Theoretical studies of the formation mechanism of protein complex by using coarse-grained models

メタデータ	
	言語: eng
	出版者:
	公開日: 2017-10-05
	キーワード (Ja):
	キーワード (En):
	作成者:
	メールアドレス:
	所属:
URL	http://hdl.handle.net/2297/40539

Theoretical Studies of the Formation Mechanism of Protein Complex by Using Coarse-grained Models

Micke Rusmerryani

Division of Mathematical and Physical Science,

Graduate School of Natural Science and Technology,

Kanazawa University

2014

博士論文

粗視化モデルを用いたタンパク質複合体形
成機構に関する理論的研究

金沢大学大学院自然科学研究科

数物科学専攻

計算バイオ研究室

学籍番号　　　　1123102010

氏名　　　　　　Micke Rusmerryani

主任指導教員氏名　長尾秀実

Acknowledgement

First of all, I would like to express my deepest gratitude to the one above all of us, the omnipresent Allah SWT, for answering my prayers and for giving me the strength to accomplish this doctoral thesis.

In this occasion, I would like to offer my sincerest gratitude to my supervisor, Prof. Hidemi Nagao, who not only served as my supervisor but also encouraged and challenged me throughout my academic program. The good advice, support and valuable discussions on both an academic and a personal level from Prof. Masako Takasu, for which I am extremely grateful. I also would like to gratitude Assistant Professor Hiroaki Saito, Assistant Professor Kazutomo Kawaguchi, and Dr. Acep Purqon for the kindness, supports, and helps me a lot not only in academic but also in my life. Also to Prof. Mineo Saito, Prof. Tatsuki Oda, and Prof. Shinichi Miura, for the valuable advices and supports.

Many thanks are given to all my members of the Computational Biophysics Laboratory for kindly supports and encouragements to make this work possible. To Seiji Matsui and Hikmah Balbeid who always support me through good time and bad time. To all members of PPI Ishikawa and Berang for many helps and enjoyable moments. It is a pleasure to thank those who made this thesis possible, hence I apologize I can not write all the names but I will always remember all your kindness.

Last but not the least, I dedicate this thesis to my parents who have given me the opportunity of an education from the best institutions and support throughout my life and to my family and Dendi who gave me their unequivocal support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

My current research is inspired by the rapid developments of nanoscience in the last decade. The concept of nanoscience was first addressed by physicist Richard Feynman in his lecture "There's Plenty of Room at the Bottom" [1]. In his talk, he considered a feasibility to manipulate matter on an atomic scale and offered some challenges which are known later as nanotechnology.

In this universe, all matter is built up of extremelly small particles called atoms. Since nanoscience deals with the nanomaterials, it requires the ability to imagine, observe, and work on the nanoscale, where the prefix "nano" refers to 10^{-9}.

Nevertheless, both experimental and theoretical studies on behavior of nanomaterials are still limited. For instance, to see such an extremely small matter like atom, we need the most advanced microscope, the scanning tunneling microscope (STM). Recently, computer simulation has emerged as the midway between the theoretical and experimental approaches to obtain better understanding of matter on the atomic scale.

In particular, my research interest lies in the field of biomolecular modeling and more specifically in the study of the formation of protein complex. Just like building is made up of bricks, cell is made up of atoms and living

organism is made up of cells. To understand the properties of the cell, it is important to understand how the atoms or molecules attract and bind to another to form a cell.

One of the most common computational methods for studying the dynamics of protein in atomic scale is molecular dynamics simulation. Although molecular dynamics simulation allow us to observe the protein dynamics in atomic details, it is limited to the size and time scales. In order to surpass these limitations, in this thesis I try to extend the biomolecular modeling to study the protein complex dynamics.

1.2   Overview of the thesis

This thesis is organized as follows:

Chapter 1 gives general introduction consisting of motivation of my research and overview of this thesis.

In chapter 2, we introduce the basic information of azurin including the structure and the biological functions. We also introduce the basic concept of coarse-graining in biomolecular modeling.

The next three chapters are presented based on my published papers as first author. In chapter 3, we present our study on unfolding process of azurin using native-center structure based model. The paper is:

M. Rusmerryani, M. T. Pakpahan, M. Nishimura, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Transition state analysis of azurin via Gō-like model*, AIP Conf. Proc., **1518**, 641 (2013).

In chapter 4, we expand this model to simulate several chains of azurin. The paper is:

M. Rusmerryani, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Coarse-grained simulation of azurin crystal complex system: Protein–protein interactions*, ISCS 2013 Selected Papers, **4** (2013).

In chapter 5, we improve our model by using Lennard-Jones potential as the intermolecular interaction in order to find more transferable coarse-grained model. The paper is:

M. Rusmerryani, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Protein–protein interactions of azurin complex by coarse-grained simulations with a Gō-like model*, JPS Conf. Proc., **1**, 012054 (2014).

Chapter 6 is conclusion of this thesis and future work.

Moreover, in Appendix B we present our extended work which will be submitted. In this work, we employ knowledge-based approach by empirically evaluate the intermolecular contacts from known crystal structure of azurin. This work will offer a new insight to approach the intermolecular potential model for unknown complex structure. Last, in Appendix A we provide brief derivation of our force field.

Chapter 2

Research Objectives

In the following sections we will briefly introduce the main objectives for our research: protein azurin and coarse-grained models in the field of bio-molecular modeling.

2.1   Azurin

Azurin is one of cupredoxin or blue copper protein that contains a single Type I copper center. Azurin molecule has low molecular weight around 14 kDa and consists of 128 amino acids [2]. Azurin is found mainly in *Pseudomonas aeruginosa* bacteria [3] which usually grows in the soil but also often found in the lungs. Azurin from *Pseudomonas aeruginosa* is known to exhibit a large stability [4].

As other cupredoxins, azurin functions in the electron transfer. Particularly, the electron transfers occur between azurin and cytochrome c-551 [5] and between azurin and cytochrome c oxidase [6]. Advanced studies have been conducted both experimentally and theoretically for further investigation of the kinetics of electron transfer.

Recently, many researches put their attention to investigate azurin since azurin may be considered as a proper candidate for treatment of cancer through nanotechnology [7]. Azurin was found to form a complex with the tumor-suppressor protein p53 [8] and to induce apoptosis in macrophage

cells [9]. Several studies on cancer treatment by azurin are performed, including melanoma [9], breast cancer [10, 11], bone cancer [12], prostate cancer [13], brain tumor [13], and leukemia cells treatment [14].

In this thesis, we use the native structure of azurin complex system obtained from X-ray crystal structure of *Pseudomonas Aeruginosa* azurin (PDB entry : 4AZU) [15]. In this azurin complex, the unit cell[1] consists of one asymmetric unit[2] where its asymmetric unit is composed of a tetramer of azurin molecules (Figure 2.1).

Each azurin is composed of eight β-strands and one helix arranged in a double wound Greek key topology [16]. A Greek key is a series of four sequential β-strands arranged in the order three up-and-down β-strands connected by hairpins are followed by a longer connection to the fourth strand which lies adjacent to the first (Figure 2.2). The structure of azurin containing Greek key motif can be seen in Figure 2.3. The copper is coordinated by three strong ligands arranged in a trigonal-planar configuration (the side chains of Cys112, His117, and His46) and a weak ligand Met121. However, in this thesis the interaction made by copper will be neglected.

---

[1]**Unit cell** is the smallest building block of a crystal structure
[2]**Asymmetric unit** is the smallest part of the crystal that is duplicated and moved by symmetry operations to form the unit cell of the crystal.
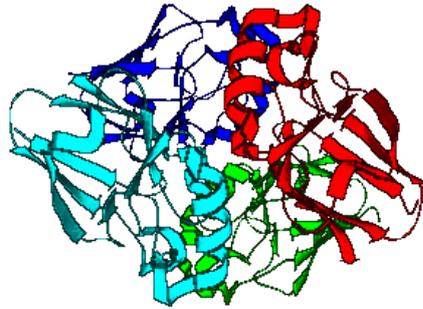
Figure 2.1: Three dimensional structure of azurin complex obtained from X-ray crystal structure of *Pseudomonas Aeruginosa* azurin (PDB entry : 4AZU)
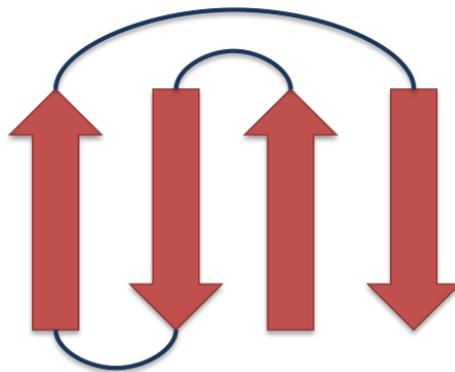
Figure 2.2: Topology of the Greek key motif.
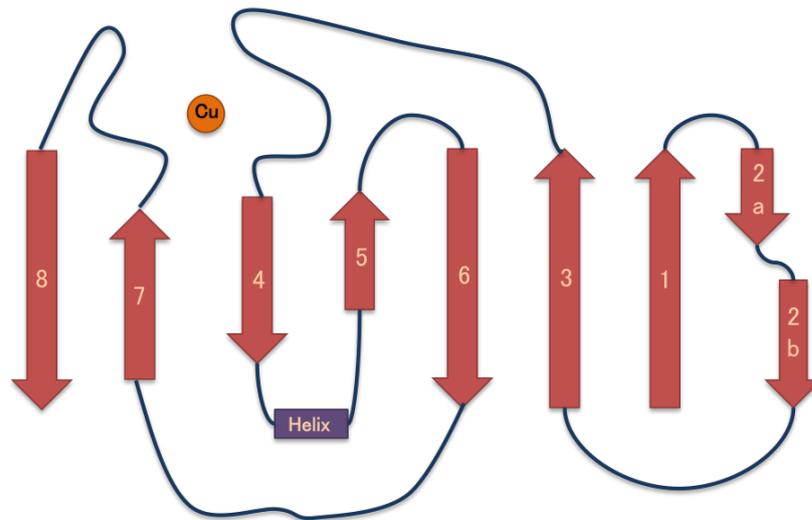
(a) Topology diagram
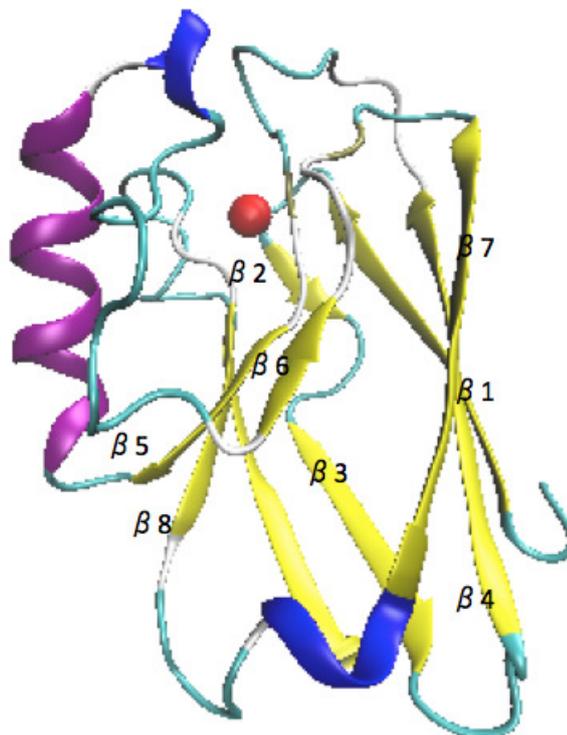
(b) Three dimensional structure

Figure 2.3: Structure of *Pseudomonas aeruginosa* azurin containing Greek key motif showing in simplified topology diagram (a) and in real three dimensional structure generated by VMD (b).

2.2 Coarse-grained models

Coarse-grained model is a lower resolution model where some of fine details are eliminated. In molecular dynamics, this model is obtained by replacing the "unnecessary" atomistic details of a biological molecule. In the past decade, coarse-grained models have gained much attention since they could overcome the spatial and temporal problems of all-atom model. All-atom simulations are limited to small systems and nanosecond time scales. Meanwhile, coarse-grained models allow us to simulate larger systems and slow processes which require micro- to millisecond time scales.

Several coarse-grained models have been developed for many classes of biomolecules: water, lipids [17–20], proteins [21–26], nucleic acids, and carbohydrates. These models are constructed with different levels of resolution and approaches. As the current work deals with protein, now we only discuss coarse-grained protein models and introduce several models that have been quite successful to characterize protein folding and dynamics.

To construct a coarse-grained model, first we have to determine the resolution of the representation for our system. In coarse-grained protein models, each amino acid can be represented by one site, usually associated with the position of α-carbon, or a few sites, usually three or more backbone sites. After that, we have to determine the appropriate interactions between coarse-grained particles. This part has become a great challenge in biomolecular modeling.

There are several approaches to develop the coarse-grained potentials. The most common way is to classify these approaches into **top-down** and **bottom-up** approaches. In bottom-up approaches, the interaction between coarse-grained particles is determined based upon the given fundamental description from a higher resolution model or classical atomistic model for the same system. Conversely, top-down models is constructed on the basis of the real experimental observation that provides the phenomena of physical principles, especially the thermodynamics properties.

These approaches have their own advantages and disadvantages. Bottom-up models can be used to predict the particular system when no such observation exist. On the other hand, while their models are restricted to their dependence on the more detailed model, the top-down models are typically under-constrained, which means that the restrictions are very small. Nevertheless, as the advancement of coarse-grained models may integrate both principles, the distinction becomes quite intuitive and blurred. For instance, the popular Martini model has retained a great success in providing transferable potential for modeling liquid and membrane by incorporating the top-down and bottom-up approach [27, 28].

Another common way is to distinguish them into **physics-based** and **knowledge-based** approaches. While the physics-based approaches employ physical theories to determine the interaction, the knowledge-based approaches employ the empirical informations provided from the experimentally determined three-dimensional structure. This distinction is also becoming blurred with the same reason. Most coarse-grained models for protein usually combine these two approaches with bottom-up approach.

Several models have been successfully applied to study the dynamics of protein. The simplest coarse-grained protein model is network model, such as Elastic Network Model (ENM) [29] or Gaussian Network Model (GNM) [30]. These models determine their coefficients based on the native contact map and employ a spring potential for modeling all interactions. Nevertheless, these models do not provide the directions of particle motions.

