Contents

| 1        | Intr                   | oduction  | 1  |
|----------|------------------------|---|----|
| <b>2</b> | Methodology            |   | 6  |
|          | 2.1                    | Model selection   | 6  |
|          | 2.2                    | Sampling and simulation   | 12 |
|          | 2.3                    | Canonical analysis  | 16 |
|          | 2.4                    | $q_1$ and $q_2$ order parameter space   | 20 |
|          | 2.5                    | Binning Jackknife error analysis  | 23 |
| 3        | Results and discussion |   | 25 |
|          | 3.1                    | Stability of polymers with weak torsion restraint                             | 27 |
|          | 3.2                    | Phase behavior of helical polymers with moderate torsion restraint            | 36 |
|          | 3.3                    | The general structural transitions of helical macromolecules with strong tor- |    |
|          |                        | sion restraint  | 43 |
|          | 3.4                    | Comparison  | 52 |
| 4        | Con                    | clusion and outlook   | 54 |
| <b>5</b> | App                    | pendix  | 57 |
|          | 5.1                    | Noticeable Issues in MPI Programming  | 62 |

### Bibliography

# List of Figures

| 1.1 | A peptide bond is the bond between an $NH_2$ group and $COOH$ group                         | 1  |
|-----|---|----|
| 1.2 | The structure of chymotrypsin in which the labelled amino acids form the                    |    |
|     | active site of the enzyme. Chymotrypsin has the tertiary structure. $\ldots$ .              | 2  |
| 1.3 | $\alpha\beta$ hemoglobin has the quarternary structure composed of four units. $\alpha$ and |    |
|     | $\beta$ subunits are in red and blue, respectively, and the iron-containing heme            |    |
|     | groups in green.  | 2  |
| 1.4 | Four levels of protein hierarchy. These are primary structure, secondary struc-             |    |
|     | ture, teritiary structure and quarternary structure   | 3  |
| 1.5 | Atomic model (all-atom model). Hydrogen bonds are represented by yellow                     |    |
|     | dashed lines in this $\alpha$ -helix structure. Oxygen atoms, nitrogen atoms, hydro-        |    |
|     | gen atoms and carbon atoms are represented in red, blue, white and green                    |    |
|     | respectively. $[9]$   | 4  |
| 2.1 | Coarse-grained representation of an $\alpha$ -helix. All different amino acids are          |    |
|     | considered as identical monomers  | 7  |
| 2.2 | The nonbonded interaction is represented by the dimensionless shifted Lennard-              |    |
|     | Jones (LJ) potential. For bonded interaction, the FENE potential is used.                   | 8  |
| 2.3 | The dimensionless shifted Lennard-Jones (LJ) potential with cutoff                          | 9  |
| 2.4 | The finitely extensible nonlinear elastic (FENE) potential                                  | 10 |
| 2.5 | The bending angle is denoted by $\theta$ (blue) and the torsion angle by $\tau$ (red).      | 11 |

| 2.6  | Left side shows the local update. The randomly picked monomer is in red  |    |
|------|--|----|
|      | and is updated within a cubic box. An example for a torsion update by  |    |
|      | $45^0$ is shown on the right. The randomly picked bond is in blue. The three   |    |
|      | monomers in light green show the new positions after the $45^0$ torsion update.  | 13 |
| 2.7  | The mechanism for the replica-exchange between neighbor threads. $\ldots$ .  | 15 |
| 2.8  | Schematic plot of the entropy $S$ of a system experiencing a first-order transition.   | 16 |
| 2.9  | Entropy S and heat capacity $C_v$ of a system at a second-order transition   | 17 |
| 2.10 | The energy histogram of a flexible polymer with 55 monomers near $T \sim 0.3$ .  |    |
|      | [a] with bonded LJ potential [b] without bonded LJ potential   | 18 |
| 2.11 | The heat capacity $C_{\rm V}$ and the fluctuation of the square radius of gyration   |    |
|      | $d\langle R_{\rm g}^2 \rangle/dT$ of a flexible polymer with 13 monomers   | 19 |
| 2.12 | The pair of order parameters $q_1$ and $q_2$   | 20 |
| 2.13 | The pair values of $(q_1,q_2)$ for [a] the single helix, [b] two-helix bundle and [c]  |    |
|      | three-helix bundle shown above are $(-5.31, -0.05)$ , $(-4.76, -1.61)$ and $(-4.24, -2.5)$ ,   |    |
|      | respectively.  | 21 |
| 2.14 | The pair values of $(q_1,q_2)$ for the single helix, two-helix bundle and three-helix  |    |
|      | bundle shown above are $(-5.31, -0.05)$ , $(-4.76, -1.61)$ and $(-4.24, -2.5)$ , respectively.   | 22 |
| 3.1  | Hyperphase diagrams of bending restrained $S_{\theta} = 200$ (left) and unrestrained   |    |
|      | $S_{\theta} = 0$ (right) with 40 monomers  | 26 |
| 3.2  | (a) Average energy and (b) heat capacity curves at confined torsion strength   |    |
|      | $S_{\tau} = 4. \ldots $ | 28 |
| 3.3  | (a) Side view and (b) top view of a 3-helix bundle. (c) Side view of a 3-helix   |    |
|      | bundle with two crossing helical ends and (d) side view of the 2-helix bundle  |    |
|      | with two crossing helical ends   | 29 |

| 3.4  | (a) The average square radius of gyration and (b) the fluctuation of the square                |    |
|------|--|----|
|      | radius of gyration at $S_{\theta}=4$   | 30 |
| 3.5  | Examples of (a) a liquid globule and (b) a random-coil structure. $\ldots$ .                   | 31 |
| 3.6  | The hyperphase diagram of the polymer with 40 monomers at confined torsion                     |    |
|      | strength $S_{\tau} = 4$ which is parameterized by the bending strength $S_{\theta}$ and the    |    |
|      | temperature $T$ . Examples of structures dominant in the different phases are                  |    |
|      | also shown.  | 33 |
| 3.7  | The helical folding pathways in the $q_1$ and $q_2$ space at fixed bending strength            |    |
|      | $S_{\theta} = 4$ by decreasing the temperature T. The black dot locates the $(q_1, q_2)$ for   |    |
|      | the 3-parallel-helix bundle  | 35 |
| 3.8  | (a) The average energy and (b) heat capacity curves for $S_{\tau} = 5.$                        | 38 |
| 3.9  | (a) Average square radius of gyration and (b) its fluctuation for $S_{\tau} = 5.$              | 40 |
| 3.10 | Dependence of the peak temperature of the heat capacity $T_{C_v^{max}}$ on the bending         |    |
|      | strength $S_{\theta}$ for the compact phase  | 42 |
| 3.11 | Representative conformations dominating the different structural phases iden-                  |    |
|      | tified for bending strengths in the interval $S_{\theta} \in [0, 60]$ at $S_{\tau} = 5. \dots$ | 42 |
| 3.12 | (a) Average energy and (b) heat capacity curves at $S_{\tau} = 8. \ldots \ldots \ldots$        | 44 |
| 3.13 | (a) Average square radius of gyration and (b) the fluctuation of the square                    |    |
|      | radius of gyration at $S_{\theta} = 8$   | 46 |
| 3.14 | The hyperphase diagram of the polymer with 40 monomers at confined tor-                        |    |
|      | sion strength $S_{\tau} = 8$ , parameterized by the bending strength $S_{\theta}$ and the      |    |
|      | temperature $T$ . Representative conformations are shown as well                               | 48 |
| 3.15 | The helical folding pathways in the $q_1$ and $q_2$ space at fixed bending strength            |    |
|      | $S_{\theta} = 8$ as the temperature T decreases  | 51 |

| 4.1 | $\beta$ -sheet structure obtained with $S_{\theta} = 200$ and $S_{\tau} = 20$ by changing the |    |
|-----|---|----|
|     | reference torsion angle $\tau_0$ and bending angle $\theta_0$ to $\pi$ and $1/3\pi$           | 56 |
| 5.1 | (a) A sample code for with MPI statements. (b) Simulation results for the                     |    |
|     | sample code   | 58 |
| 5.2 | (a) Modified program for case A. (b) Case A results   | 64 |
| 5.3 | (a) Modified program for case B. (b) Case B results   | 66 |
| 5.4 | (a) Modified program for case C. (b) Case C results   | 68 |

### Chapter 1

# Introduction

Amino acids are the building blocks of proteins. There are 20 different types of amino acids in bioproteins and all of them contain an  $NH_2$  group (amino) and a *COOH* group (carboxylic). A peptide is a short chain of fewer than fifty amino acids linked by peptide bonds (Fig. 1.1), in which the  $NH_2$  group of one amino acid binds to the *COOH* group of another amino acid [1].



Figure 1.1: A peptide bond is the bond between an  $NH_2$  group and COOH group.

A polypeptide is a longer, continuous and unbranched chain of fifty or more amino acids and a protein is a polypeptide. For example, Chymotripsin (Fig. 1.2) [2,3] is a single protein with about 100 to 500 amino acids which acts as a biological catalyst [4].



Figure 1.2: The structure of chymotrypsin in which the labelled amino acids form the active site of the enzyme. Chymotrypsin has the tertiary structure.

In contrast, hemoglobin (Fig. 1.3) [5,6] in red blood cells consists of four units and each unit contains a compound known as heme. Each heme contains an iron molecule in the center which can bind an oxygen molecule. Hemoglobin plays the key role in transporting oxygen throughout the body.



Figure 1.3:  $\alpha\beta$  hemoglobin has the quarternary structure composed of four units.  $\alpha$  and  $\beta$  subunits are in red and blue, respectively, and the iron-containing heme groups in green.