Native-centric models also have been greatly used to study the protein folding. Similar to the network models, native-centric model also determine the parameters on the basis of the native contacts. This model, usually referred as Gō model, represents the interaction as bonded and non-bonded interactions [21]. At this time, this model may be the most realistic coarse-grained model for protein. However, these models can not be applied to the unknown structure.

Other promising knowledge-based models have been developed to pro-

vide more transferable coarse-grained model. The best known model is Miyazawa and Jernigan model with the statistical contact potentials. This model has widely been used as a first estimate of the interaction between particles in coarse-grained model [23,31,32]. On the implementation, people usually combine those models with bottom-up approach to construct the potential strength. The commonly used strategies are Iterative Boltzmann Inversion (IBI) [33], Inverse Monte Carlo (IMC) [34], and the Force-Matching [35, 36].

After all, that is why the coarse-grained is favored for solving many biomolecular problems. The flexibility to improve their models by integrating several approaches based on their research purposes has become a great advantage of coarse-grained. This way researchers can optimize the effectiveness of their models.

Chapter 3

Implementation of Coarse-Grained Model to Probe Unfolding Process of Azurin

In this chapter, we employ our structure-based coarse-grained simulation by adopting Gō potential model to examine the effects of mutated azurin to the unfolding process.

3.1   Introduction

Protein assembles to the unique three-dimensional structure called the native state to perform its biological function. The understanding of conformational transition from denatured to native state, or usually known as folding process, is very important. While the native state is unique, the transition state is not just a single conformation. Multiple folding pathways can lead the protein sequences toward the native state or in contrast the pathways may be trapped in the non-native conformation.

The study of protein folding was pioneered by Anfinsen [37] on his observation on the refolding of ribonuclease molecule. His famous **"Thermodynamic Hypothesis"** has become a fundamental keystone to the development on the study of protein folding. Its statement that the native structure

of an amino acid sequence in its normal surrounding is the one which has the lowest free energy, also has been supported by the funneled energy landscape theory [38].

In recent times, major developments on the study of protein folding dynamics has been greatly advanced into the fast and time-resolved techniques. Along with the prior advances in the experimental and theoretical studies, those studies can be combined to develop the computational studies of folding mechanism at the residue or atomic level. For instance, structure-based simulation, pioneered by Gō [21], has successfully employed underlying two essential theories of folding mechanism: the principle of minimal frustration [39] and the funneled energy landscape [38, 40]. In addition, transition state theory also has been widely used to probe the folding/unfolding mechanism with the computational studies [41, 42]. Together, the structure-based simulation and transition state theory have become powerful tools to examine the folding/unfolding process in the multiscale level.

In the last decade, understanding the effects of mutation on protein is one of great issues both in experimental and computational studies. It will provide many valuable insights to understand stability and kinetics of protein such as azurin. In this chapter, we will discuss our coarse-grained simulation on azurin. Azurin is known as an extremely stable protein as (see Section 2.1 for further explanation). However, experimental study of mutated azurin was found that the mutation of His117 to Gly on the apo-form affects the stability of azurin whereas the unfolding proceeds much faster [43]. Currently, we implement the off-lattice Gō-like potential [21, 22] to probe the unfolding dynamics of a mutated azurin and a wild-type azurin.

Experimentally, azurin is known to exhibit two-state folding/unfolding process: native and denatured states [44]. Here, we observed the change in activation free energy relative to the change in stability of the transition state to locate three state ensemble and compared with the wild-type azurin. Moreover, we also considered how the temperature affects the unfolding process of this mutated azurin. For structural description, we probed the unfolding

pathways of azurin using protein engineering technique, Φ-value. Our result has found to be in agreement with both experimental and theoretical data. Present study also shows that the helix region, known as p28 peptide fragment of azurin, remains stable in both mutated and wild-type azurin.

3.2 Material and methods

Protein folding/unfolding is a process of unstructured (unfolded) amino acid sequences transforming into structured state or usually called native state, and vice versa. To understand the folding mechanism, sometimes we need to observe the unfolding mechanism beforehand. In this study, we probe the unfolding mechanism of a mutated azurin and compare it to the wild-type azurin via coarse-grained simulation. More detailed explanation of our model system and simulation method is given below.

3.2.1 Protein

As described in Chapter I, we limit our objective by observing dynamics of azurin. In this chapter, we choose the mutated azurin obtained by changing His117 for a glycine. His117 is one of the three main ligands on the copper binding site. This mutation increases the flexibility on the loop containing those ligands and is less rigid compared to the wild-type azurin. Regarding to the folding process, the folding speed of the mutated azurin is known to be quite similar to the wild-type azurin. In contrast, the unfolding speed is found to be faster than the wild-type [43].

Currently, we simulate single apo-azurin for both mutated and wild-type azurin. The initial structures are taken from protein data bank[1] with PDB ID: 3N2J for H117G azurin [45] and 4AZU for wild-type azurin [15]. The crystal structures of both azurins are almost the same since the position of residue 117 is on the loop which is out of the main β-strand of the azurin.

---

[1]http://www.rscb.org/

3.2.2 Coarse-grained model of protein

The basic concept of coarse-graining is to simplify high-resolution details that are not necessary to understand the particular process. Coarse-grained models of biomolecules usually represent groupings of two or more atoms into a single bead. In our study, we develop coarse-grained model of protein at the residue level in which each residue is represented only with C_α atoms. We set each particle with the same mass.

3.2.3 Unit of coarse-grained model

We determine the units for our coarse-grained model using basic quantities, which are length (σ_0), mass (m), time (τ), and derived quantity, energy (ε_0). The values of our coarse-grained units are listed in Table 3.1. The values of σ_0, m, and ε_0 are determined from the radius of protein, the average mass of amino acids, and the temperature of system, respectively. Meanwhile, the time unit (τ) is calculated by $\sigma_0\sqrt{m/\varepsilon_0}$.

3.2.4 Potential model

We applied the off-lattice model founded by Gō [21] to mimic the perfect funnel aspect of folding energy landscape for our coarse-grained model. We adapt Gō model interaction energy which is developed by Clementi et

Table 3.1: Units of coarse-grained model

coarse-grained units		Experimental units
length	σ_0	1.0 Å
mass	m	137 amu
energy	ε_0	0.6 kcal/mol
time	τ	2.0 ps

al. [22]. This model explicitly maintains the stability of native contacts by eliminating the energetic frustration from the non-native interactions. Until now, this model has retained great success on the folding studies.

The potential energy between particles involves bonded and nonbonded interaction energy as shown in detail in Table 3.2. Bonded potential energy between particles describes spring potential for two successive particles, angle potential energy describes bending motion between two successive virtual-bonds, and dihedral potential describes the rotation of the four subsequent residues. Meanwhile, nonbonded interaction is distinguished into two categories, native interaction and non-native interaction.

In Table 3.2, r, θ, and ϕ represent the distance between two successive residues, the angles formed by three successive particles, and the dihedral angle defined by four subsequent residues along the chain at the given configuration, respectively. The non-bonded interaction implement 10-12 Lennard-Jones potential for native interactions and a short-range repulsive between non-native pairs, where r_{ij} represents the distance between i-th and j-th unsubsequent residues. We define a pair to be in native contact if r_{0ij} is less than 6.5 Å. Otherwise, it will be categorized as non-native pairs. All variables with subscript "0" mean the values of the corresponding variables at the native conformation. Detailed potential parameters are shown in Table 3.3.

Table 3.2: Potential model

Type of interaction		Potential energy
Bonded	Virtual bond-stretching	$K_{\text{bond}}(r - r_0)^2$
	Virtual bond-angle bending	$K_\theta(\cos(\theta) - \cos(\theta_0))^2$
	Virtual bond-torsional term	$K_\phi[1 - \cos(\phi - \phi_0)] +$
		$\frac{K_\phi}{2}[1 - \cos(3 \times (\phi - \phi_0))]$
Non-bonded	Native	$\varepsilon_{\text{nat}}\left[5\left(\frac{r_0^{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_0^{ij}}{r_{ij}}\right)^{10}\right]$
	Non-native	$\varepsilon_{\text{non-nat}}\left(\frac{C}{r_{ij}}\right)^{12}$

Table 3.3: Potential parameter

Parameter		Value in kcal/mol
Virtual bond-stretching	K_{bond}	100.0 Å^{-1}
Virtual bond-angle bending	K_θ	20.0
Virtual bond-torsional term	K_ϕ	1.0
Native	ε_{nat}	0.3
Non-native	$\varepsilon_{\text{non-nat}}$	0.2

3.2.5 Equation of motion

The main idea of our simulation is to predict the dynamics of the protein. For every time step, the position and velocity of each residue are calculated using Newton's second law:

$$m\frac{d^2\vec{r}}{dt^2} = \vec{F}, \tag{3.1}$$

where $\vec{r}$ is the vector of Cartesian coordinate of the particle, and $\vec{F}$ is the gradient of the potential energy at the given particle. In our coarse-grained simulation, we implement the Langevin equation of motion to mimic the non-conservative forces from the solvent which is describe in the following equation:

$$m\frac{d^2\vec{r}}{dt^2} = \vec{F} - \zeta\frac{d\vec{r}}{dt} + \xi(t), \tag{3.2}$$

where ζ, the damping friction coeffient, is set to be $0.25(\tau^{-1})$. Meanwhile $\xi(t)$ represents the random force which satisfy:

$$\langle\xi(t)\rangle = 0; \tag{3.3}$$

$$\langle\xi(t)\xi(t')\rangle = 2m\zeta k_B T\delta(t-t'), \tag{3.4}$$

where k_B is the Boltzmann constant.

We use the simple and widely used numerical integration algorithm, *leap-frog* algorithm, to solve the equation 3.2. Then the position and velocity of each particle can be obtained. This algorithm is computationally less expensive and less "storage consuming" than the predictor-corrector algorithm, yet is still accurate.

3.3 Analysis methods

The kinetic free energy relation can be used to obtain the position of the funnel transition state [38, 41, 42]. While the native state is unique, the

transition state is not just a single conformation which can be defined as an ensemble. The transition state ensemble (TSE) consists of relatively large number of configurations described by specific order parameter that measures its nativeness. Most of small proteins have a two-state folding/unfolding process. In such a case, three states appear and are defined as: native, transition, and denatured states.

3.3.1 Measure of nativeness

One way to measure the nativeness of the given configuration is by the fraction of the native contacts [38]. A pair residue is counted to be in native contact if the distance is less than 6.5 Å in the native state. Related to the kinetic free energy, this order parameter also can be defined as the reaction coordinate. For our convenience, we define it as Q, which mathematically can be written as

$$Q = \frac{\text{number of native contacts in a given configuration}}{\text{number of native contacts in native state}}.$$ (3.5)

This value of Q ranges from 0 to 1, in which Q close to unity represents the similarity to the native structure. In reverse, Q close to zero shows the dissimilarity to the native structure.

By the histogram method, the free energy profile ($F(Q)$) can be obtained [46]. The relation between free energy and the position of the funnel transition state allows us to locate the three states ensemble, which are denatured, transition, and native state.

3.3.2 Φ-analysis

The free energy profile is a good tool to provide us the general description of the funnel of transition state. Nevertheless, it does not provide us structural description. Therefore, further analysis is needed to characterize the TSE. In experimental studies, currently the only way to probe the transition state of the folding process in depth is the protein engineering method,

Φ-analysis. It is defined as the ratio of change of the folding barrier energy to stability upon the mutations, which is represented by following equation

$$\Phi = \frac{\Delta\Delta G^{\ddagger}}{\Delta\Delta G^{0}}, \tag{3.6}$$

where $\Delta\Delta G^{0}$ is the difference in the total free energy between mutant and wild-type proteins, and $\Delta\Delta G^{\ddagger}$ is the free energy changes of the folding barrier.

In the same objective, the theoretical Φ-analysis technique is introduced by Fersht and colleagues [47, 48] to characterize the TSE. This technique has been successfully applied to analyze folding TSE [22,49]. The change in free energy barrier can be interpreted by a single simple reaction coordinate. Then, the Φ-value is defined by:

$$\Phi_i = \frac{\langle E_i \rangle_{TS} - \langle E_i \rangle_D}{\langle E_i \rangle_N - \langle E_i \rangle_D}, \tag{3.7}$$

where E_i is the sum of interaction energies of i-th residue with any other residues and the bracket $\langle\;\rangle$ means average of the quantity over an ensemble. The subscripts represent its states: TS, D, and N, for transition, denatured, and native state, respectively.

This statistical mechanical description of Φ has been widely used for comparison with the experimental Φ-value. Meanwhile the free energy profile allows us to locate the TSE, Φ-value describes the contribution of each residue at the transition state. Besides, it also can be used to measure the changes in TSE upon single or multiple mutations on the folding rate and stability.

3.4   Results

Several short simulations are performed under various temperatures, chosen by bisection method over range of temperatures, to estimate the folding transition temperature (T_f). We start with the low temperature which

gives us high population of native state (high Q) and also the high temperature which gives the opposite condition. Then we select the subinterval to narrow the range of temperature. Repeatedly we apply the bisection method over the new subinterval until the criteria for folding transition temperature is satisfied. The folding transition temperature itself is determined when the native state and denatured state are equipopulated. The folding temperature of wild type azurin is considered to be referenced to a set of states. Then several simulations of mutated azurin are performed under constant temperatures: $T = T_f$, $T < T_f$, and $T > T_f$, for longer simulation time.

In order to obtain the free energy profile, we observed the thermodynamic configurations as a function of the reaction coordinate along simulation time which is represented by the fraction of native contacts formed in a given conformation as we mentioned in the previous section. At the T_f, the Q-score fluctuates along simulation and almost equipopulated between native and denatured states. In the native structure, 186 contacts exist.

As we mentioned before, under thermodynamic conditions most of folding process is known as a two state reaction. In such a case, the free energy profile has double minimum corresponding to the ensembles of native state and denatured state with varying degrees of ordering. In our case, all simulations indicate the two state reactions as shown in Figure 3.1(a). This result is in agreement with the experimental measurements where both of the wild-type and H117G azurins unfold in two-state without intermediates [43].