Proteins, which widely exist in cells or organisms, perform a dazzling array of functions. A protein chain folds into a distinct three-dimensional structure which is stabilized by noncovalent interactions. The function of a protein is determined by the geometric structure and the distinctive chemical properties and sequences of the amino acids it is composed of. Conceptually, there are four hierarchical levels of protein structures (Fig. 1.4) [7,8].



Figure 1.4: Four levels of protein hierarchy. These are primary structure, secondary structure, teritiary structure and quarternary structure.



Figure 1.5: Atomic model (all-atom model). Hydrogen bonds are represented by yellow dashed lines in this  $\alpha$ -helix structure. Oxygen atoms, nitrogen atoms, hydrogen atoms and carbon atoms are represented in red, blue, white and green respectively. [9]

The primary structure is the linear sequence of amino acids linked together by peptide bonds. Secondary structures are represented by local  $\alpha$ -helices (Fig. 1.5) or  $\beta$ -sheets and stabilized by hydrogen bonds (Fig. 1.5) between CO and NH groups of different amino acids. The tertiary structure holds together the elements of secondary structure ( $\alpha$ -helix and  $\beta$ -sheet), turns and coils. The formation of the tertiary structure is the result of a single polypeptide chain stabilized by the attractive interaction between polar residues and the surrounding polar solvent (in the simplest case, water). The quaternary structure is composed of several polypeptides. The interactions to stabilize tertiary and quaternary structure are weak and the difference in free energy between the folded and unfolded states in typical proteins is in the range of only 20 to 65 kJ/mol comparing to the thermal energy at room temperature. As a consequence, the native conformation of a protein is only marginally stable and is not rigidly fixed but undergoes thermal fluctuations. Structural fluctuations have significant influence on the function of proteins. When the structures are dissolved, it will be dysfunctional. This can eventually result in severe diseases such as Alzheimer's disease, Parkinson's disease and Huntington's disease. Therefore, characterizing the structures of proteins is a fundamental prerequisite for the understanding of how biological systems work and for the development of effective disease protection and therapies.

With this motivation in mind, we performed parallel-tempering replica-exchange Monte Carlo studies of a coarse-grained model for helical polymer with 40 monomers. We introduced four different types of potentials in the model, i.e., interactions between nonbond monomers  $(U_{LJ})$ , interactions between bonded monomers  $(U_{FENE})$ , bending potential  $(U_{bend})$  and torsion potential  $(U_{tor})$ . Previous studies focused on the structural transitions in the spaces of torsion strength and temperature for flexible and semi-flexible polymers. However, the dependence of structural phase on the bending strength has not yet been studied in detail. By employing canonical analysis and adequate order parameters, we aim at constructing the hyperphase diagrams. We find in the spaces of the bending stiffness and temperature at confined torsion strengths, an array of phases such as two-helix bundle, two-crossing-helix bundle, three-helix bundle, amorphous, liquid and random coils. A novel structure we call as helix-liquid was discovered at intermediate values of the bending stiffness.

### Chapter 2

# Methodology

#### 2.1 Model selection

The model selection is the first step to investigate the structural transitions of the polymer system. Atomic models, also known as all-atom models (Fig. 1.5), are particularly popular among biochemists. Biological details are tried to be maintained on a semi-classical level mimicing quantum effects. As a consequence, atomic models are extremely complicated. Some of them include extensive physical interactions such as electrostatic potentials, solvent buffer effects etc. Due to its complexity, simulations of such models require extensive system resources. Typically, atomic models cannot provide generic features or statistical information for different biological systems. However, these approaches are only useful for the characterization of specific system properties. Hundreds to thousands of force-field parameters have to be introduced, but are very difficult to evaluate. Two different atomic models often will not yield the same results.

Complex macromolecules consist of thousands of atoms. However, they are finite systems and the general structural behaviours do not necessarily depend on atomic details. The chemical group(s) forming a monomer can be considered as a coarse-grained entity. Therefore, we employ a generic coarse-grained model (Fig. 2.1) for the polymer [17–20]. This model provides the generic facts which cause the polymer folding and it also helps us obtain the prominent secondary and tertiary structures as well as statistical information. An important fact is that the coarse-grained models are simple and feasible in computer simulations.



Figure 2.1: Coarse-grained representation of an  $\alpha$ -helix. All different amino acids are considered as identical monomers.

Here, we focus on a model of a polymer chain with N=40 monomers. The conformation of our 40mer can be expressed by the vector  $\overrightarrow{X} = (\overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_{40}})$  where  $\overrightarrow{x_i}$  is the position of the ith monomer. The energy of the polymer chain contains non-bonded, bonded, bending and torsion interactions. Therefore, the total energy of the polymer with conformation X can be written as:

$$E(\mathbf{X}) = S_{\rm LJ} \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} U_{\rm LJ}(r_{ij}) + S_{\rm FENE} \sum_{i=1}^{N-1} U_{\rm FENE}(r_{ii+1})$$

$$+S_{\theta} \sum_{l=2}^{N-1} U_{\text{bend}}(\theta_l) + S_{\tau} \sum_{k=3}^{N-1} U_{\text{tor}}(\tau_k)$$

The interactions between all nonbonded monomers are represented by the dimensionless shifted Lennard-Jones (LJ) potential [42, 43] with cutoff (Fig. 2.2 and Fig. 2.3)

$$U_{\rm LJ}(r) = \begin{cases} 4[(\sigma/r)^{12} - (\sigma/r)^6] - V_{\rm LJ}(r_c), & r \le r_c, \\ 0, & \text{otherwise.} \end{cases}$$
(2.1)



Figure 2.2: The nonbonded interaction is represented by the dimensionless shifted Lennard-Jones (LJ) potential. For bonded interaction, the FENE potential is used.

The length scale of the nonbonded interaction is represented by  $r_0$ , which is set to unity in the simulations and, is where the minimum of the LJ potential is located. The LJ potential is zero at the van der Waals radius  $\sigma = r_0/2^{1/6}$ . If the distance r between any pair of monomers is greater than  $r_c = 2.5\sigma$ , the LJ potential will be zero. The  $r_c$  is called the cutoff distance and the LJ potential is shifted by  $V_{LJ}(r_c) = 4[(\sigma/r_c)^{12} - (\sigma/r_c)^6] \approx -0.0163169$  to avoid a discontinuity at the cutoff point. The non-bond interaction strength  $S_{LJ}$  is the reference energy scale for all energetic quantities, and it is set to unity in our simulation.



Figure 2.3: The dimensionless shifted Lennard-Jones (LJ) potential with cutoff.

The interaction between any pair of bonded monomers can be described by the finitely extensible nonlinear elastic (FENE) potential [21,22] (Fig. 2.2 and Fig 2.4)

$$U_{\text{FENE}}(r) = \ln\{1 - [(r - r_0)/R]^2\}.$$
(2.2)

The minimum and the maximum bond lengths are limited by the parameter R since the FENE potential diverges as  $r \to r_0 \pm R$ . The bond interaction strength is  $S_{FENE} = -KR^2/2$  for which we fixed the parameter values  $R = (3/7)r_0$  and  $K = (98/5)r_0^2$  [10].



Figure 2.4: The finitely extensible nonlinear elastic (FENE) potential.



Figure 2.5: The bending angle is denoted by  $\theta$  (blue) and the torsion angle by  $\tau$  (red).

The bending potential [15] is defined as

$$U_{\text{bend}}(\theta) = 1 - \cos(\theta - \theta_0). \tag{2.3}$$

where  $\theta$  is the bending angle formed by two successive bonds (Fig. 2.5). The reference bending angle is set to  $\theta_0 = 1.742$ .

Similarly, the torsion potential [16] is given by

$$U_{\rm tor}(\tau) = 1 - \cos(\tau - \tau_0), \tag{2.4}$$

where  $\tau$  is an out-of-plane torsion angle formed by three successive bonds (Fig. 2.5). The reference torsion angle is  $\tau_0 = 0.873$ . With these reference values of  $\theta_0$  and  $\tau_0$ , helical segments in our polymer model resemble  $\alpha$ -helices with about 4 monomers per turn in the ground state.  $S_{\theta}$  is the bending strength and  $S_{\tau}$  is the torsion strength. We systematically investigate the system's structural behaviours by varying the values of  $S_{\theta}$  from 0 to 200 at confined values of  $S_{\tau}=4$ , 5 and 8, respectively.

#### 2.2 Sampling and simulation

The polymer system is composed of 40 identical monomers and we choose a random initial microstate in our simulations. The standard importance-sampling method typically used in Monte Carlo simulations is the Metropolis method [44]. At the given temperature T, it is based on the canonical microstate probability

$$P(\mathbf{X}) \sim e^{-\beta E(\mathbf{X})}$$

where  $\beta = 1/k_B T$  and  $k_B$  is the Boltzmann constant. For numerical reasons, we set  $k_B = 1$  in the simulations.

In our simulations, we use simple Monte Carlo updates to sample the conformational space. Each update results in a change in the polymer's energy by an amount  $\Delta E$ , where  $\Delta E = E(\mathbf{X'}) - E(\mathbf{X})$ ,  $\mathbf{X'}$  is the updated state and  $\mathbf{X}$  is the previous state. The update is accepted with probability  $P_{\text{acc}}^{\text{M}}$  according to the Metropolis sampling criterion,

$$P_{\rm acc}^{\rm M}(\mathbf{X} \rightarrow \mathbf{X'}) = \min(1, e^{-\beta \Delta E})$$

Obviously, the update will be accepted when  $\Delta E < 0$ . If the updated state has a greater energy, i.e.,  $\Delta E > 0$ , we randomly generate a number  $r \in [0, 1)$  and accept the updated state only if  $r < e^{-\beta \Delta E}$ . The proposed update is rejected otherwise.