Furthermore, the three ensemble of states based on the ranges of Q-scores can be identified by this profile. The denatured state is determined by the well curve which is close to zero, in this case we have $Q_D \approx 0.18$. Conversely, the native state is determined by the well curve near the position where the folded state appears around $Q_N \approx 0.72$, and the transition state which is defined by the position of the free energy barrier in the $Q_{TS} \approx 0.3$.
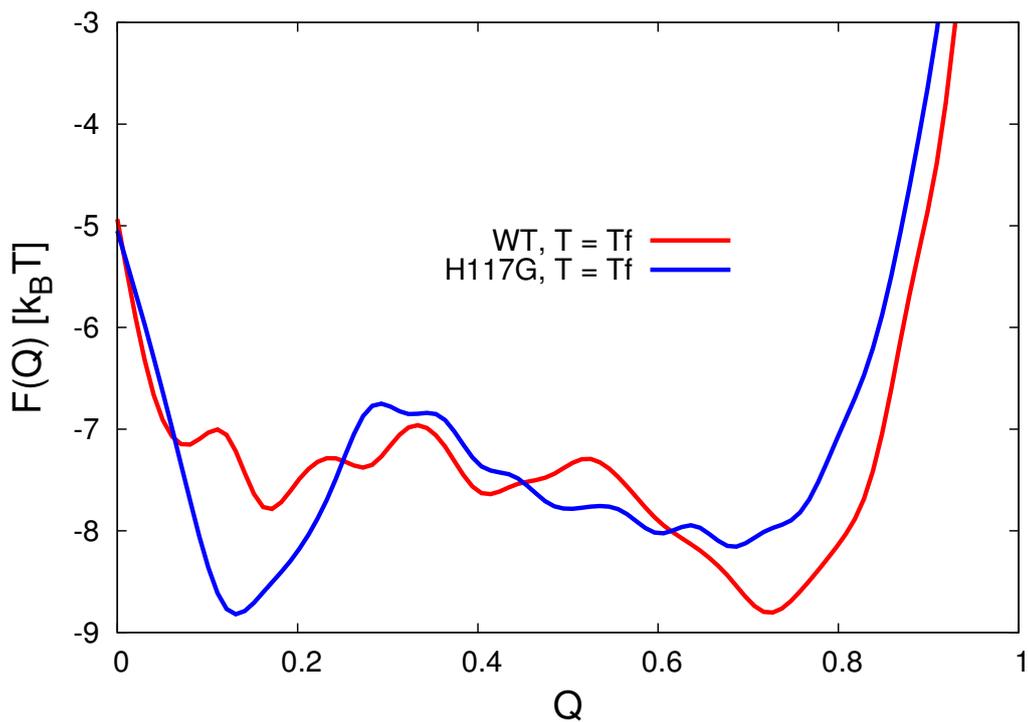
(a) At folding temperature.
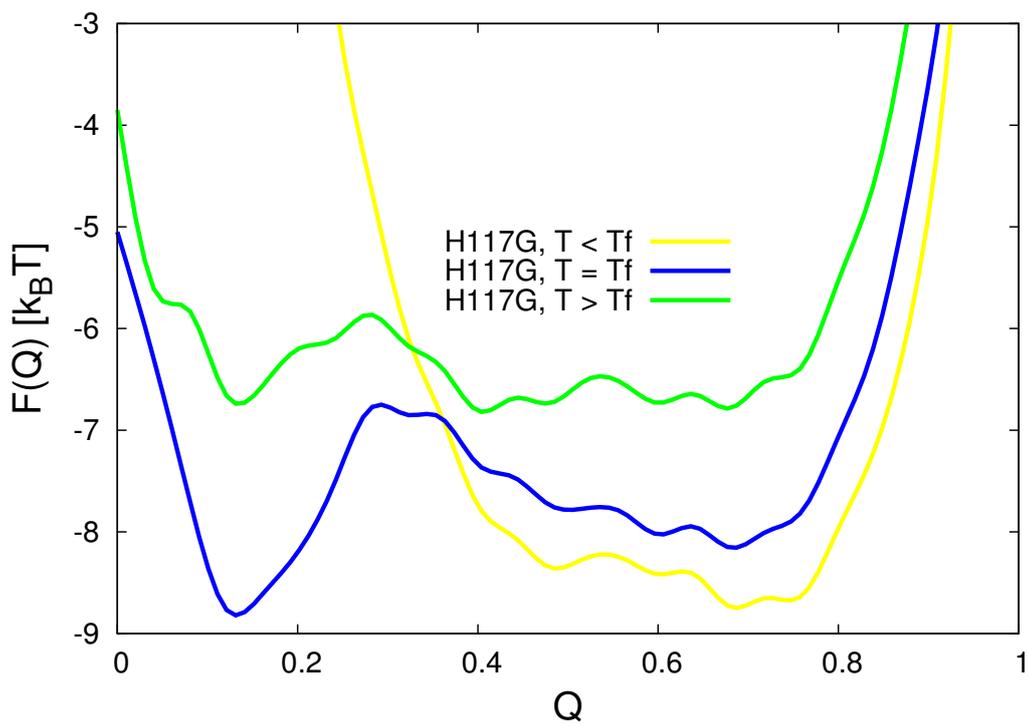
(b) At various temperature.

Figure 3.1: Free energy as a function of reaction coordinate Q. Comparison between mutated azurin and wild-type (a) shows a significant difference on its double well minimum. Meanwhile, (b) shows the dependence on the temperature where the simulation were done at $T = 0.98\,T_f$, $T = T_f$, and $T = 1.01\,T_f$, respectively.

Figure 3.1(a) also shows that the mutation of H117G gives changes in the stability of unfolding azurin. Since Q_D is getting closer to zero, the mutated azurin gains more stability in the denatured state and less similarity with the native state. The lower $F(Q)$ at denatured state of unfolding H117G azurin shows that the unfolding of mutated azurin is faster than the wild-type azurin. In addition, at the folding temperature the mutated azurin gives sharper transition state than the wild-type. These findings obtained from free energy profile are in agreement with the experimental results [43].

As well as the mutation, temperature also affects the unfolding of azurin as shown in Figure 3.1(b). In the lower temperature ($T < T_f$) the native state is found to be more stable. In contrast, higher temperature ($T > T_f$) gives us broader distribution of free energy and smaller free energy barrier. The double minimum also is not clear and its indicates that the azurin may be trapped in non-native conformation.

To observe the structural description of the unfolding process of azurin, the unfolding pathways were quantified using Φ-analysis. Φ-value of each residue was calculated by using equation (3.7). The result in Figure 3.2 shows that the helix region, which contains mainly local interactions, is the most native-like compared to other regions in both wild-type and mutated azurin, as expected. The helix is found to fold faster than strands because its structure contains mainly local interactions [50]. Figure 3.2 also gives us more detailed information related to our findings from the free energy profile. In agreement with the free energy profile, the mutated azurin unfolds faster, specifically at $\beta 3$ and $\beta 5$, yet remains more native-like at $\beta 7$ compared to the wild-type. Here $\beta 3$, $\beta 5$, and $\beta 7$ are the positions of β-strands of azurin (see Figure 2.3).
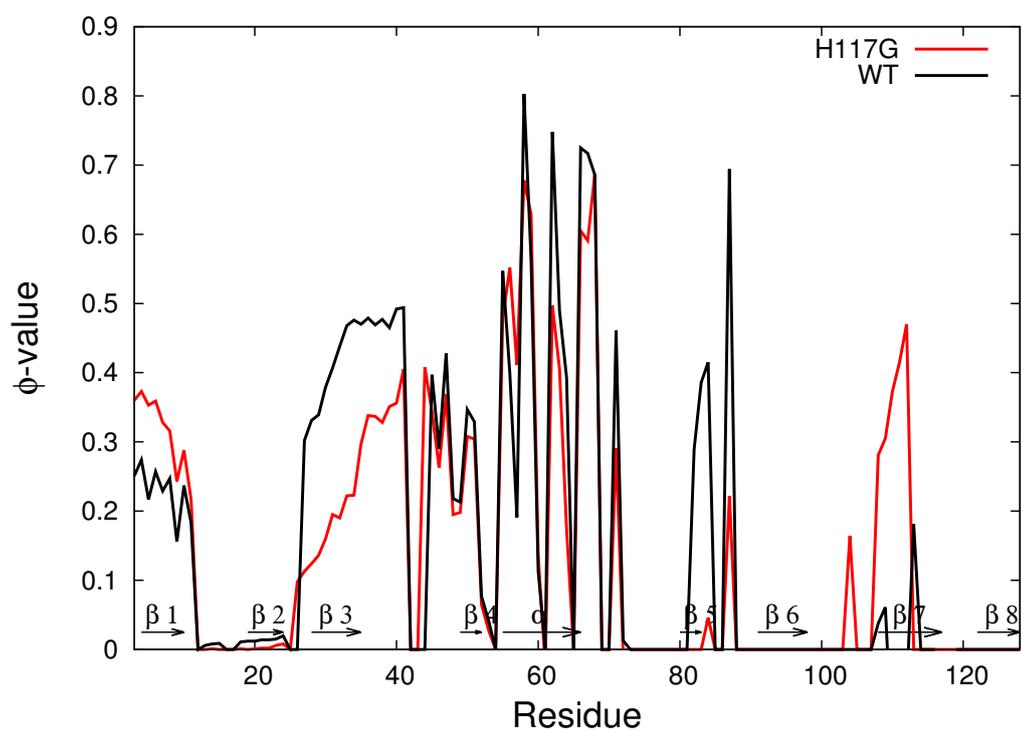
Figure 3.2: Comparison of Φ-value for each residue between mutated azurin and wild-type azurin at $T = T_f$.

Moreover, we also compare our result with the experimental data. From our mutated azurin unfolding simulation at folding transition temperature, the average Φ-value is found to be 0.148. It is appropriate with the experimental Φ-value for mutation His117 to Gly which is 0.1 ± 0.03 [51] or 0.1 ± 0.06 [52], and also with the theoretical data which is $\Phi \approx 0$ [51]. This mutation of His117 to Gly is found to give more stability to its nearby region, $\beta 7$, even though the mutated residue actually has almost no native contact with other residues. It means the non-native interaction also plays a role in our case.

3.5 Conclusion

Advancing computational study of protein folding is a great issue in biomolecular study. Based on two fundamental theories of principle of minimal frustration and energy landscape theory, we have performed the structure-based simulation of wild-type and mutated azurin. Present study shows that the mutation of His117 to Gly affects the stability of the denatured state. Both free energy profile and protein engineering method, Φ-analysis confirmed that the mutated azurin folds faster than the wild-type. In particular, the $\beta 7$ region, which is near the mutated residue, is found to be more stable compare to the wild-type. Nevertheless, in both types of azurin, the helix region which contains more local interactions has become the most native-like region at the transition state.

In short, our findings have found to be in agreement with both experimental and theoretical studies. Even so, currently we only use single order parameter Q, defined as the measure of nativeness, to characterize the changes in free energy and observe the native and denatured states. Further analysis may be needed to gain insights into the folding/unfolding mechanism of azurin.

Chapter 4

Implementation of Gō Model on Azurin Complex System

Previously we have applied Gō model on single chain of azurin via coarse-grained simulation. In this chapter, we will discuss the implementation of Gō model on multiple chains of azurin.

4.1 Introduction

Proteins play extremely important roles not only in human but also in other living organisms. They usually form complex interactions with other macromolecules, such as lipids, nucleic acids, or other proteins, to perform their biological functions [53]. In the last decades, this intermolecular interaction has become a great issue in the biophysics field. Other studies have conducted to advance the computational study on intermolecular interaction [23, 31, 54–56].

In our present study, we will focus on protein–protein interaction. Several studies have found that the formation of protein complex is affected by the presence of other proteins [31, 56]. Their interactions will tend to force the proteins to form compact configuration [57]. On the study of folding process, Gō has found that the long-range interactions play an important contribution on the stability of native conformation [21]. Inspired by his study, we predict

that the long range interaction formed by the intermolecular interaction also contributes to the conformational stability of the protein.

Here, our goal is to investigate the effects of protein–protein interaction on the conformational stability of protein complex. We developed a topology-based coarse-grained model to simulate several identical chains of azurin. In previous chapter, we have applied Gō-like model to simulate single chain of azurin. This model employed the principle of minimal frustration and the funneled energy landscape. In the similar way, we will treat the intermolecular interaction as we have treated the non-bonded interaction on the intramolecular interaction describing in chapter 3. These studies will provide important insights into the importance of native contacts into the stability of protein complexes.

4.2 Material and simulation methods

4.2.1 Model system

In this chapter, the native structure of azurin complex system was obtained from X-ray crystal structure of *Pseudomonas Aeruginosa* azurin (PDB ID: 4AZU) [15]. This crystal structure is composed of a tetramer of identical azurin molecules. Using this conformation, we build several systems consisting of dimer, trimer, and tetramer of azurin as shown in Figure 4.1. The blue chain is chosen as the representative chain that will be our focus in this observation. Meanwhile other chains act as the crowding agents.

Two dimer systems are presented here, dimer I and dimer II. Dimer I is an independent system where the distance between the dimer is more than the cutoff. Otherwise, dimer II is the interacted system obtained from the original crystal structure. We add one more chain in trimer system and two more chains in tetramer system. Both systems are also obtained from the crystal structure, so that the tetramer system actually is the unit cell of the crystal structure.
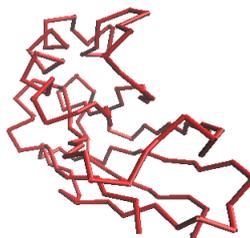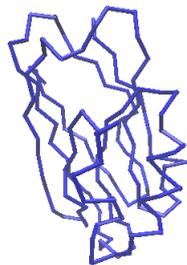
(a) Dimer system I
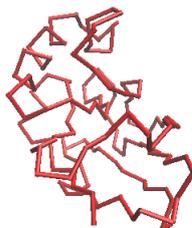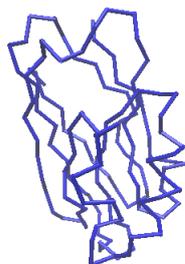
(b) Dimer system II
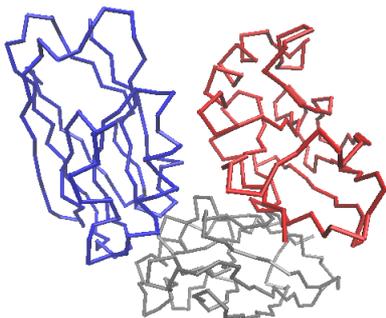
(c) Trimer system
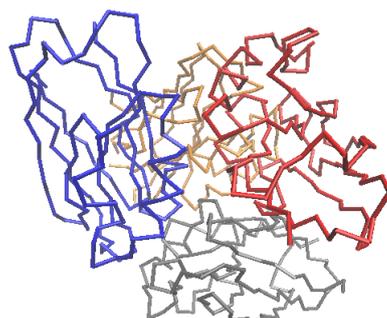
(d) Tetramer system

Figure 4.1: Initial configurations.

4.2.2 Potential model

We performed CG simulation with an implementation of native-structure based potential interaction to observe the dynamics of each configuration system. Our potential interaction is distinguished into intramolecular and intermolecular interactions. The off-lattice Gō-like model is employed using the same formula as in Table 3.2 to represent the intramolecular interactions. For the intermolecular interaction energy, we adopt the non-bonded term from Gō-like potential in Table 3.2 which can be written in the following formula:

$$E_{ij}^{\alpha\beta}(r) = \varepsilon_{\mathsf{nat}} \left[5 \left(\frac{\sigma}{r} \right)^{12} - 6 \left(\frac{\sigma}{r} \right)^{10} \right] + \varepsilon_{\mathsf{non\text{-}nat}} \left(\frac{C}{r} \right)^{12}. \qquad (4.1)$$

To avoid the ambiguity, we introduce the superscript $^{\alpha\beta}$ to distinguish the non-bonded interaction in intramolecular and intermolecular interactions. This indicator denotes the different chain. So, in Eq. (4.1), i and j represent i-th and j-th Gō particles of chain α and β respectively. Meanwhile in Table 3.2, i and j represent i-th and j-th Gō particles of unsubsequent particles in the same chain.

4.2.3 Simulation condition

As we have done in the non-bonded potential shown in Table 3.2, σ is set to be the reference pairwise distance obtained from the crystal structure. The same definition of native and non-native contact is also applied. By using this potential model, we simulate all systems with the same potential parameters and simulation condition as in the previous chapter. In addition, we avoid the translational and rotational movement of the system by setting the momentum and angular momentum of the whole system to zero during the simulation [58, 59] for every several steps. Our CG simulations were performed under constant temperature on the folding temperature as in Chapter 3.

4.3 Analysis methods

We calculated several properties to investigate the roles of the intermolecular interactions. We monitored the interchain dynamics by the autocorrelation of the distance between the centers of mass of pair-chains. Autocorrelation is a correlation between a time series with itself, so in our case this property can give us information whether the system remains in the same state from time to time. The autocorrelation can be calculated by the sufficient statistical average of the time series, as follows:

$$A(t') = \frac{\langle (x(t) - \langle x \rangle) \cdot (x(t + t') - \langle x \rangle) \rangle}{\langle x(t) - \langle x \rangle \rangle^2}, \tag{4.2}$$

where x is the time series property and t' is the time lag.

As our previous study in Chapter 3, we also compare the thermodynamical property using the free energy profile. The same method was used in this study, where the free energy profile is obtained by the histogram method [46] with the fraction of nativeness (Q) as the reaction coordinate.

4.4 Results

Each system was simulated at the residue level under constant folding temperature T_f. For our convenience, we will focus on two representative chains, called A and B, and compare their dynamics in all systems. Their interchain distances (d_{AB}) can be seen in Figure 4.2. These distances represent the distance between the centers of mass of those chains.

Figure 4.2 shows that the interchain distance of dimer system I which has no intermolecular interaction is more fluctuating than the other dimer systems. In the dimer system I, the non-bonded interaction only involves the non-native interaction which is a repulsive interaction, since in this system the dimer does not have native contact. So when the dimer becomes closer, it tends to repel and move away. On the other hand, the dimer system II

which has native contacts is much more stable. It shows that native contact plays a significant role on the interchain interaction. Besides, the crystal structure is believed to be the most stable conformation of the system.

Furthermore, the stability is also affected by the crowding system. If we compare the dimer system II, trimer system, and tetramer system which are all taken from the crystal structure, Figure 4.2 shows that the interchain distance becomes less fluctuating as the system becomes more crowded. This comparison clearly shows the importance of native contacts on the dynamics of protein complex as well as our comparison of the dimer systems. The standard deviation in Table 4.1 also confirms that the system with higher compactness has smaller deviation and reaches equilibrium time faster as shown in Figure 4.3.
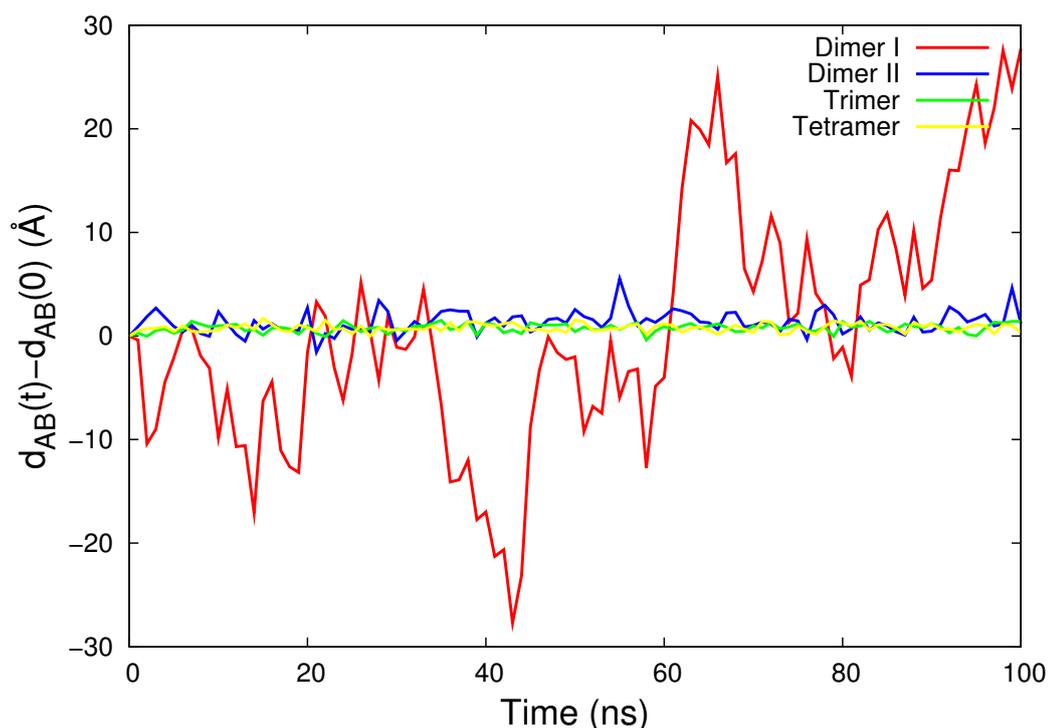
Figure 4.2: Time series of the interchain distances.

Table 4.1: Standard deviation of the interchain distances

System	chain A-B	chain A-C	chain A-D
Dimer I	11.448	-	-
Dimer II	0.944	-	-
Trimer	0.492	0.329	-
Tetramer	0.39	0.243	0.441
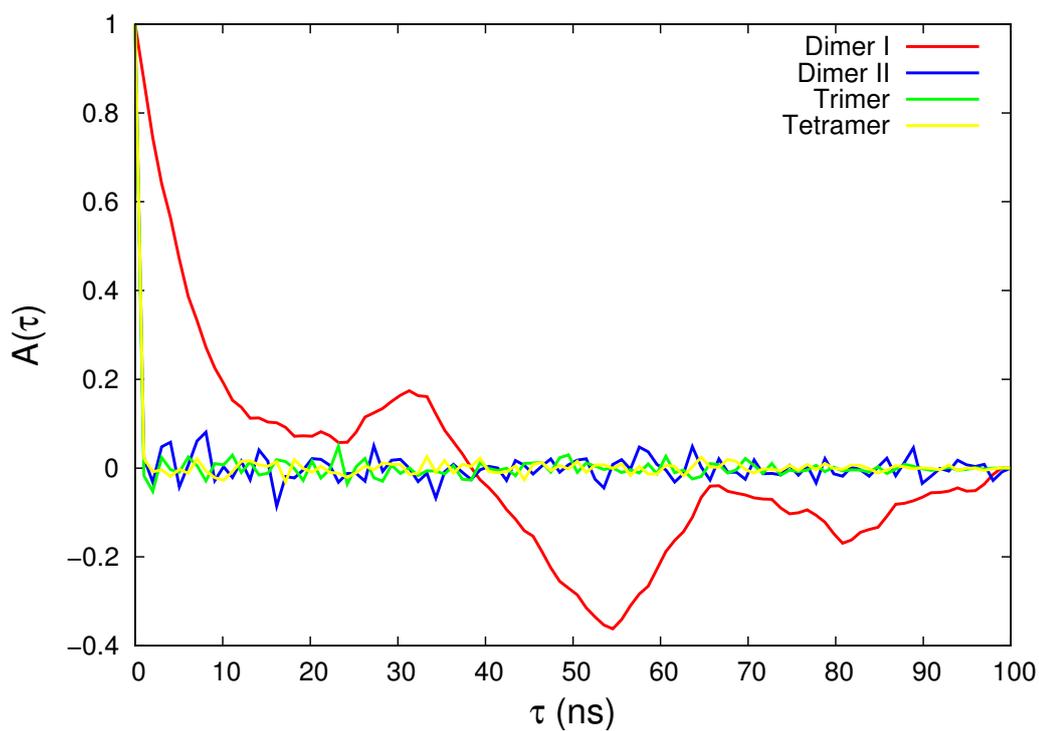
Figure 4.3: Autocorrelation of the interchain distance as a function of time lag (τ).
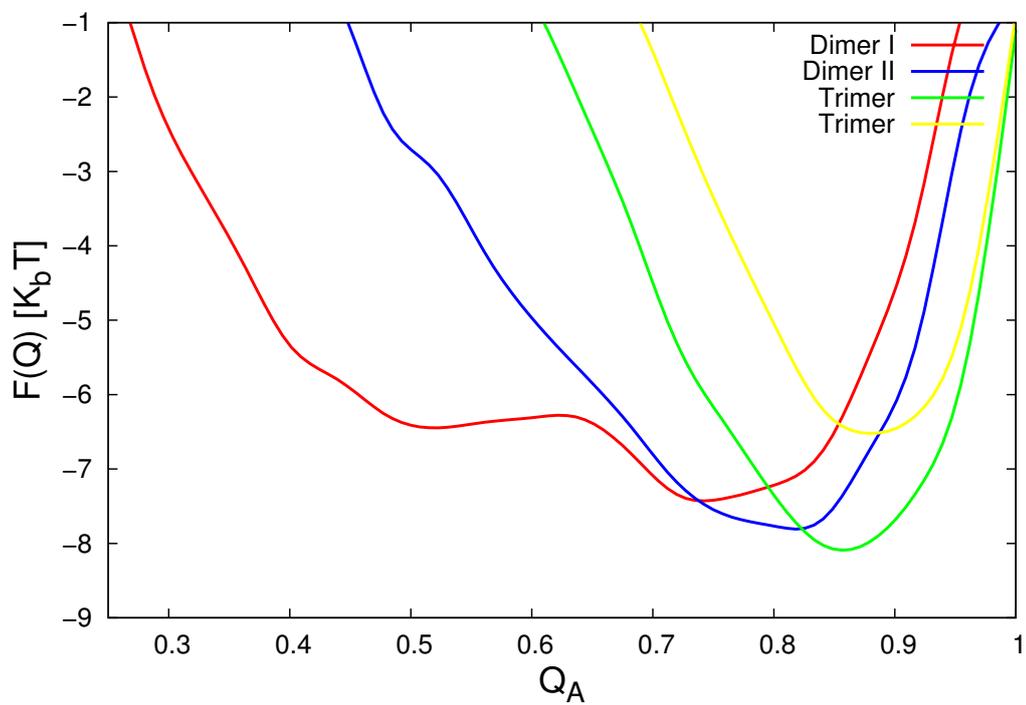
Figure 4.4: Free energy profile as a function of reaction coordinate Q which is the measurement of the nativeness.

Moreover, we also investigated the conformational change to see how the crowding affects the structure of individual chain. As in previous chapter, we obtained the free energy profile by using the nativeness measurement (Q) as the reaction coordinate. This free energy profile is shown in Figure 4.4. The tetramer system is found to be the most native-like configuration. It is very natural since in the more crowded or more compact system, each residue will have less space to move, so they tend to keep the optimal distance.

From the viewpoint of the potential model, the systems will tend to keep as nearly as possible to the native structure because Gō-like potential model minimizes the topological frustration. Nevertheless, it means that this model is very dependent to the native structure. So when the native structure do not have native contact, as in dimer system I, the system will not have the attractive interaction. Meanwhile in the real system, when they become closer and reach the contact distance, the attractive interaction exists.

4.5   Conclusions

We found that the native contact plays an important role on the dynamics of the protein complex system. Our studies also found that more crowded and compact system affects the protein movement as well. In consequence, the tetramer system which is the unit cell of the crystal structure, naturally has the most stable and native-like configuration over other systems.

Gō-like model can be used to reproduce the native crystal structure very well. However, we have to consider the dependence of current intermolecular potential model to the presence of native structure. More general potential model might be considered to represent more realistic interaction when we start from an independent system.

Chapter 5

The Effects of Intermolecular Interactions to The Conformational Changes of Azurin Complex

In the following sections, we employ widely used Lennard-Jones potential for intermolecular interaction and investigate the conformational changes of azurin complex.

5.1   Introduction

Most proteins perform their biological function by associating to form protein complex. This association involves the protein–protein interaction. Recent studies show that noncovalent binding can influence the protein stability [21, 60]. Therefore, advancing study of protein–protein interactions becomes very important for better understanding of protein function.

From the viewpoints of coarse-grained simulation, determining the protein–protein interactions is still a great mystery. Many studies have modeled the intermolecular interactions in protein complexes [31,56,61]. Each model has the advantages and disadvantages depending on the goal of their studies.

In the previous chapter we have shown that non-bonded potential adopted from the Gō-like model gives good result to reproduce the native conformation by assuming that the crystal structure is the native structure. Nevertheless, the dependence on the native structure restrict us for more general implementation, such as for a larger system than the native structure. Here, we will apply more general intermolecular potential and investigate the conformational changes of azurin complex. We treat each chain as rigid as possible by employing the off lattice Gō-like model to represent the bonded and non-bonded intramolecular interactions [21,22]. Meanwhile the intermolecular interaction is represented by the 6-12 Lennard-Jones (LJ) potential with general parameters [62]. We will observe the stability of azurin complex by analyzing the conformational change and total surface area.