There are several different types of updates which can be used in sampling, and their choice impacts the efficiency of the simulations. Here, we employed both the local displacement update and the torsion update (Fig. 2.6). For the displacement update, we randomly choose a monomer (red) and randomly move it within a cubic box with edge lengths  $d = 0.3r_0$ . If the box space is too large, the acceptance rate becomes low and it will

take more time to update the system microstate. If the box size is small, it will lead to a high acceptance rate but the change of the polymers structure will be very small and it still takes longer to make a significant change of the state. Therefore, a reasonable box size can lead to gains in simulation efficiency.

For the torsion update, we randomly select a bond and rotate the whole monomers after the bond by a random angle. It is not difficult to understand that the acceptance rate for such global updates is low. However, it increases the simulation efficiency by a lot once the global update was accepted due to the large structural change. It helps the system overcome the local minimums i.e. the local free energy barriers.



Figure 2.6: Left side shows the local update. The randomly picked monomer is in red and is updated within a cubic box. An example for a torsion update by  $45^{\circ}$  is shown on the right. The randomly picked bond is in blue. The three monomers in light green show the new positions after the  $45^{\circ}$  torsion update.

However, even with global updates, the Metropolis method is still one of the least efficient methods, because the widths of canonical distributions at low temperatures are extremely small and  $\beta \sim (1/T)$  is very large. When the updates lead to an increment in energy, they will be strongly suppressed due to the Boltzman weight  $e^{-\beta\Delta E} \rightarrow 0$ . As a result, the system will be trapped in a local energy minimum for a long time. In addition, at a first order transition, the energy distribution is bimodal, i.e., there are two peaks with a highly suppressed energetic region in between. The system will be easily trapped in either one of peaks if the Metropolis sampling is the standalone method. In order to enhance the sampling in the low temperature region and improve the simulation efficiency, we employ the replica-exchange (parallel tempering) Monte Carlo method. In our parallel simulation, multiple threads are utilized where each thread is used to sample the conformation space at a single fixed temperature. After a sufficient number ( $\sim 2000$ ) of Monte Carlo updates, we attempt to exchange the configuration in each temperature thread with one of its neighbors. The mechanism of the replica exchange is visualized in Fig. 2.7. The red box represents high temperature threads while the blue box represents low temperature threads. Since the lowest temperature threads and the highest temperature threads only have one neighbor, they will be idle for half of the exchange attempts.

- The recipe of replica-exchange parallel tempering simulation
  - 1. 36 temperatures are assigned to 36 threads.
  - 2. Randomly initialize the configuration of the polymer.
  - 3. Using local or global updates to update the polymer structure.
  - 4. Calculate the energy change  $\Delta E = E(\mathbf{X'}) E(\mathbf{X})$
  - 5. Accept or reject according to the Metropolis probability  $P_{\rm acc}^{\rm M}(\mathbf{X} \rightarrow \mathbf{X'}) = \min(1, e^{-\beta \Delta E})$
  - 6. Repeat steps 2 to 4 for sufficient Monte Carlo updates.
  - 7. Every 80000 steps, attempt to exchange configurations between two neighbor cores with the acceptance probability  $P_{\rm acc}^{\rm PT}(i \leftrightarrow j) = \min(1, e^{-(\beta_i - \beta_j)[E(\mathbf{X}_j) - E(\mathbf{X}_i)]})$



### Simulation Time

Figure 2.7: The mechanism for the replica-exchange between neighbor threads.

#### 2.3 Canonical analysis

The canonical ensemble is the statistical ensemble that represents the possible states of a system in thermal equilibrium with a heat bath at a fixed temperature. A system in the canonical ensemble with fixed temperature T, system volume V and the number of particles N has the Helmholtz free energy of F(N, V, T). In the thermodynamic limit, it is possible to employ an order parameter to distinguish the phase transitions in the system.

In the classification scheme, phase transitions are categorized into two classes. The first order transition, also called as discontinuous transition, is defined by a discontinuity in the first derivative of free energy, i.e., entropy S at transition temperature  $T_{tr}$  (Fig. 2.8)

$$\left(\frac{\partial F}{\partial T}\right)_{N,V} = -S(N,V,T) \tag{2.5}$$



Figure 2.8: Schematic plot of the entropy S of a system experiencing a first-order transition.

where  $\Delta S$  measures the distance of the gap. First order phase transitions involve the latent heat  $\Delta Q$ :

$$\Delta Q = T_{tr} \Delta S \tag{2.6}$$

which is the measurement of the energetic separation between the two phases.

The second order transition, also called a continuous transition, in which the entropy S is continuous but the second derivative of the free energy, i.e., the heat capacity  $C_{\rm v}$  is discontinuous at the critical temperature  $T_c$  (Fig. 2.9):

$$\left(\frac{\partial^2 F}{\partial T^2}\right)_{N,V} = -\left(\frac{\partial S}{\partial T}\right)_{N,V} = -\frac{1}{T}C_v(T).$$
(2.7)



Figure 2.9: Entropy S and heat capacity  $C_v$  of a system at a second-order transition.

At the critical temperature  $T_c$ , the heat capacity follows a power law  $C_v(\tau) \propto |\tau|^{-\alpha}$ , where  $\tau = (T - T_c)/T$  and  $\alpha$  is the critical exponent. In the thermodynamic limit, the size of a system is considered as infinitely large. However, the system we investigate is finite and finite size effects are significant and the quantities discussed above do not exhibit non-analyticities for finite systems. In the finite system, instead of singularity, peaks of order parameters may indicate the pseudophase transition. It is difficult to distinguish the order of transition because both first and second order transition exhibit peaks in the fluctuation quantities. However, we can determine the order of transition from the distribution of the internal energy near the transition temperature (Fig. 2.10).



Figure 2.10: The energy histogram of a flexible polymer with 55 monomers near  $T \sim 0.3$ . [a] with bonded LJ potential [b] without bonded LJ potential.

Bimodal distributions indicate a first-order transition. The energy difference between the two peaks is the associated latent heat  $\Delta Q = E_2 - E_1$ . The energy histogram at the critical temperature associated with a latent heat  $\Delta Q = 0$ . Apparently, power law behaviour cannot be expected for finite systems. In some cases, canonical analysis may not be sufficient to categorize the phases for finite systems because peaks do not always exist in heat capacity when the pseudophase transition does occur. Therefore, signals from different quantities are crucial in order to obtain an accurate pseudophase diagram. For this purpose, structural order parameters such as the fluctuation of end-end distance  $d\langle R_{ee}^2 \rangle/dT$  and radius of gyration  $d\langle R_g^2 \rangle/dT$  are employed in our study. In Fig. 2.12, we show an example of a flexible polymer with 13 monomers. Near the temperature T = 1, there is only a very weak signal (shoulder) in the heat capacity  $C_v$  curve and it is difficult to tell whether or not a "phase" transition occurs. However, there is a prominent peak in the radius of gyration curve in the same temperature region, which is a strong evidence for a "phase" transition. By examining the structures of the system near the temperature region T = 1, we find that there is a liquid-gas transition.



Figure 2.11: The heat capacity  $C_{\rm V}$  and the fluctuation of the square radius of gyration  $d\langle R_{\rm g}^2 \rangle/dT$  of a flexible polymer with 13 monomers.

### **2.4** $q_1$ and $q_2$ order parameter space

By varying the bending strength  $S_{\theta}$  at different confined torsion constraints  $S_{\tau} = 4$ ,  $S_{\tau} = 5$ and  $S_{\tau} = 8$ , we discovered a number of interesting structures at low temperatures such as three-helix bundles, two-helix bundles, three-non-helix bundles, two-crossing-helix bundles. However, the traditional order parameters such as heat capacity  $(C_v)$ , end-end distance  $(R_{ee})$ , radius of gyration  $(R_g)$  are not sufficient to distinguish structural phases associated with all these different conformations. In addition, it is very important to find a structural transition path to understand the process of protein folding [31–35]. As a result, we introduce a pair of order parameters denoted as  $q_1$  and  $q_2$  (Fig. 2.12). Here,  $q_1$  is defined as the average total energy of the  $U_{LJ}$  between nonbonded monomers separated from the selected monomer (red) by 6 or less bonds (blue). We do not count the bonded  $U_{LJ}$  between monomers (green) into  $q_1$ . Consequently,  $q_2$  is defined as the average total energy of the  $U_{LJ}$  between the selected monomer (red) and all other monomers (yellow) which are more than 6 bonds away.



Figure 2.12: The pair of order parameters  $q_1$  and  $q_2$ .

The order parameters  $q_1$  and  $q_2$  are given by

$$q_1(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} \Theta_{6,j-i} v_{LJ}(r_{ij}), \qquad (2.8)$$

$$q_2(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} \Theta_{j-i,7} v_{LJ}(r_{ij}), \qquad (2.9)$$

where  $\Theta$  is defined as

$$\Theta_{k,l} = \begin{cases} 1, & \text{if } k \ge l, \\ 0, & \text{otherwise.} \end{cases}$$
(2.10)

[c]



Figure 2.13: The pair values of  $(q_1,q_2)$  for [a] the single helix, [b] two-helix bundle and [c] three-helix bundle shown above are (-5.31,-0.05), (-4.76,-1.61) and (-4.24,-2.5), respectively.

To understand  $q_1$  and  $q_2$ , we first take a single helix as an example. Once a certain monomer **M** is chosen, only the monomers in the above or the below turn will contribute to  $q_1$  by taking into account the fact that there are about 4 monomers in each turn. The monomers which are 6 or more bonds away from **M** contribute 0 to  $q_2$  by considering the cutoff distance of  $U_{LJ}$ . Therefore, in a single long helix,  $q_1$  is minimum and  $q_2$  is maximum (Fig. 2.13). When the number of helical segments increases, two-helix as an example, the monomers which are 6 or more bonds away from **M** may contribute to  $q_2$  if they are in different helical segment. As a consequence, some of the local energy  $U_{LJ}$  (the interaction with the monomers which are 6 or fewer bonds away from **M**) will be sacrificed. Therefore, the  $q_1$  value will increase if the interaction along the chain becomes weaker and  $q_2$  will decrease if there are more energetic contacts between helical segments.



Figure 2.14: The pair values of  $(q_1,q_2)$  for the single helix, two-helix bundle and three-helix bundle shown above are (-5.31,-0.05), (-4.76,-1.61) and (-4.24,-2.5), respectively.

#### 2.5 Binning Jackknife error analysis

Error analysis is a very important step to evaluate the uncertainty associated with a data set. Without the error analysis, it is hard to examine the precision of the data and the results are not trustworthy. Here, we employ binning jackknife error analysis method to obtain reliable error bars for the fluctuation quantities [14].

We have total M correlated data points and each data point is labeled as  $O_m$  where m = 1, 2, ..., M. The whole data set is cut into K bins and each subset contains  $M^B$  data points and we have  $M = KM^B$ . Therefore, a set of uncorrelated data subsets has been created.

The average of the quantity O in each bin is given by

$$\overline{O}_{k}^{B} = \frac{1}{M^{B}} \sum_{m^{B}=1}^{M^{B}} O_{(k-1)M^{B}+m^{B}}, \quad k = 1, ..., K$$
(2.11)

and

$$\overline{O} = \frac{1}{K} \sum_{k=1}^{K} \overline{O}_{k}^{B}$$
(2.12)

The binning jackknife method is the complementary to the binning method. Instead of calculating the average of the observable O in a single bin, we use the data of all bins except the one which is labeled as k. Therefore, we define the kth individual jackknife average by taking all the data points and subtracting those in the kth bin

$$\overline{O}_{k}^{J} = \frac{M\overline{O} - M^{B}\overline{O}_{k}^{B}}{M - M^{B}}$$
(2.13)

The grand average of O is calculated by averaging  $\overline{O}_k^J$  over all jackknife bins

$$\overline{O} = \frac{1}{K} \sum_{k=1}^{K} \overline{O}_{k}^{J}$$
(2.14)

The reliable error for  $\overline{O}$  then is given by

$$\epsilon_{\overline{O}} = \sqrt{\frac{K - 1}{K} \sum_{k=1}^{K} (\overline{O}_k^J - \overline{O})^2}$$
(2.15)

The equation (2.15) is the general function to calculate the error because the observable O can be any fluctuation quantity.

# Chapter 3

### **Results and discussion**

Our study extends previous work [10–12], in which the bending strength was confined at both  $S_{\theta} = 0$  (flexible) [36–38] and  $S_{\theta} = 200$  (semiflexible) [39–41], respectively (Fig. 3.1),. The strength of torsion was altered to explore the phase transition from the torsion-free case  $S_{\tau} = 0$  to a very strong helix potential,  $S_{\tau} = 30$ . Hyperphase diagrams were constructed for both flexible and semi-flexible cases in the space of torsion strength  $S_{\tau}$  and the temperature T. In the semi-flexible phase diagram, an array of robust organization of unique structural phases was found: stable single-helix, 2-helix bundle [23–29], 3-helix bundle and 4-helix bundle conformation dominate in the low temperature region. However, for the flexible case, the folding process [45–51] is not influenced by the torsion strength at all, meaning the dominant phase is an amorphous phase for all values of the torsion strength. Therefore, it is very interesting to explore the  $S_{\theta}$  space between flexible  $S_{\theta} = 0$  and semi-flexible  $S_{\theta} = 200$ cases in more detail which is the focus of our study.


Figure 3.1: Hyperphase diagrams of bending restrained  $S_{\theta} = 200$  (left) and unrestrained  $S_{\theta} = 0$  (right) with 40 monomers.

# 3.1 Stability of polymers with weak torsion restraint

In this section we systematically investigate the structural behaviour of a homopolymer with confined torsion strength  $S_{\tau} = 4$  [52]. A hyperphase diagram is constructed by varying the bending strength  $S_{\theta}$ . We identify an array of structures at low temperature as well as the folding paths in the  $q_1$  and  $q_2$  order parameter space.

## **3.1.1** Energetic order parameters

In Fig.3.2(a), we show the average energy for six different bending strength values i.e.  $S_{\theta} = 1, 3, 5, 10, 30, 60$  and 80. In general, the average energy increases as  $S_{\theta}$  increases for all temperature. The curves are flatter at low and high temperature than at moderate temperatures in the range  $T \sim [0.4, 1.2]$ . The energy difference among different bending strengths is smaller in the moderate temperature region than at low or high temperature. The heat capacity curves are shown in Fig. 3.2(b).



Figure 3.2: (a) Average energy and (b) heat capacity curves at confined torsion strength  $S_{\tau} = 4$ .

At low temperature, we find a noticeable peak for  $S_{\theta} = 1$ . It indicates the structural transition from the 3-helix bundle (Fig. 3.3(a)) to the amorphous phase. The helical segment is not the  $\alpha$ -helix because the number of monomers per turn is not 4 (Fig. 3.3(b)) and, therefore, bending angles do not coincide with the bending reference angle  $\theta_0$ . Since the bending strength is small ( $S_{\theta} = 1$ ), the penalty (increment) from the bending potential is smaller than the decrement from LJ potential. Therefore, it has lower energy and is more preferred by the system at low temperature. There is also a signal for  $S_{\theta} = 30$  which indicates the transition from the 3-helix bundle phase with two crossing helical ends (Fig. 3.2(c)) to 2-helix bundles with two crossing helical ends (Fig. 3.2(d)). In the region  $T \sim [0.2, 0.6]$ , the heat capacity has the most prominent peak for  $S_{\theta} = 1$ . The magnitude of the peaks decrease as the  $S_{\theta}$  values increase. For  $S_{\theta} = 5$  and 10, the signals are weaker and have shifted. The peaks become more significant as the  $S_{\theta}$  values get larger. However, when  $S_{\theta}$  is 30 or larger, there is no difference among the magnitude of the peaks. Near  $T \sim 1.2$ , all heat capacity curves show weak signals. In order to determine whether there are pseudophase transitions, we will need to employ other order parameters, i.e., the structural order parameters.



Figure 3.3: (a) Side view and (b) top view of a 3-helix bundle. (c) Side view of a 3-helix bundle with two crossing helical ends and (d) side view of the 2-helix bundle with two crossing helical ends.

# 3.1.2 Structural order parameters

The average square radius of gyration  $\langle R_g^2 \rangle$  curves for different values of bending strength are shown in Fig. 3.4(a). At lowest temperature, all curves converge around the value of 2. Near  $T \sim 1.2$ ,  $\langle R_g^2 \rangle$  increase sharply. In the high temperature region, the  $\langle R_g^2 \rangle$  values increase as  $S_{\theta}$  values increase. The fluctuation of the square radius of gyration  $d\langle R_g^2 \rangle/dT$ curves are shown in Fig. 3.4(b).



Figure 3.4: (a) The average square radius of gyration and (b) the fluctuation of the square radius of gyration at  $S_{\theta}=4$ .

At low temperature, the blue curve for  $S_{\theta} = 1$  exhibits a clear peak which is followed by a trough as T increases. This means when the temperature increases, the conformation becomes more compact. It further provides evidence for the phase transition from 3-helix bundle to the amorphous phase since the amorphous "phase" is more compact. For the curve of  $S_{\theta} = 30$ , there is a peak which is consistent with the signal in the heat capacity curve that confirms the transition from 3-helix bundle with two crossing helical ends to 2-helix bundle with two crossing helical ends. Near  $T \sim 1.2$ , all curves show pronounced peaks and indicate a strong structural phase transition from liquid globules(Fig. 3.4(a)) to random coils (Fig. 3.4(b)). The height of the peak increases as the value of  $S_{\theta}$  increases. The structural order parameters are complementary to the energetic order parameter  $\langle E \rangle$ where there are shoulders in the heat capacity curves instead of the significant signals in the  $d\langle R_g^2 \rangle/dT$  curves.



Figure 3.5: Examples of (a) a liquid globule and (b) a random-coil structure.

## 3.1.3 The hyperphase diagram

Equipped with the information from both the heat capacity  $C_v$  curves and the fluctuation of the square radius of gyration  $d\langle R_g^2 \rangle/dT$  curves, we construct the hyperphase diagram shown in Fig. 3.5. Colored regions are distinguished in the spaces of the bending strength  $S_{\theta}$  and the temperature T.

First of all, we analyze the phase diagram vertically, i.e., the bending strength  $S_{\theta}$  is considered as a material parameter. In the lowest temperature region, we start with the amorphous phase at  $S_{\theta} = 0$ . As the bending strength  $S_{\theta}$  increases up to 40, the system undergoes transitions from the 3-helix bundle phase to 2-helix bundles with two crossing helical ends phase and 3-helix bundles with two crossing helical ends. The most attractive structural behavior is that as the bending strength  $S_{\theta}$  is larger than 40, the system revisits the 2-helix bundle phase with two crossing helical ends, but if  $S_{\theta} > 65$ , the system falls into 3-helix bundle again. The reason behind this is that the 3-parallel-helix bundle conformation is more preferred by the system for very large bending strength value. In order to make this happen, the helical segment in the middle of the 3-helix bundle with two crossing helical ends must be dissolved and reform again for the large bending strength values  $S_{ heta}$  > 65 by paralleling with one of the two ends. As the bending strength  $S_{\theta}$  gets even larger, the other end gradually lines up with these two helical segments, and the dominant phase of the system is the 3-parallel-helix bundle. The 3-parallel-helix bundle at the top left of the phase diagram (Fig. 3.6) is an example found at  $S_{\theta} = 200$ . In the moderate temperature region  $T \sim [0.5, 1.2]$  we find the liquid phase while in the high temperature region, i.e., T > 1.2, the random coil is the dominant phase. These two phases are not influenced by the bending strength  $S_{\theta}$ .