5.2   Material and simulation methods

5.2.1   Model systems

As in the previous chapter, currently we build several systems consisting of dimer, trimer, and tetramer of identical azurin as shown in Figure 5.1. System I, IV, and V are taken from the original crystal structure (4AZU). System II and III are modified dimer systems which have intermolecular interaction with different contact areas.
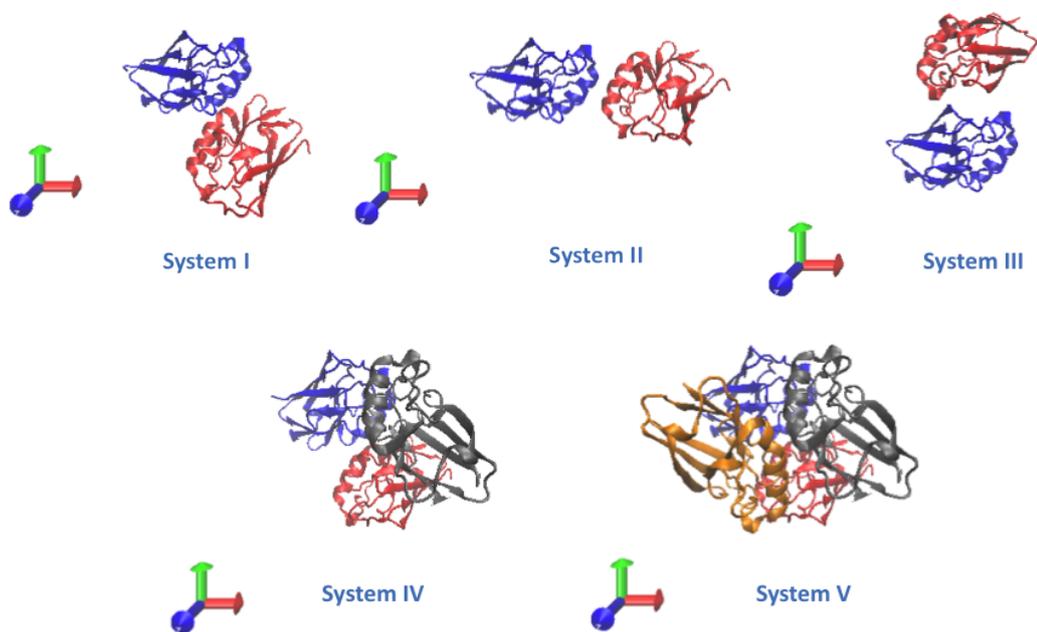
Figure 5.1: Various initial conformations. System I, II, and III are dimer systems with different contact orientation.

5.2.2 Potential model

We carried out coarse-grained simulations by combining native-structure based potential interaction for the intramolecular interaction and physics-based potential for the intermolecular interaction. The same intramolecular potential will be used to minimize the topological frustration of each chain, since we will focus more on the protein–protein interaction. The widely used 6-12 Lennard-Jones potential will be applied as the intermolecular interaction to describes the van der Waals term [62, 63]. This potential is defined as follows:

$$U_{LJ}(r) = 4\varepsilon_{LJ} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{5.1}$$

where r represents the pairwise distance between two residues from different chains and σ is the distance where the intermolecular potential between two residues is zero.

5.2.3 Simulation condition

In current study, we redefine the units for our coarse-grained model using the same quantities as our previous units, which are length (σ_0), mass (m), time (τ), and energy (ε_0). The values of our new coarse-grained units are listed in Table 5.1. The values of σ_0, m, and ε_0 are determined from the average van der Waals radii of azurin, the average mass of azurin, and the temperature of system, respectively. The time unit (τ) is calculated by the same method as in Chapter 3. We also redefine the potential parameter and simulation condition as in Table 5.2.

Table 5.1: New units of coarse-grained model

CG units		Experimental units
length	σ_0	5.7 Å
mass	m	110 amu
energy	ε_0	0.6 kcal/mol
time	τ	3 ps

Table 5.2: Potential parameter and simulation condition

Parameter		Value in kcal/mol
Virtual bond-stretching	K_{bond}	$50.0\ \text{Å}^{-1}$
Virtual bond-angle bending	K_θ	10.0
Virtual bond-torsional term	K_ϕ	1.0
Native	ε_{nat}	1.0
Non-native	$\varepsilon_{\text{non-nat}}$	1.0
Intermolecular	ε_{LJ}	0.4 or 0.13
Others		**Value**
LJ distance	σ	$6.5\ \text{Å}$
Friction coefficient	ζ	$0.5\ (\tau^{-1})$
Temperature	T	$300\ \text{K}$

For the estimation of σ and ε_{LJ} for intermolecular interaction, we consider the correlation with the non-bonded potential model for intramolecular interactions as shown in Table 5.3. It is not easy to clearly obtain the value of those parameters. The r_0 on intramolecular interaction represents specific distance obtained from the crystal structure where each pair has different value. Meanwhile σ, set as general parameter for all pairs, usually represents the particle size. Here we use $\sigma = 6.5$ Å, because the interacting residues within 6.5 Å is found to contribute significantly to the protein–protein association [61]. By comparing the minimum value of the potential, ε_{LJ} should be smaller than ε_1 since intermolecular interaction is weaker than the intramolecular interaction.

Table 5.3: Comparison of two non-bonded potential

	Intramolecular 10-12 LJ potential	Intermolecular 6-12 LJ potential
Potential ($U(r)$)	$\varepsilon_1\left[5\left(\frac{r_0}{r}\right)^{12} - 6\left(\frac{r_0}{r}\right)^{10}\right]$	$4\varepsilon_{LJ}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$
$U(r) = 0$	$r = \sqrt{5/6}\,r_0$	$r = \sigma$
Minimum $U(r)$	when $r = r_0$	when $r = 2^{1/6}\sigma$
	$U(r_0) = -\varepsilon_1$	$U(2^{1/6}\sigma) = -\varepsilon_{LJ}$

5.3 Analysis methods

We measured the structural stability of azurin complex through the root mean square displacement (RMSD) with respect to the initial structure. A least-square fitting on given structure to the initial structure is performed to obtain minimal RMSD. We also analyzed the total surface area of the system to gain insight into the accesibility of the system to a solvent. This concept was first introduced by Lee and Richards [64]. Our calculation applied statistical approach for faster calculation of accesible surface area, proposed by Wodak and Janin [65], and was performed by using POPS program [66].

5.4 Results

In this section we will explain our analyses into two part. First, we investigate the effects of intermolecular interaction strength to the stability of the system. We will compare two parameter values of ε_{LJ} as in Table 5.2 for the simulation of system I. Later, we found that $\varepsilon_{LJ} = 0.13$ kcal/mol is better and we will use it to simulate the other systems. Second, we evaluate physical properties for all systems as described in previous section.

Higher intermolecular potential parameter represents stronger interaction, so we expect that two chains will tend to get closer and the buried area will increase. Our expectation is well confirmed as shown in Figure 5.2. This figure shows that the conformation of both chains are starting to denature as indicated by the steady increment of RMSD values of both chains. This denaturation is appropriate with the aggregation possibility that is shown by the decrease of total surface area (Figure 5.3). On the other hand, Figure 5.4(a) shows that the simulation with smaller ε_{LJ} gave more stable conformation. These findings show that strong attractive interaction may lead the system to aggregation.
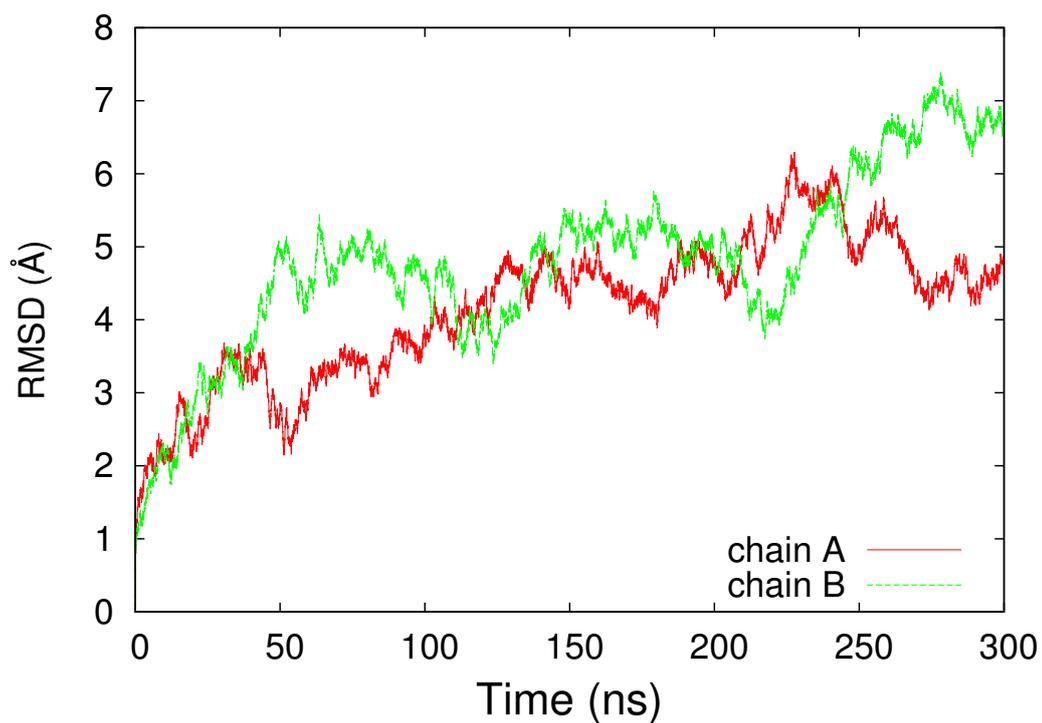
Figure 5.2: The root mean square displacement shows the conformational changes compared to the given initial configuration caused by the strong intermolecular interaction.
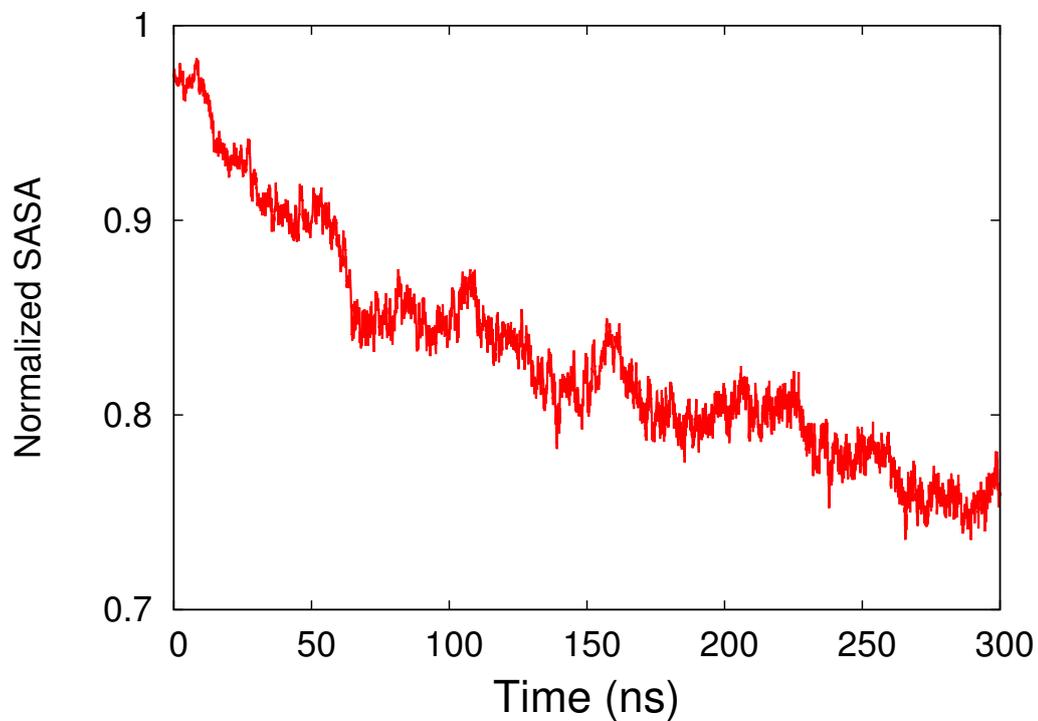
Figure 5.3: The total surface area is calculated at the residue level and is normalized by the total surface area of independent chains where the actual surface area of one chain is 6735.87 Å^2.
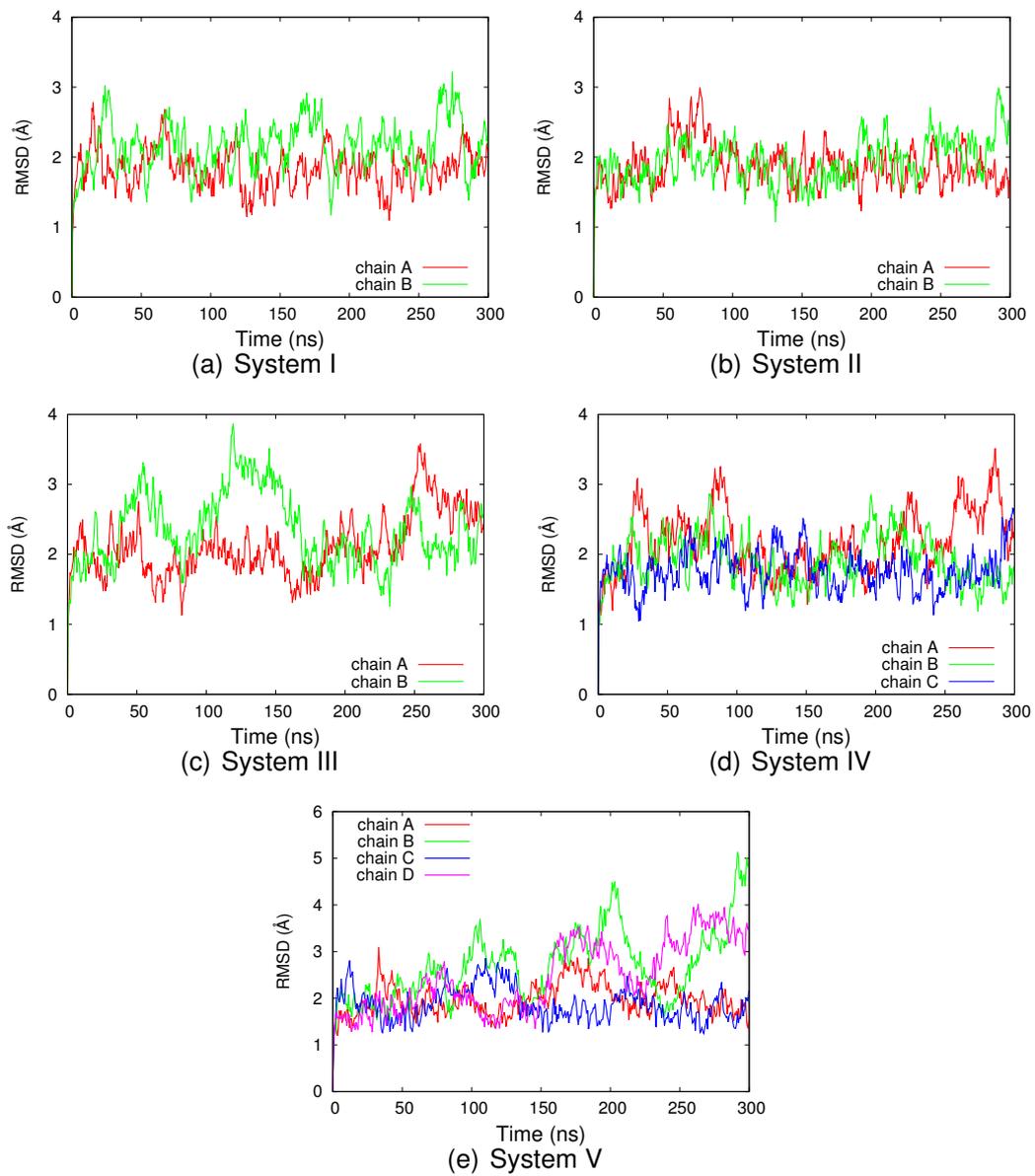
Figure 5.4: The root mean square displacement of the simulations with $\varepsilon_{LJ} = 0.13$ kcal/mol.