Now let's cross the phase diagram horizontally, i.e., the temperature T is the control parameter. For  $S_{\theta} > 50$ , the phases can be clearly distinguished as random coil, liquid, 3-helix bundle with two crossing helical ends and 2-helix bundle with two crossing helical ends. For  $S_{\theta} \sim [20, 40]$ , the system visits 4 phases as *T* increases: 3-helix bundles with two crossing helical ends, 2-helix bundles with two crossing helical ends, liquid globules and random coils. Near  $S_{\theta} \sim 10$ , the dominant structures are 2-helix bundles with two crossing helical ends, liquid globules and random coils. For small bending strength values between 1 and 3, there are 4 phases: 3-helix bundles, amorphous, liquid globules and random coils.



Figure 3.6: The hyperphase diagram of the polymer with 40 monomers at confined torsion strength  $S_{\tau} = 4$  which is parameterized by the bending strength  $S_{\theta}$  and the temperature T. Examples of structures dominant in the different phases are also shown.

# **3.1.4** Folding path in $(q_1,q_2)$ order parameter space

We employ the structure parameter space  $q_1$  and  $q_2$  to further study the folding process and distinguish the phases. In our simulation, we stored 600 structures for a fixed bending strength  $S_{\theta}$  at a given temperature. Then we calculate the energy for each of the 600 structures and plot the histogram. We take a subunit of the structures with whose energy is near the peak of the histogram and average the  $q_1$  and  $q_2$  values of the structures in the subunit. This makes a dot for a fixed  $S_{\theta}$  in the  $q_1$  and  $q_2$  order parameter space which represents the global free energy minimum. Therefore, the pathway is the line by connecting all the free energy minima at different temperatures. We show the pathways for 6 different bending strengths  $S_{\theta}$  in Fig. 3.6 and the black dot locates the  $(q_1, q_2)$  of the 3-parallel-helix bundle when  $S_{\theta} = 200$ . All the pathways begin in the high temperature region which is the upper right corner of Fig. 3.7. In the high temperature region, all 6 pathways are in the random coil phase. As the temperature decreases, they undergo the liquid phase which is in the middle of the plot. In the low temperature region i.e. the lower left of the plot, the pathways go into different branches depending on the values of  $S_{\theta}$ . The pathway for  $S_{\theta} = 0$ ends in the amorphous phase while the trajectory of  $S_{\theta} = 3$  ends in the 3-helix bundle phase. However, we notice that all other 4 trajectories for  $S_{\theta} = 10, 30, 60$  and 80 end in the same area. The dominant structures in the low temperature region for the four different  $S_{\theta}$  are "similar". Therefore, it is not a surprise that they have the pair of  $q_1$  and  $q_2$  with close values. The merge also provides us the evidence that the phase for  $S_{\theta} \sim [20, 40]$  (the yellow region in Fig. 3.5) and the phase for  $S_{\theta} > 65$  at low temperature are identical and confirms our most important discovery of the phase revisiting.



Figure 3.7: The helical folding pathways in the  $q_1$  and  $q_2$  space at fixed bending strength  $S_{\theta} = 4$  by decreasing the temperature T. The black dot locates the  $(q_1,q_2)$  for the 3-parallel-helix bundle.

# 3.2 Phase behavior of helical polymers with moderate torsion restraint

In this section we explore the formation of structural phases for helical polymers with confined torsion strength  $S_{\tau} = 5$  by varying the bending strength  $S_{\theta}$  [13]. We used the external energetic and structural fluctuations to indicate the significant phase changes. These analysis supports our expectation for the stabilization of helical phases due to the existence of bending stiffness and we find an array of dominant structures in the low temperature region. Moreover, we find a novel phase at the intermediate values of the bending strength which provides the insight into the process of helix bundle formation.

## **3.2.1** Energetic order parameters

We first discuss the results of energetic canonical statistical analysis applied to our model. In the average energy graph (Fig. 3.8(a)), we notice that the system energy increases when the temperature increases for all values of bending strength. When the temperature become higher ( $T \sim 1.5$ ), the slope of the energy curve decreases which indicates that the energy increase more slowly. At any temperature position, when the polymer is flexible, i.e.,  $S_{\theta} = 0$ the energy of the system is smallest while for bending strength  $S_{\theta} = 60$  the system energy is largest. We also notice that there is a clear gap between  $S_{\theta} = 0$  and  $S_{\theta} = 3$ . This is a clear evidence that tells when the bending constraint starts to apply to the flexible system, the energy of the system dramatically increases. However, when the bending strength is 10 times greater than the torsion constraint, i.e.,  $S_{\theta} > 60$ , increasing bending constraint will not result in a noticeable change in energy. This is because large bending strength governs the system energy and there is no competition with both torsion and non-bonded interaction effects.

The heat capacity  $C_v$  curves as functions of temperature are shown in Fig. 3.8(b). We find

prominent peaks for both small bending strength  $S_{\theta} \leq 5$  and large bending strength  $S_{\theta} \geq 10$ near  $T \sim 0.5$ . The peak suggests the existence of the solid-liquid transition. For the bending strength  $S_{\theta}$  between 5 and 10, we observe broad peaks. A flexible polymer, i.e.,  $S_{\theta} = 0$  has the largest height of peak and decrease with the bending strength increasing. However when the bending strength is  $S_{\theta} = 7$  there is a clear shift and the heat capacity increases with the bending strength. We see a clear amorphous phase develops at low temperature (T < 0.37) when bending strength is  $S_{\theta} < 5$ . For  $S_{\theta} > 10$  we see a distinct 3-helix structure form at low temperature (T < 0.5). It is because with larger bending strength, breaking helix structures will result in larger energetic penalty and therefore the system will more favor helical structures. When the bending strength is between 5 and 10, the structural phase seems to be a mixed phases of both. We could also clearly identify the shoulder for small bending strength  $S_{\theta} \leq 5$  and large bending strength  $S_{\theta} \geq 10$  but it is less obvious for bending strengths between 5 and 10. This shoulder suggests there may occur a "phase" transition and we will employ the structural order parameters to perform an in-depth analysis.



Figure 3.8: (a) The average energy and (b) heat capacity curves for  $S_{\tau} = 5$ .

## **3.2.2** Structural order parameters

Fig 3.9(a) gives us the information about the average square radius of gyration  $\langle R_q^2 \rangle$ . We can see all curves converge at 2 in the low temperature region. There is a very steep slope between temperature 1 and 1.5 for all values of bending strength which suggests there is a structural transition in this temperature region. However, we do not see any signal at low temperature. This is because the transition at low temperature is the solid-liquid transition and the structures in both solid and liquid phase are compact and the change in radius of gyration is not noticeable. The derivative of the average square radius of gyration is shown in Fig. 3.9(b). We see the prominent peaks between temperature 1 and 1.5 for all values of bending strength which is a strong evidence that suggests a structure phase transition in this region and confirms the shoulder in the heat capacity curves. It indicates that the transition causes the major changes in the structures and can clearly be associated with the liquid-gas transition between globular and random coil. The liquid phase of our polymer is overall disordered but may contain short, locally ordered segments such as helical turns. The random-coil conformations do not have any indication of order and the fluctuation behaviour of the individual monomers is only constrained by the bond potential, i.e., the FENE potential.



Figure 3.9: (a) Average square radius of gyration and (b) its fluctuation for  $S_{\tau} = 5$ .

# **3.2.3** The dependence of the peak temperature of $T_{C_v^{max}}$

It is worth to take a look at the details for polymers with intermediate bending strength  $S_{\theta}$ to better understand how the different solid phases are formed. A deeper investigation into the shift region of the liquid-solid transition temperature gives us a first hint. By using the peak temperature of the heat capacity curves in the low temperature region, we obtain the dependence of the peak temperature of the heat capacity  $T_{C_{ii}^{max}}$  for all bending strengths  $S_{\theta}$  which is shown in Fig. 3.10. Starting from the flexible polymer case, i.e.,  $S_{\theta} = 0$ , we notice that the transition point shifts to lower temperatures as well as the peak height decreases as the bending strength  $S_{\theta}$  increases. The signal has shifted to  $T \approx 0.31$  when  $S_{\theta} = 5$ . Following that, there is a clear visible gap that another signal appears at about  $T \approx 0.45$  when the bending strength  $S_{\theta} = 7$ . Then the signal becomes stronger for larger  $S_{\theta}$ and shifts to the higher temperatures. This behaviour gives rise to the assumption that an intermediate phase exists for  $5 < S_{\theta} < 7$ , which mediates the crossover between amorphous and the helical structures. Indeed, by visual inspection of the dominant structures, we find a novel structure type which has not been discovered before in the limiting cases of flexible and semiflexible polymers. It has a compact and hybrid helix-liquid structure which possesses an ordered helical segment attached to a disordered liquid body (Fig. 3.11). This intermediate structural phase is rather weak but it does show the evidence that how the helical bundles form out of the liquid phase, provided the bending stiffness is strong enough to counter the non-bonded interactions which cause the formation of most compact amorphous structures.



Figure 3.10: Dependence of the peak temperature of the heat capacity  $T_{C_v^{max}}$  on the bending strength  $S_{\theta}$  for the compact phase.



Figure 3.11: Representative conformations dominating the different structural phases identified for bending strengths in the interval  $S_{\theta} \in [0, 60]$  at  $S_{\tau} = 5$ .

# 3.3 The general structural transitions of helical macromolecules with strong torsion restraint

In this section, we study the impact of the competition between bending strength and torsion effects upon the formation of the structural phases for the helcial polymer with confined torsion strength  $S_{\tau} = 8$  [53]. We employ the conventional canonical statistical analysis by introducing both the energetic and structural thermodynamic quantities. These analyses help us construct the hyperphase diagram in spaces of temperature and bending stiffness. We take a detailed look in the low temperature region and distinguish an array of dominant structures. Further, we explore the polymer folding path by using the order parameters  $q_1$ and  $q_2$ .

## **3.3.1** Energetic order parameters

We first show the result of the average energy in Fig. 3.12(a). Generally speaking, the average energy  $\langle E \rangle$  increases as the bending strength increases across the temperature T space. Also, the average energy  $\langle E \rangle$  increases as the temperature increases for all bending strength values. In the temperature  $T \sim [0.4, 1.3]$ , the slope for all curves is steeper than the slope in low temperature region i.e. T < 0.4 and in the high temperature region i.e. T > 1.4. The average energy change between  $S_{\tau} = 0$  and  $S_{\tau} = 3$  is much more pronounced than the change between  $S_{\tau} = 30$  and  $S_{\tau} = 100$ , because there is still significant competition between bending and torsion potential. However, when the bending strength is large enough, it dominates the system energy. Any further increment in  $S_{\tau}$  does not have a noticeable impact on the system behaviour.



Figure 3.12: (a) Average energy and (b) heat capacity curves at  $S_{\tau} = 8$ .

The heat capacity curves are shown in Fig. 3.12(b). In the low temperature region, i.e.,  $T \sim 0.15$ , we notice the weak signals at the bending strengths  $S_{\theta} = 1$  and  $S_{\theta} = 3$ . There is a transition from the 3-helix bundle to the amorphous phase. Near  $T \sim 0.25$ , we notice another weak signal at  $S_{\theta} = 30$ . It indicates the phase transition from the 2-helix bundle with two crossing helical ends to the 2-helix bundle. Near  $T \sim 0.4$ , we clearly identify a prominent peak at small bending strengths, i.e.,  $S_{\theta} \leq 3$ . The height of the peak increases when the polymer changes from flexible, i.e.,  $S_{\theta} = 0$  to  $S_{\theta} = 1$  and then decreases as the  $S_{\theta}$  is getting larger. At  $S_{\theta} = 8$  the peak becomes flat and wide. For bending strength  $S_{\theta} = 16$ , there is a clear shift for the critical temperature and then the height of the peak increases as the bending strength increases further. We characterize it as the amorphous-liquid transition. In the temperature region  $T \sim [0.8, 1.2]$ , we notice that there are prominent peaks for large bending strengths  $S_{\theta} \geq 8$  while only shoulders are visible for small bending strengths  $S_{\theta} \leq 3$ . In order to confirm the "phase" transition we perform a structural analysis.



Figure 3.13: (a) Average square radius of gyration and (b) the fluctuation of the square radius of gyration at  $S_{\theta}=8$ .

## **3.3.2** Structural order parameters

The average square radius of gyration  $\langle R_g^2 \rangle$  curves are shown in Fig. 3.13(a). We see all curves converge to 2 at low temperature except for the curve of  $S_{\theta} = 100$  which converges to 3. The reason is that the dominant structure at the lowest temperature for  $S_{\theta} = 100$  is a 2-helix bundle which is less compact than the structures found at smaller bending strength values. A noticeable slope occurs near  $T \sim 0.25$  at the bending value  $S_{\theta} = 30$  which bridges the curves at small values of bending strengths and the curve at  $S_{\theta} = 100$ . A very steep slope occurs for all bending strength values between temperature 0.8 and 1.2 which indicates there is a structural transition in this temperature region.

We show the result of the fluctuation of the square radius of gyration  $d\langle R_g^2 \rangle/dT$  in Fig. 3.13(b). At the low temperature, i.e.,  $T \sim 0.1$  there are small peaks for the curves of bending strengths  $S_{\theta} = 1$  and  $S_{\theta} = 3$  and as the temperature increases slightly there is a noticeable down trend. This is a strong evidence which indicates the "phase" transition from 3-helix bundle to the amorphous phase, where conformations are more compact than 3-helix bundle which explains the down trend of the curve when the temperature increases. In the temperature region  $T \sim [0.25, 0.5]$ , we notice a very clear peak in the curve for  $S_{\theta} = 30$  which indicates the transition from the 2-helix bundle with two crossing helical ends to the 2-helix bundle phase and suggests that this is a structural transition. However, for the curves of  $S_{\theta} = 8$  and  $S_{\theta} = 16$ , the signals are not that strong and may suggest a mixed phase here. In the temperature region  $T \sim [0.8, 1.2]$ , we see prominent peaks for all values of bending strength which are strong evidences for the phase transition from liquid to random-coil phase.



Figure 3.14: The hyperphase diagram of the polymer with 40 monomers at confined torsion strength  $S_{\tau} = 8$ , parameterized by the bending strength  $S_{\theta}$  and the temperature T. Representative conformations are shown as well.

# 3.3.3 The hyperphase diagram

After the careful analysis of both energetic and structural order quantities, we construct the hyperphase diagram in the space of bending strength  $S_{\theta}$  and temperature T (Fig. 3.14). We first look into the phase diagram vertically, i.e., the bending strength  $S_{\theta}$  is varied as a material parameter. In the lowest temperature region, the system stays in the amorphous phase for the flexible polymers,  $S_{\theta} = 0$ . As the bending strength  $S_{\theta}$  increases slightly, the system enters the 3-helix bundle phase. The 2-helix bundle with two crossing helical ends becomes the dominant configuration when  $S_{\theta} \sim [15, 50]$ . As the bending strength  $S_{\theta}$  is getting even larger, i.e.,  $S_{\theta} > 65$ , there is a clear dominant conformation which is identified as 2-helix bundle. Near  $T \sim 0.7$ , the dominant phase is qualitatively recognizable as the liquid phase. However, as the bending strength  $S_{\theta}$  increases, the liquid phase shrinks. This is because the helical segments are more stable when the bending strength is large and it will require more energy, for example higher temperatures to dissolve the ordered conformation. If T > 1.2, the random-coil phase is the dominant phase for all bending values.

We then analyze the phase diagram horizontally, i.e., the temperature is controlled as an external parameter. For the flexible polymer,  $S_{\theta} = 0$ , the dominant conformations are amorphous, liquid and random-coils phase as temperature increases. For  $1 < S_{\theta} < 5$ , there are 4 phases, i.e., 3-helix bundle, amorphous, liquid and random-coil phase. As  $S_{\theta}$ increases up to 8, the amorphous phase disappears with the other 3 phases remained. For the  $S_{\theta} \sim [15, 55]$ , the system is in the 2-helix bundle with two crossing helical ends phase at the low temperature and enters 2-helix bundle, liquid and random-coil phases, respectively, by increasing *T*. For large bending strengths  $S_{\theta} > 65$ , we discover 3 dominant phases: 2-helix bundle, liquid and random-coil phase. Indeed, these phases across the temperature space imply the specific folding path for a given  $S_{\theta}$  which can be further verified in the  $q_1$  and  $q_2$ order parameters space.

# **3.3.4** The exploration of folding paths in $(q_1, q_2)$ order parameters space

The  $q_1$  and  $q_2$  plot is shown in Fig. 3.15 which is used to investigate the polymer folding processes and further verify the dominant phases. Each dot in the plot represents the global free energy minimum. We have discussed the method in detail in section 3.1.4 for how to determine the dots. The pathway is obtained by connecting all free energy minima at different temperatures for a given bending strength  $S_{\theta}$ . Here, we show 6 pathways for 6 different bending strength  $S_{\theta}$ . The upper right corner is the high temperature region and all 6 pathways start from there, i.e., in the random-coil phase. As temperature decreases, they enter liquid phase in the center of the plot, and we do not see the paths diverge. However, in the lower left of the plot i.e. in the low temperature region, we find the paths separate away. For  $S_{\theta} = 0$ , the trajectory ends in the amorphous phase. For  $S_{\theta} = 3$ , the path goes through the amorphous phase but near the end, it is stabilized in the 3-helix bundle phase. For  $S_{\theta} = 8$ , the path ends between the 3-helix bundle phase and 2-helix bundle with two crossing helical ends phases which is not a surprise. Because it is near the phase transition band. For the  $S_{\theta} = 16$ , it clearly ends in the 2-helix bundle with two crossing helical ends phase. The pathway for bending strength  $S_{\theta}$  is very interesting as it exhibits a clear jump down from 2-helix bundle phase into the phase of the 2-helix bundle with two crossing helical ends phase. For  $S_{\theta} = 100$ , the trajectory ends in the 2-helix phase. The results from  $q_1$  and  $q_2$  plot are perfectly consistent with our hyperphase diagram.



Figure 3.15: The helical folding pathways in the  $q_1$  and  $q_2$  space at fixed bending strength  $S_{\theta} = 8$  as the temperature T decreases.

# 3.4 Comparison

In this section, we will systematically compare the generic structural behaviour of helical polymers between torsion strength  $S_{\tau} = 4$  and  $S_{\tau} = 8$ . System behaviours at  $S_{\tau} = 5$  is very similar to those at  $S_{\tau} = 4$  and thus, we exclude this case from the discussion here.