Table 5.4: Quantitative comparison of the RMSD of each system is represented by the calculation of average (Å) and standard deviation (Å) of RMSD.

System	Average(Standard deviation)			
	Chain A	Chain B	Chain C	Chain D
I	1.84(0.28)	2.15(0.37)	-	-
II	1.88(0.30)	1.94(0.30)	-	-
III	2.11(0.44)	2.35(0.50)	-	-
IV	2.17(0.45)	1.91(0.34)	1.77(0.28)	-
V	1.98(0.34)	2.68(0.76)	1.87(0.35)	2.46(0.74)

Table 5.5: Comparison of the initial and average surface area (Å^2), initial number of pair contacts, and average total energy (kcal/mol).

System	Total SASA (Å^2)		initial	Av. total
	Initial	Average	#contacts	energy
I	12982.76	13750.75	13	125.99
II	13178.79	14186.64	8	131.03
III	12897.96	13827.58	19	126.78
IV	18615.81	19931.83	35	164.86
V	24108.12	25705.4	82	211.12

Now let us investigate the dynamics of azurin complex regarding to the role of initial contacts. We measure the structural stability of each chain from the given configuration against the initial state and compare the behavior of all systems as shown in Figure 5.4. Figure 5.4 shows that the conformations of azurin remain relatively stable at all systems. This finding is also confirmed quantitatively in Table 5.4. The averages of RMSD are below 3 Å with relatively small deviation.

We also monitor the total surface area (SASA) of azurin complex at residue level. By this calculation we can investigate the buried area in the binding site. From our calculation, the total surface area of all system increases about 1,000 Å^2 (see Table 5.5). Meanwhile, Figure 5.5 shows that the ratio of total SASA of the azurin complex to total SASA of independent chains decreases. This decrease represents the increase of buried area, which indicates the strong attractive interaction in the binding site.

Moreover, Figure 5.6 shows that the ratio of SASA decreases along with
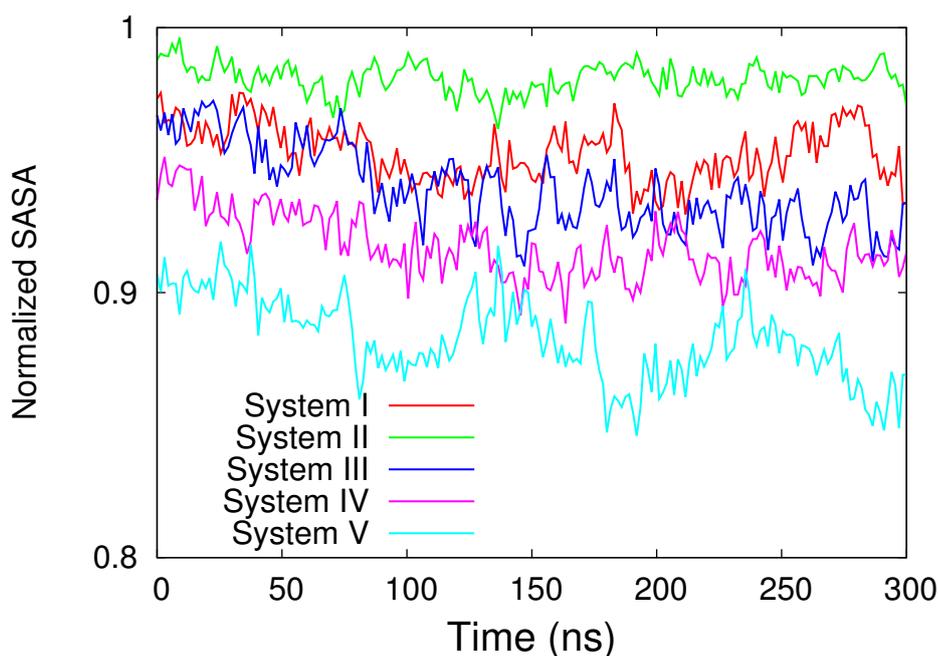
Figure 5.5: The normalized total surface area of the simulations with $\varepsilon_{LJ} = 0.13$ kcal/mol.

the increase of the number of pair contacts. However, the dimer systems have steeper slope than the larger systems, which means that the number of chains also plays an important role in the intermolecular interaction. The relation between ratio of SASA and number of contacts is very reasonable since when the contact between two particles is formed, they become closer and the buried area becomes larger. On the large systems, the space for their movements becomes less so that even though the initial contacts is large, the change during simulation may not be significant. In addition, the time series of the number of contacts in binding area (Figure 5.7) confirms that as the system becomes more crowding, the increase of the number of contacts becomes slower.
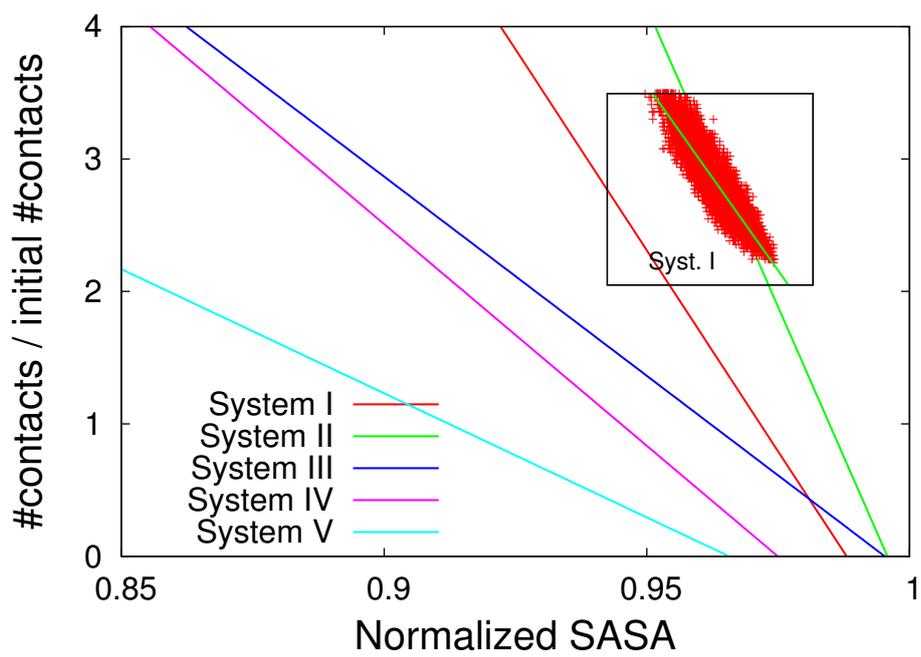
Figure 5.6: The regression of normalized SASA to the ratio of number of pair contacts. The inset shows the data set with fitted regression line.
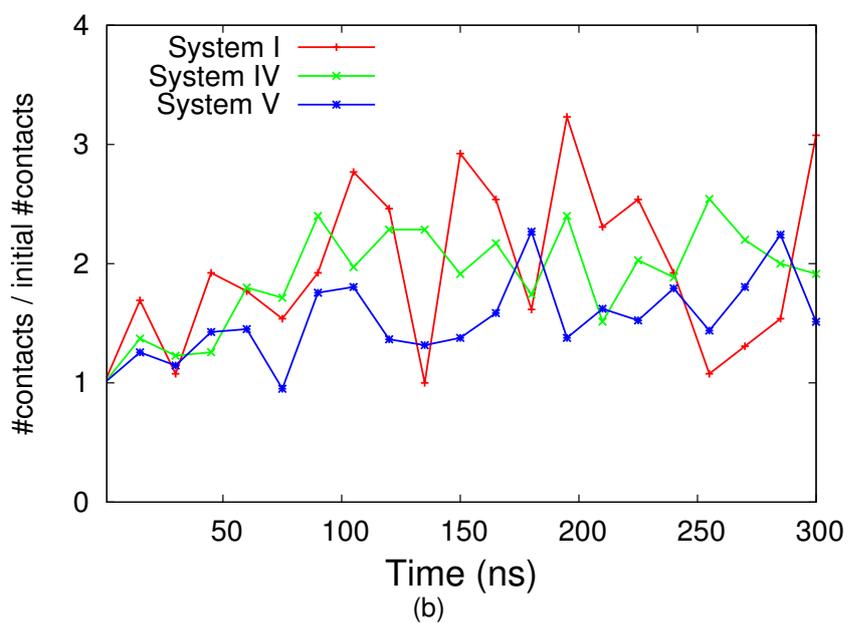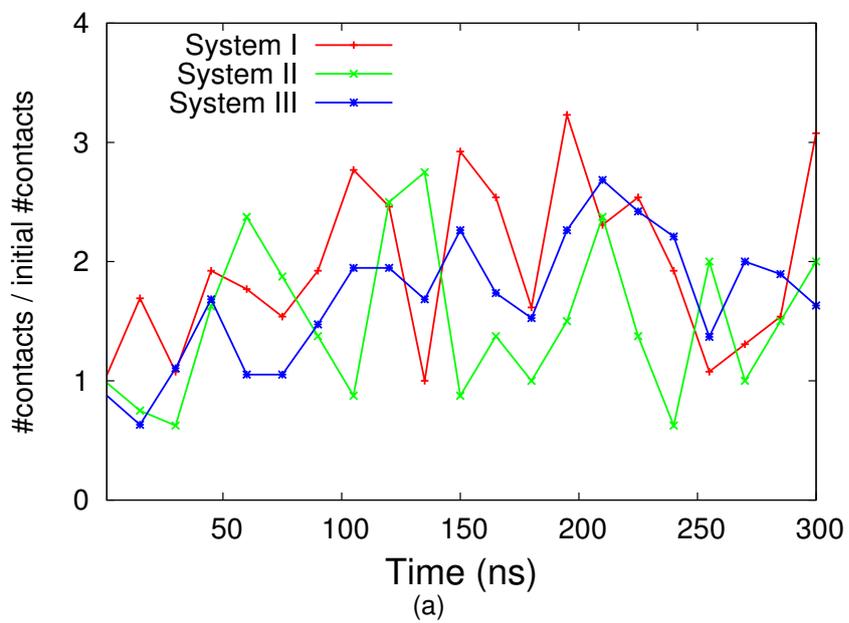
Figure 5.7: Number of contacts in the binding area. (a) Comparison among dimer systems. (b) Comparison among systems from the original PDB file.

45

5.5 Conclusion

Development of intermolecular potential model is very important in protein–protein interaction studies. To use general form of potential model which do not depend on the native structure, we have to consider what kind of properties should be involved and what properties can be neglected. Our study suggested that the choice of coefficient parameter for intermolecular interaction can cause the azurin to bind to the other chain.

One of two potential parameters that we use in this study indicates the possibility of binding chains in azurin complex even though both parameters satisfy our limitation that intermolecular interaction should be weaker than intramolecular interaction. We suppose that we need to treat the intermolecular interaction, especially in the binding area, more carefully by considering more physical informations from the crystal structure.

Moreover, due to the initial conformation, we found that crowded system also plays an important role on the stability of protein complex. In this study, all systems tend to form more packed conformation. However, as the system is getting more crowded, the increase of the number of contacts becomes slower. Overall, more considerations are needed to develop the potential model that can be used for more general implementation, especially on the estimation of the parameters.

Chapter 6

General Conclusion

In this thesis, we observed the formation mechanism of azurin complex by using coarse-grained models. We started it by performing coarse-grained simulation of a single azurin as described in chapter 3. We carried out coarse-grained simulation via implementation of Gō model to observe the unfolding process of wild-type azurin and mutated azurin (H117G). This model was developed based on two fundamental theories of principle of minimal frustration and energy landscape theory. We found that the mutation of His117 to Gly affects the stability of the denatured state and the mutated azurin folds faster than the wild-type.

In the next step, our studies are followed by developing coarse-grained models for azurin complexes as explained in chapter 4 and 5. In chapter 4, we adopted the non-bonded interaction of the Gō-like model into the intermolecular interaction in azurin complex. This potential is very good to maintain the stability of protein complexes since this model minimize the topological frustration. Nevertheless, this potential also has limitation. Even though this model can reproduce the azurin complex, it lacks of transferability. We can only apply this model to a particular system which means that the existence of known native structure is a must on this model.

To overcome this limitation, we explored more general potential model by adopting the widely used Lennard-Jones potential as described in chapter 5. However, determining the parameters becomes a great challenge. We found

that this approach has not yet accurately reproduced the azurin complex. Intermolecular interactions in protein complexes often can not be derived into a simple model. Being simple and oversimplified are two different things. We have to carefully determine which phenomena can be simplified and which should be described with more complicated models.