From our study, we find the following in the phase behavior at  $S_{\tau} = 4$  and  $S_{\tau} = 8$ :

- There are robust helical structures in the low temperature region. Some of the dominant structures are in common such as the amorphous phase, 3-helix bundle when the S<sub>θ</sub> is small and 2-helix bundle with two crossing helical ends.
- The liquid phase dominates the moderate temperature region for all bending strength values.
- In the high temperature region, they all have random-coil phase independently of the bending strength  $S_{\theta}$ .

There are also obvious differences which can be summarized as below:

In the low temperature region with very large bending strength S<sub>θ</sub> > 70, the dominant conformation at S<sub>τ</sub> = 4 is the 3-helix bundle while it is a 2-helix bundle for S<sub>τ</sub> = 8. The reason is that the formation of the helical segment depends on the torsion strength. The increment in the number of helical segments will increase the penalty of torsion potential but may reduce the LJ potential. Therefore, the outcome is the result of the competition of torsion energy and interaction between non-bonded monomers. When S<sub>τ</sub> = 4, the system sacrifices some torsion potential and benefits from LJ potential due to the relatively small value of S<sub>τ</sub>. However, if S<sub>τ</sub> = 8, i.e., the torsion strength is relatively large, the gain from LJ potential cannot offset the loss in torsion potential by having more helical segments. Thus, the dominant structure in S<sub>τ</sub> = 8 case is 2-helix

bundle instead of 3-helix bundle. Similarly, if the torsion strength is even larger, such as  $S_{\tau} = 30$ , the dominant structure is a single helix.

 The number of dominant phases in the low temperature region with S<sub>τ</sub> = 4 is larger than the number of phases at S<sub>τ</sub> = 8. It is not surprising because smaller torsion strength means the system may have more flexibility to form preferred structure with lower free energy.

# Chapter 4

# **Conclusion and outlook**

In this thesis, we focused on exploring the structural behaviour of helical polymers. For this purpose, we performed parallel-tempering Monte Carlo simulations of a coarse-grained model of a homopolymer with 40 monomers. In our model, the interaction between the neighboring monomers is represented by a FENE potential and the Lennard-Jones potential is used to mimic the interaction between non-bonded monomers. We further included the torsion potential with the dihedral angle and the bending potential with the bending angle to support the helicity. In our study, the torsion strength values are confined at  $S_{\tau} = 4, 5$ and 8 respectively. By varying the bending strength  $S_{\theta}$ , we systematically investigated the structural transitions and stability. In addition, we gained the insight into the folding process in the order parameter  $q_1$  and  $q_2$  space.

We investigated the helical polymers with confined torsion strength  $S_{\tau} = 4$ . By using both energetic and structural canonical analysis, we constructed the hyperphase diagram in the space of  $S_{\theta}$  and the temperature T. We find various helical structures including 3-helix bundles with two crossing helical ends, 2-helix bundles with two crossing helical ends and 3-parallel-helix bundles. The helical segment are formed when both torsion and bending constraints are included. The system is in the amorphous phase at low temperature if either of them is zero.

The macromolecules with confined torsion strength  $S_{\tau} = 5$  has also been studied. We explore the significant difference in the shapes of the solid structures. By using both canonical statistical analysis of energetic and structural fluctuation quantities, we discover the novel structural phase we call helix-liquid phase, which is found in a certain intermediate interval of bending stiffness values i.e.  $S_{\theta} \in [5,7]$  where the conformation are composed of a helical segment attached to a disordered liquid body. This feature may lead to a better understanding of the general consequences for folding process of polymer under the influence of competing interactions and constraints.

Lastly, we explored the homopolymer with confined torsion strength  $S_{\tau} = 8$ . We constructed the hyperphase diagram and found amorphous, 3-helix bundle, 2-helix bundle with two crossing helical ends and 2-helix bundle at low temperature. As expected, in the moderate or high temperature region, the dominant phases are liquid and random-coil, respectively. The helical segments are stabilized by both torsion and bending strength. Our result may lead an insight into most protein structures.

Further, we discovered that the reason of the formation of helical order is because of the torsion and bending potential. In other words, the correct torsion and bending restraints mimic the hydrogen bond which supports the helicity in our model. The number of the helical segments decreases as the torsion strength increases in order to reduce the penalty from the torsion potential. Our coarse-grained homopolymer model is used for tertiary structures of helical bundles. However, for secondary structure, there are two types, i.e., helices and sheets. Only backbone atoms in amino acids are necessary for the formation of secondary structure, not the atoms in the side chain i.e. the residue. Because the amino acids' backbones are identical for all 20 amino acids, our homopolymer model works perfectly for the individual helix segment. The far LJ potential glue the segments together in order to form the bundles. In this context, we believe our model will also work for the other type of

protein secondary structure i.e.  $\beta$ -sheet by modifying certain parameters or reference values. In Fig. 4.1, we show the potential sheet-like structure with different input reference values. The future studies are desirable to construct a whole picture of the structural behaviour for the  $\beta$ -sheet polymers.



Figure 4.1:  $\beta$ -sheet structure obtained with  $S_{\theta} = 200$  and  $S_{\tau} = 20$  by changing the reference torsion angle  $\tau_0$  and bending angle  $\theta_0$  to  $\pi$  and  $1/3\pi$ 

# Chapter 5

# Appendix

To discuss MPI statements, let us start with a simple C program in Fig. 5.1 (a). This code aims in exchanging an array with two elements in two threads. The original arrays and the arrays after exchanging are shown in Fig. 5.1(b). Every MPI program must contain the



Figure 5.1: (a) A sample code for with MPI statements. (b) Simulation results for the sample code.

preprocessor directive

## #include "mpi.h"

This file, "mpi.h", contains the definitions and declarations for compiling any MPI program. MPI use the consistent scheme for identifiers. All MPI identifiers start with string "MPL" and if it is followed by the name of MPI function, the first character of the function name is capitalized and subsequent characters are lowercase (e.g. "MPI\_Send"). If "MPL\_" followed by a constant, all characters should be uppercase (e.g. "MPI\_INT").

### MPI\_Init(&argc, &argv);

The function "MPI\_Init()" must be called first before any other MPI functions and "MPI\_Init()" can only be called once. Its parameters are pointers to the main function's parameters(argc and argv). It allows the systems to do any necessary setup so that MPI library can be used.

#### MPI\_Finalize()

Once a program has finished using MPI library, it must call "MPI\_Finalize()". The statement "MPI\_Finalize()" cleans up any unfinished "MPI\_" function and frees memory allocated by MPI. So far, the previous discussions are about the most basic MPI frame. Next, we will discuss some common used MPI functions.

### MPI\_COMM\_rank(MPI\_COMM\_WORLD, &my\_rank)

The function "MPI\_COMM\_rank()" returns the rank of a process to its second parameter and the rank starts from zero. The first parameter is a communicator which is a collection of processors that can communicate with each other. "MPI\_COMM\_WORLD" is the only communicator we need and it has been predefined in the library which consists of all the running processors when the program begins.

#### MPI\_COMM\_size(MPI\_COMM\_WORLD, &my\_size)

The function "MPI\_COMM\_size()" returns the total number of processors in the communicator to its second parameter. The first parameter represents a communicator as well. The processors in a communicator use functions "MPLSend()" and "MPLRecv()" to send, receive or exchange information with each other. The exact syntax for the two functions is given below:

| MPI_Send( |              |              |
|-----------|--------------|--------------|
|           | Void*        | message,     |
|           | int          | count,       |
|           | MPI_Datatype | datatype,    |
|           | int          | destination, |
|           | int          | tag,         |
|           | MPI_Comm     | communitor ) |
| MPI_Recv( |              |              |
|           | Void*        | message,     |
|           | int          | count,       |
|           | MPI_Datatype | datatype,    |
|           | int          | source,      |
|           | int          | tag,         |
|           | MPI_Comm     | communitor,  |
|           | MPI_Status*  | status)      |

For the function "MPL\_Send()", the first parameter "message" refers the block of memory which stored the content that is going to be sent. The second parameter "count" tells the system how long is the message and the third parameter determines the data type of the message e.g. "MPL\_INT", "MPL\_FLOAT" and "MPL\_DOUBLE" etc. The forth parameter "source" refers the rank of the processor which is going to receive the message. The fifth parameter "tag" is the label of the message which makes the message be unique. The last parameter in "MPL\_Send()" represents the collection of the processors.

For the function "MPLRecv()", the forth parameter "source" refers the rank of the

processor which is going to send the message. We should notice that the fifth parameter "tag" here has to match up with the "tag" in the function "MPI\_Send()" so that messages won't get mixed up. The last parameter "status" in "MPI\_Recv()" returns the information about the data which is actually received. For other parameters in "MPI\_Recv", they have the same definitions as those in the function "MPI\_Send()".
# 5.1 Noticeable Issues in MPI Programming

In practical, we need the processors communicate with each other and exchange information instead of running independently. However, when the processors are attempting to exchange messages, some issues might arise due to synchronization. Let's first assume that there are only two processors A and B running on distinct nodes. One mechanism is that processor A sends "request to send" to processor B and waits until it receives "ready to receive" from B, at which point processor A begins to send the actual message to processor B. This approach is called synchronous communication. It requires both processors to be at the same stage. Alternatively, the message sent by processor A can be copied to a system-controlled block of memory. Instead of waiting processor B's response, processor A can continue executing. When the processor B arrives at the point that it is ready to receive message, the system simply copies the buffered message into processor B's memory location. The second approach is called asynchronous or buffered communication. The advantage of second approach is that the sending processor is able to continue executing even if the receiving processor is not ready. The disadvantages are also clear. This approach will occupy more system resources. Further, if the receiving processor is already ready, it will still copy the message to the block of memory which takes longer running time.

Another issue we need to address is the semantics of the "MPI\_Send()"/"MPI\_Recv()" pairing which has occurred in our simulation. Suppose processor A calls "MPI\_Send" but processor B does not execute a receive. Although it won't cause the program to crash, the message will simply sit in the buffer until the program ends. When the capability of buffer is not enough then the program may crash. If the system does not provide buffering, the situation becomes even worse since processor A will be waiting forever for "ready to receive" from processor B and the program is actually dead.

Let's look back to the program in the Fig. 2.1 (a). The purpose of this program is to

exchange the array a between processor 0 and processor 1. For processor 0, the value of array a is  $\{0,0\}$  while for process0r 1, its array a is  $\{1, 1\}$ . The correct result is shown in Fig. 2.1 (b) after the exchange between two processors. By changing the sequence of "MPI\_Send()" and "MPI\_Recv()" in the program, we will see how it affects the results.

• Case A

---- processor 1 -----

MPI\_Recv(a,2,MPI\_INT,0,33,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE); MPI\_Send(a,2,MPI\_INT,0,11,MPI\_COMM\_WORLD);

— processor 0 —

MPI\_Recv(a,2,MPI\_INT,1,11,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE);

MPI\_Send(a,2,MPI\_INT,1,33,MPI\_COMM\_WORLD);

We exchange the sequence of "MPLSend()" and "MPLRecv()" for both processors in the original program. After the modification, the program won't exchange the array a between processor 0 and processor 1 at all. We could not even get a result shown in Fig. 2.2 (b). The reason is that both processor 0 and processor 1 call "MPLRecv()" function first without executing "MPLSend()". Therefore, both processors won't receive anything from the buffer and will keep waiting forever. Since the "MPLRecv()" function has not been cleared, both processors are not able to move forward to execute "MPLSend()" which causes the program dead.



Figure 5.2: (a) Modified program for case A. (b) Case A results.

• Case B

— processor 1 —

## MPI\_Recv(a,2,MPI\_INT,0,33,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE);

## MPI\_Send(a,2,MPI\_INT,0,11,MPI\_COMM\_WORLD);

— processor 0 —

#### MPI\_Send(a,2,MPI\_INT,1,33,MPI\_COMM\_WORLD);

#### MPI\_Recv(a,2,MPI\_INT,1,11,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE);

We exchange the sequence of "MPLSend()" and "MPLRecv()" for processor 1 while the processor 0 remains the same. We find that pseudo exchange happens! Based on Fig. 2.3 (b), We get a wrong result. Instead of exchanging array a between processor 0 and processor 1, processor 0 sends its values to processor 1 but array a in processor 0 stays the same. The reason causes the pseudo exchange is that processor 1 receives the array a from processor 0 first and replaces its own values. Then, processor 1 sends array a to processor 0. Obviously, it will be identical with the array a in processor 0.



Figure 5.3: (a) Modified program for case B. (b) Case B results.

• Case C

— processor 1 —

### MPI\_Send(a,2,MPI\_INT,0,11,MPI\_COMM\_WORLD);

MPI\_Recv(a,2,MPI\_INT,0,33,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE);

— processor 0 —

MPI\_Recv(a,2,MPI\_INT,1,11,MPI\_COMM\_WORLD,MPI\_STATUS\_IGNORE); MPI\_Send(a,2,MPI\_INT,1,33,MPI\_COMM\_WORLD);

We exchange the sequence in processor 0 and processor 1 stays the same. We find the pseudo exchange again! The result is shown in Fig. 2.3 (b) which is not what we expected. Processor 1 sends its values to processor 0 but array a stays the same in processor 1. The reason is the same as the previous case which is that the processor 0 receives the value from processor 1 before sending out its own. The array a in processor 0 will be replaced by processor 1's and then if processor 0 executes "MPLSend()", it actually sends processor 1's array a back.



Figure 5.4: (a) Modified program for case C. (b) Case C results.

These outstanding issues we discussed above are very common in parallel computing and they are hard to debug most time. Especially for case B and C, the program will run to the end and you will obtain the results without even realizing the results are wrong. Therefore, crossing check is very important. Also, to avoid these bugs, executes simple parallel programs first and make sure you could get the expected results before running an extensive simulation.

# Bibliography

- [1] Hamley, Ian W. Introduction to Peptide Science. Wiley, 2020.
- [2] S. Freer, J. Kraut, J. Robertus, Wright HT and Xuong NH, *Biochemistry*. 9 9 (1997)
- [3] Jcwhizz, https://commons.wikimedia.org/wiki/File:Chymotrypsin.png
- [4] Anthonsen 2000 Applied Biocatalysis (Taylor and Francis. ISBN 9789058230249)
- [5] Patricia-R, https://commons.wikimedia.org/wiki/File:1GZX\_Haemoglobin.png
- [6] H. Lehmann and R. W. Carrell, Br Med J. 4 748 (1968)
- [7] Cmdr Jameson https://commons.wikimedia.org/wiki/File:Protein\_structure.png
- [8] L Pauling, Proc Natl Acad Sci USA. **37** 205 (1951)
- [9] InfoCan https://commons.wikimedia.org/wiki/File:Alpha\_helix.png
- [10] M. J. Williams and M. Bachmann, Phys. Rev. Lett. **115** 048301 (2015)
- [11] M. J. Williams and M. Bachmann, Phys. Rev. E **93** 062501 (2016)
- [12] M. J. Williams and M. Bachmann, Polymers 8 245 (2016)
- [13] Y. Dai and M. Bachmann, IOP Conf. Series: Journal of Physics: Conf. Series 1012 012007 (2018)

- [14] M. Bachmann 2014 Thermodynamics and Statistical Mechanics of Macromolecular Systems (Cambridge: Cambridge University Press)
- [15] T. W. Nee, J. Chem. Phys. **59** 2244 (1973)
- [16] D. C. Rapaport, Phys. Rev. E 66 011906 (2002)
- [17] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. 57 2607 (1986)
- [18] C. J. Greyer, Computing Science and Statistics: Proceedings of the 23rd symposium on the Interface edited by Keramidas E M (1991)
- [19] K. Hukusima and K. Nemoto, J. Phys. Soc. Jpn. 65 1604 (1996)
- [20] K. Hukusima and K. Nemoto, Int. J. Mod. Phys. C 7 337 (1996)
- [21] K. Kremer and G. S. Grest, J. Chem. Phys. **92** 5057 (1990)
- [22] A. Milchev, A. Bhattacharya and K. Binder, Macromolecules **34** 1881 (2001)
- [23] T. Bereau, M. Bachmann and M. Deserno, J. Am. Chem. Soc. **132** 13129 (2010)
- [24] C. Zhang, J. Hou and S. H. Kim, Proc. Natl. Acad. Sci. USA 99 3581 (2002)
- [25] S. W. Bruun, V. I. and F. M. Poulsen, Proc. Natl. Acad. Sci. USA 107 13306 (2010)
- [26] Z. Guo and E. M. Boczko, Proc. Natl. Acad. Sci. USA **94** 10161 (1997)
- [27] N. L. Harris and F. E. Cohen, J. Nol. Bio. **236** 1356 (1994)
- [28] S. Wallin, Proc. Natl. Acad. Sci. USA **97** 13614 (2000)
- [29] S. R. Presnell, Proc. Natl. Acd. Sci. USA 86 6592 (1989)
- [30] T. Bleha, Com. The. Chem. **1006** 62 (2013)

- [31] G. A. Carri and M. Muthukumar, Phys. Rev. Lett. 82 5405 (1999)
- [32] Z. Qin and M. J. Buehler, Eur. Phys. J. E. **36** 53 (2013)
- [33] P. G. Wolynes, J. Phys. Chem. **99** 2177 (1995)
- [34] T. Bereau and M. Bachmann, Biophys. J. 100 2764 (2011)
- [35] V. Varshney and G. A. Carri, Phys. Rev. Lett. 95 169304 (2005)
- [36] S. Schobl and W. Janke, Phys. Rev. E 84 051805 (2011)
- [37] S. Schnabel, M. Bachmann and W. Janke, J. Chem. Phys. 131 124904 (2009)
- [38] M. P. Taylor and K. Binder, J. Chem. Phys. **131** 114907 (2009)
- [39] J. A. Martemyanova and K. Binder, J. Chem. Phys. **122** 174907 (2005)
- [40] P. A. Wiggins and P. C. Nelson, Phys. Rev. E **73** 031906 (2006)
- [41] D.T. Seaton, M. Bachmann and D. P. Landau, Int. J. Mod. Phys. 23 1240004 (2012)
- [42] J. E. Lennard-Jones, Proceedings of the Royal Society of London 106 738 (1924)
- [43] J. E. Lennard-Jones, Proceedings of the Physical Society. 43 5 (1931)
- [44] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys. 21 1087 (1953)
- [45] C. Levinthal, J. Chim. Phys. 65 44 (1968)
- [46] K. A. Dill, Biochemistry **24** 1501 (1985)
- [47] J. D. Bryngelson and P. G. Wolynes, Proteins **21** 167 (1995)
- [48] M. Karplus, Fold. Des. **2** S69 (1997)

- [49] M. Bachmann and W. Janke, Phys. Rev. E **73** 020901 (2006)
- [50] G. C. Rollins and K. A. Dill, J. Am. Chem. Soc. **136** 11420 (2014)
- [51] T. Vogel and M. Bachmann, J. Chem. Phys. 142 104901 (2015)
- [52] Y. Dai and M. Bachmann, "Stability of helical polymers with weak torsion restraint" (Preprint, 2021)
- [53] Y. Dai and M. Bachmann, "The general structural transitions of helical macromolecules with strong torsion restraint" (Preprint, 2021)