Developing accurate and transferable coarse-grained potential for protein complex remains a challenge. In the future, knowledge-based approaches may be used by employing physical informations from known PDB structure to develop a set of transferable and more appropriate interactions for azurin complex. When this problem is addressed, it would significantly improve the scope of coarse-grained protein model to be able to predict unknown structure for new systems.

Appendix A

Extended research:
An improved coarse-grained model of azurin complex via bottom-up approach

A.1 Introduction

In recent times, coarse-grained simulation has gained much attention due to its ability to overcome the time and size problems of the all-atom molecular dynamics simulation. Many researchers have advanced coarse-grained models to study the protein dynamics by various approaches [22, 23, 67]. Several models have been successfully applied to study the protein folding. Nevertheless, coarse-grained models involving protein–protein interaction are still limited. Meanwhile in real system, proteins tend to form a complex to perform its function where the formation of complex system involves protein–protein interaction.

In our previous work, we found that Gō-like model can be applied to represent the intermolecular interaction. This model has good accuracy but lacks on transferability. This model can not be applied to the unknown struc-

ture. On the other hand, we can not easily simplify the intermolecular interaction. Therefore, development of transferable coarse-grained model which represents the interactions of protein complex system is needed.

In recent work, we develop a coarse-grained potential model to simulate azurin complex, which is known as a rigid protein [4]. We introduce new parameter representing the strength of attractive interaction. To estimate our parameters, the intermolecular contact is evaluated by bottom-up approach from the native structure. Since azurin complex is known as a close-packed complex, our model will be tested to reproduce a native tetramer of azurin. This study will offer a new coarse-grained model with better accuracy and transferability. Moreover, our model will be a promising approach for the intermolecular potential model for unknown structure.

A.2   Material and simulation methods

A.2.1   Model systems

As we have mentioned in previous section, our first goal in this work is to reproduce a native tetramer azurin. Currently we also employ crystal structure of wild-type azurin with pdb entry: 4AZU (Figure 2.1) [15]. This conformation consists of four identical chains of azurin.

A.2.2   Potential model

In current work, we treat the individual chain as a rigid monomer. Therefore, we employ native-center based Gō-like model as in previous works to represent the intramolecular interaction. Meanwhile, we extend the intermolecular potential in Chapter 5 by knowledge-based approach. This mathematical formulation of modified Lennard-Jones potential is represented in the following equation:

$$U_{ij}^{\alpha\beta}(r) = 4\varepsilon_{\text{inter}} \left[\left(\frac{\sigma}{r} \right)^{12} - a_{ij}^{\alpha\beta} \left(\frac{\sigma}{r} \right)^{6} \right]. \tag{A.1}$$

Our previous studies in Chapter 4 and 5 suggested that we need to develop more specific potential. In this work, we introduce our new parameter, $a_{ij}^{\alpha\beta}$, to specify the attractive term. This parameter represents the strength of attractive interaction.

In order to define this parameter, first we determine the zero-potential and the maximum depth of the potential well of our potential model. From Equation (A.1), we have:

1. The zero-potential, $U_{ij}^{\alpha\beta}(r) = 0$, is obtained when:

$$\left(\frac{\sigma}{r}\right)^{12} - a_{ij}^{\alpha\beta}\left(\frac{\sigma}{r}\right)^6 = 0$$

$$\left(\frac{\sigma}{r}\right)^6 = a_{ij}^{\alpha\beta}$$

$$r = \frac{\sigma}{(a_{ij}^{\alpha\beta})^{1/6}}; \tag{A.2}$$

2. The maximum depth of the potential well is obtained when $U_{ij}^{\alpha\beta}(r)$ reaches its minimum. This condition is satisfied when:

$$\frac{\delta}{\delta r}\left(U_{ij}^{\alpha\beta}(r)\right) = 0$$

$$4\varepsilon_{\text{inter}}\frac{1}{r}\underbrace{\left[12\left(\frac{\sigma}{r}\right)^{12} - 6a_{ij}^{\alpha\beta}\left(\frac{\sigma}{r}\right)^6\right]}_{=0} = 0$$

$$12\left(\frac{\sigma}{r}\right)^{12} = 6a_{ij}^{\alpha\beta}\left(\frac{\sigma}{r}\right)^6$$

$$2\left(\frac{\sigma}{r}\right)^6 = a_{ij}^{\alpha\beta} \tag{A.3}$$

$$r^6 = \frac{2\sigma^6}{a_{ij}^{\alpha\beta}}$$

$$r = \frac{2^{1/6}\sigma}{(a_{ij}^{\alpha\beta})^{1/6}}. \tag{A.4}$$

So, the maximum depth of the potential well becomes:

$$U_{ij}^{\alpha\beta}\left(\frac{2^{1/6}\sigma}{(a_{ij}^{\alpha\beta})^{1/6}}\right) = 4\varepsilon_{\text{inter}}\left[\left(\frac{\sigma(a_{ij}^{\alpha\beta})^{1/6}}{2^{1/6}\sigma}\right)^{12} - a_{ij}^{\alpha\beta}\left(\frac{\sigma(a_{ij}^{\alpha\beta})^{1/6}}{2^{1/6}\sigma}\right)^{6}\right]$$

$$= 4\varepsilon_{\text{inter}}\left(\frac{(a_{ij}^{\alpha\beta})^2}{4} - \frac{(a_{ij}^{\alpha\beta})^2}{2}\right)$$

$$= -\varepsilon_{\text{inter}}(a_{ij}^{\alpha\beta})^2 \tag{A.5}$$

From the relation (A.3), we could determine the parameter for attractive term, $a_{ij}^{\alpha\beta}$. Since $a_{ij}^{\alpha\beta}$ will be specified for each inter-particle of two interacting chains, we can rewrite the relation (A.3) into:

$$a_{ij}^{\alpha\beta} = \left(\frac{2^{1/6}\sigma}{r_{ij}^{\alpha\beta}{}_{crys}}\right)^{6}, \tag{A.6}$$

where $r_{ij}^{\alpha\beta}{}_{crys}$ is obtained from the native structure. To accomplish the transferability problem, again we will simplify this parameter by bottom-up strategy.

Originally $a_{ij}^{\alpha\beta}$ is a parameter for particular known structure. In our case, we have six dimers and each inter-particle for each dimer has unique parameter. Our mission is to provide parameter that can be applied for every dimer, yet unique for each inter-particle. Or mathematically we can explain it as follows. Let α, β, and γ be three different chains. Our parameter should satisfy:

$$a_{ij}^{\alpha\beta} = a_{ij}^{\alpha\gamma} = a_{ij}^{\beta\gamma}, \tag{A.7}$$

with i and j are two particles from different chains, e.g. $a_{56}^{12} = a_{56}^{13} = a_{56}^{23}$.

Those six dimers are packed in the condition where the dimer interface between two chains is quite similar to the dimer interface between two remaining chains. So now we can reduce our cases into three dimers: dimer A–B, dimer A–C and dimer A–D. In this work, we choose the max $a_{ij}^{\alpha\beta}$ among three dimers for each inter-particle.

The remaining problems are the estimation of other two parameters, σ

and $\varepsilon_{\text{inter}}$. We first determine σ as the particle size by taking the average of van der Waals (vdW) radii of azurin. From the vdW radii in Table A.1, we get $\langle \sigma_i \rangle = 5.74$ Å. Meanwhile, in order to estimate $\varepsilon_{\text{inter}}$, we adapt the Lennard-Jones interaction strength introduced by Kim and Hummer [31].

Table A.1: Van der Waals radii (in Å) for 20 amino acids [68].

ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE
5.0	6.6	5.7	5.6	5.5	6.0	5.9	4.5	6.1	6.2
LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
6.2	6.4	6.2	6.4	5.6	5.2	5.6	6.8	6.5	5.9

Estimation of potential parameter $\varepsilon_{\text{inter}}$

In their work, they modeled the intermolecular interaction in the following way. If a pair-residue experiences an attractive interaction, the potential form is given by:

$$\phi_{ij}(r) = 4|\varepsilon_{ij}| \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} \right], \ \varepsilon_{ij} < 0. \tag{A.8}$$

Otherwise for repulsive interaction, $\varepsilon_{ij} > 0$,

$$\phi_{ij}(r) = \begin{cases} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} \right] + 2\varepsilon_{ij}, \text{ if } r < r_{ij}^0, \\ -4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} \right] \text{ if } r \geq r_{ij}^0, \end{cases} \tag{A.9}$$

where $r_{ij}^0 = 2^{1/6}\sigma_{ij}$, and σ_{ij} is residue-dependent interaction radius given by $(\sigma_i + \sigma_j)/2$.

They adapted knowledge-based statistical contact potentials obtained by Miyazawa and Jernigan to determine the potential parameters ε_{ij} [23]. The original pair contact potentials, e_{ij}, are empirically rescaled as follows:

$$\varepsilon_{ij} = \lambda(e_{ij} - e_0), \tag{A.10}$$

where e_0 is an offset parameter that balances the preference of inter-particle interactions relative to particle–solvent interactions, while λ scales the strength of the inter-particle interactions compared to the physical electrostatic inter-

actions. These two parameters are determined by fitting against experimental data. In Kim's paper [31], they used various values. Consistently good results are found on the following two models:

- $\lambda = 0.159$ and $e_0(k_BT) = -2.27$;

- $\lambda = 0.192$ and $e_0(k_BT) = -1.85$.

In principle, Miyazawa and Jernigan approximated the contact potentials by the number of contacts [23]. They define a contact between two residues if their distance is less than 6.5 Å. Then they estimate e_{ij} by Bethe approximation:

$$\exp(-e_{ij}) = \frac{\bar{n}_{ij}\bar{n}_{00}}{\bar{n}_{i0}\bar{n}_{j0}}. \tag{A.11}$$

n_{ii} is the total number of contacts between two residues of the same type, i. Meanwhile $n_{ij} + n_{ji}$ is the total number of contacts between i and j types of amino acids, which means i and j are ranged from 1 through 20. Subscript 0 is used to represent effective solvent.

The average of e_{ij} lies about -4.0 in k_BT. Then $\langle \varepsilon_{ij} \rangle$ will be:

$$\begin{aligned}
\langle \varepsilon_{ij} \rangle &= \langle \lambda \rangle \left(\langle e_{ij} \rangle - \langle e_0 \rangle \right) \\
&\approx 0.2(-4.0 - (-2.0)) \\
&\approx -0.4 \ k_BT \\
&\approx -0.24 \ \text{kcal/mol}
\end{aligned} \tag{A.12}$$

By this approximation, in the present work, $\varepsilon_{\text{inter}}$ is set to be 0.2 kcal/mol. Other parameters and simulation conditions are set to be the same values as in Chapter 5.

A.3   Analysis methods

We performed analysis on stability of tetramer azurin by measuring several structural properties. We analyze the motion of azurin through the root mean square displacement (RMSD) with respect to the initial structure as

we have done in Chapter 5. We also analyze the fluctuation of each particle of azurin by calculating the root mean square fluctuation (RMSF). We also monitor the number of contacts in the contact regions where a pair-residue belongs to a contact region if the distance is under 6.5 Å [61]. Moreover, the surface area of tetramer azurin is also calculated [64–66].

A.4   Results

We carried out coarse-grained simulation of tetramer azurin for 30 ns under constant temperature. Final conformation of tetramer azurin after simulation is shown in Figure A.1(b). Compared with the initial conformation in Figure A.1(a), the final conformation is less compact.
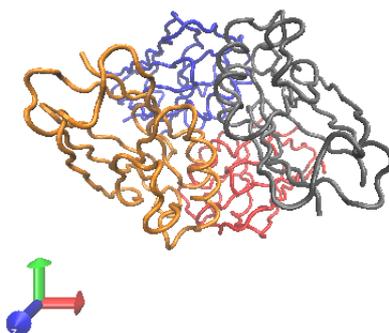
(a) Initial conformation (4AZU)
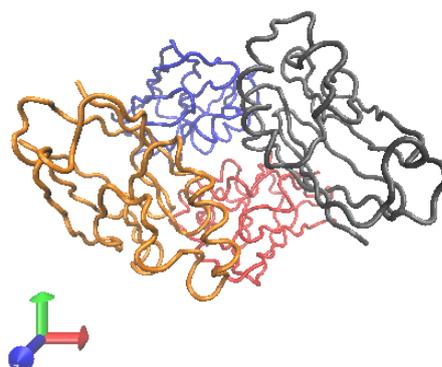
(b) After 30 ns

Figure A.1: Snapshots of the tetramer azurin from our simulation.
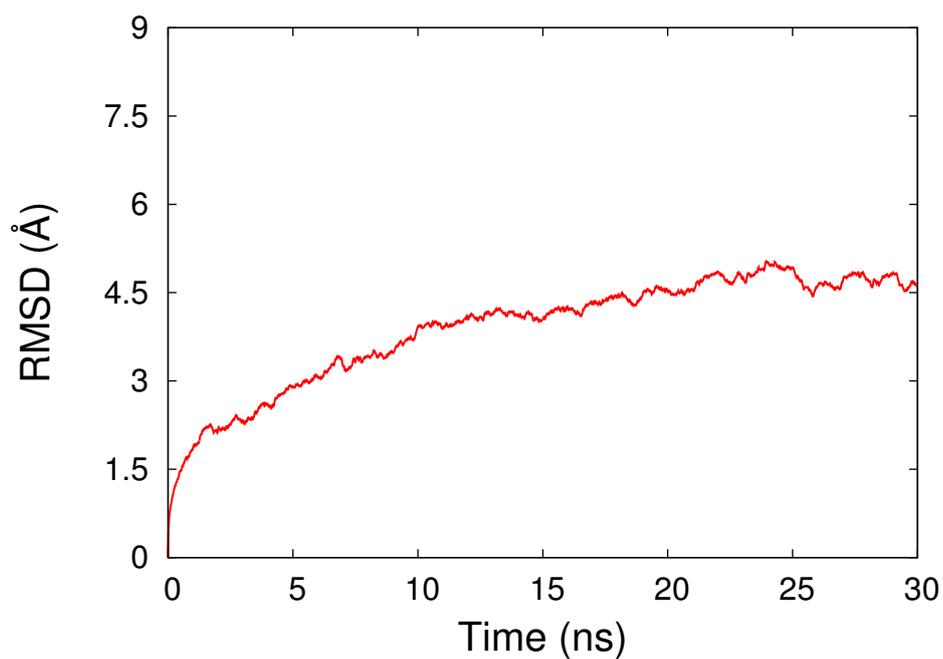
Figure A.2: RMSD profile for whole system.
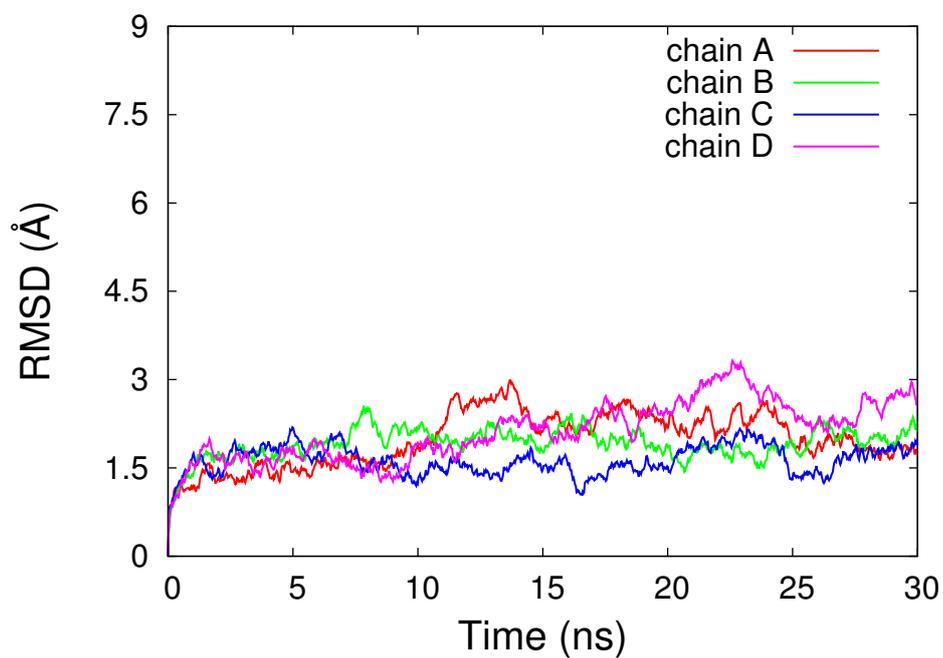
Figure A.3: RMSD profile for each chain.

Nevertheless, the RMSD measurement shows that the displacements are still tolerable since it ranged below 5 Å as shown in Figure A.2. We also measured the RMSD for each chain and our calculations show that there is no significant deformation on each individual chain as shown in Figure A.3.

Compared with our study in Chapter 5, we found that our new model is better to reproduce native tetramer azurin. In Chapter 5 we found the indication of deformation of each chain on simulation of tetramer azurin. Our new potential model is able to overcome the aggregation possibility that appears in the previous study.

In addition, we also performed an all-atom molecular dynamics simulation of tetramer azurin with explicit water solvent as comparison. The all-atom simulation was conducted using NAMD version 2.9 [69] with the CHARMM27 force field [70]. This simulation was performed under constant temperature, 300 K, controlled by Langevin dynamics.

We measured the residue fluctuation for each individual chain as shown in Figure A.4 from our coarse-grained simulation and in Figure A.5 from all-atom simulation. Those figures show that proteins in our coarse-grained simulation are more fluctuating than in all-atom simulation. However, RMSF from both simulations have almost similar pattern. We also monitored the number of contacts in the contact regions as shown in Figure A.6. Even though the number of contacts decreases, it becomes stable after around 5 ns simulation time.

These results indicate that the tetramer azurin is relatively stable. Nevertheless the final conformation is less packed than the native conformation. It is well confirmed by the calculation of surface area shown in Figure A.7. The surface area of tetramer azurin gradually increases during simulation time. Meanwhile RMSD and number of contacts imply that our system becomes relatively stable, and the increase of surface area indicate that our system is starting to separate. We suggest that longer simulation time is surely needed to investigate more deeply the stability of azurin complex.
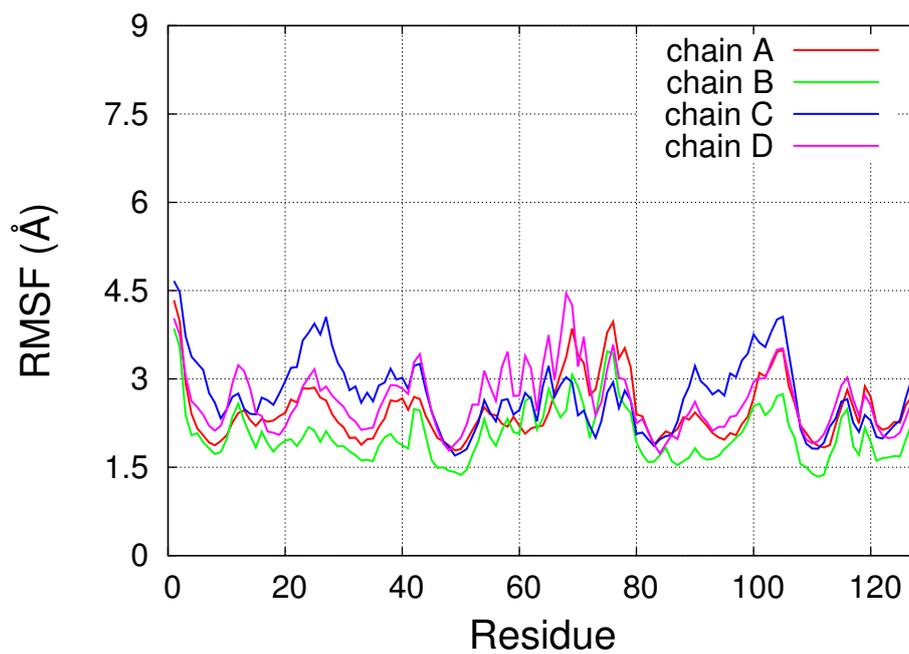
Figure A.4: RMSF profile for each chain.
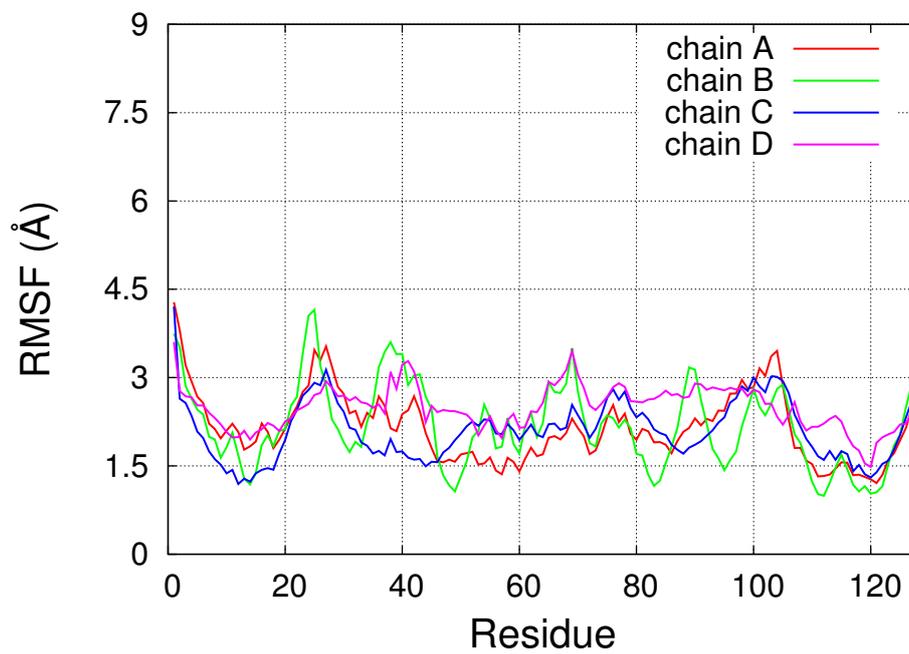
Figure A.5: RMSF profile for each chain from all atom simulation.
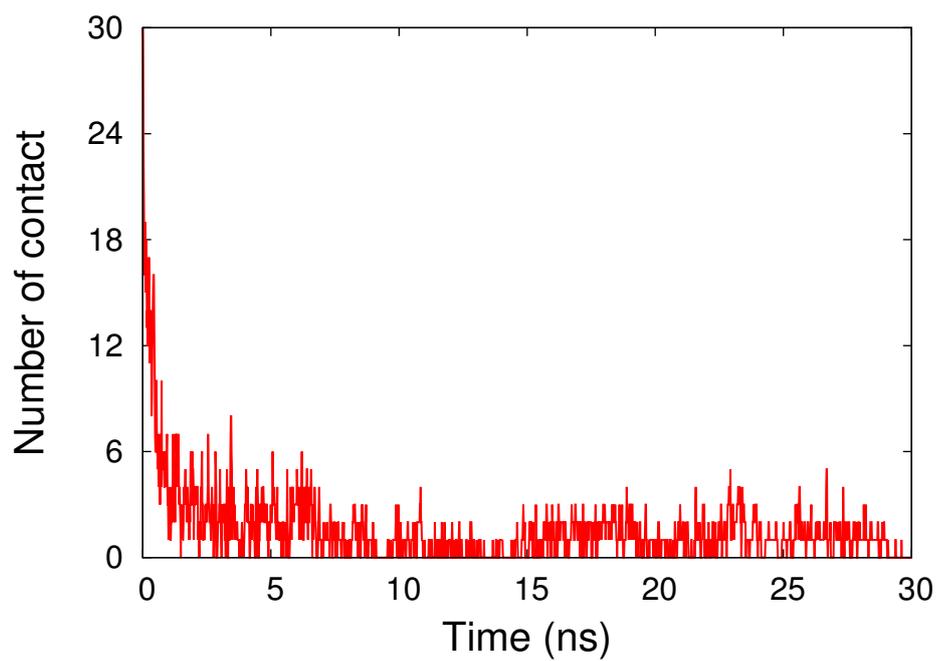
Figure A.6: Number of intermolecular contact.
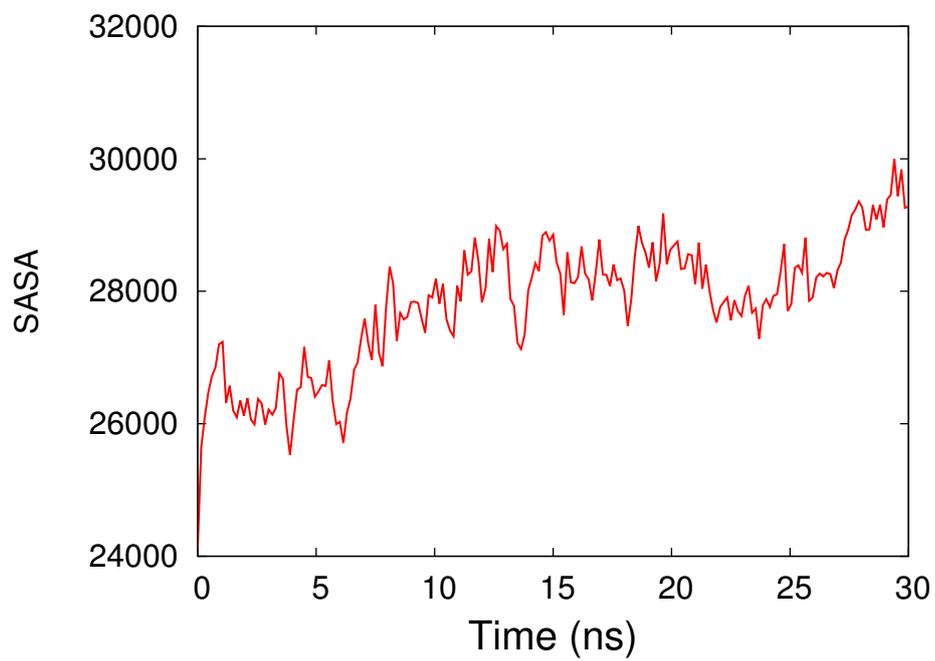
Figure A.7: Total surface area

A.5 Conclusion

In this work we offer new scheme to approach coarse-grained model of azurin complex. We modified the best known Lennard-Jones 6-12 potential by introducing new parameter for the attractive term. This parameter is determined by the native structure, yet is simplified to be more transferable to the unknown structure. Other parameters are determined by adapting Miyazawa-Jernigan contact potential and Kim-Hummer scaling.

However, it is still too premature to say that our potential is suitable to model the azurin complex. The structural properties show that our system is relatively stable during simulation time. But in contrast, the surface area of our system gradually increases indicating that the system is starting to separate.

Therefore, we need to simulate for a longer time to observe the stability and dynamics of azurin complex more deeply. Other approaches may be needed, such as force matching approach to determine parameter $a_{ij}^{\alpha\beta}$ or iterative Boltzmann inversion to determine σ. Further studies on this field will provide important contribution to advance nanoscience, particularly to biomolecular modeling.

Appendix B

Force field

Force, F, can be defined as the negative of the derivative of the potential function U. This negative sign shows that if the potential U increases along with increasing of the distance (r), the force will tend to move the particles to become closer to decrease the potential energy. Since we work on the three dimensional structures, the force will also be a three-dimensional vector.

$$\vec{F} = -\nabla U(r) \tag{B.1}$$

In the following sections, we will provide the derivation of each potential that we used in this thesis.

B.1   Gō potential

Go model potential for a given configuration of a protein can be seen in Table 3.2. The force field for each interaction will be described as follows:

Virtual bond-stretching

Virtual bond-stretching is represented by a harmonic potential well.

$$U_{\text{bond}}(r) = K_{\text{bond}}(r - r_0)^2, \tag{B.2}$$

with r is length of bonds between two adjacent particles, which can be written as

$$r_i = [(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2]^{1/2},$$ (B.3)

and $\vec{r}_i$ is the vector connecting those two particles, can also be written as

$$\vec{r}_i = \begin{bmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \\ z_{i+1} - z_i \end{bmatrix}$$ (B.4)

The force with respect to x direction becomes

$$F_{\mathsf{bond}}(r)|_x = -2K_{\mathsf{bond}}(r - r_0)\frac{\partial r}{\partial x}$$ (B.5)

The derivation of the equation B.3 gives us

$$\begin{aligned} \frac{\partial r_i}{\partial x} &= \frac{1}{2}[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2]^{-1/2} \cdot 2(x_{i+1} - x_i) \cdot (-1) \\ &= -\frac{(x_{i+1} - x_i)}{r_i}. \end{aligned}$$

(B.6)

In the same way for the y and z directions, we have

$$\begin{aligned} \nabla r_i &= - \begin{bmatrix} \frac{(x_{i+1}-x_i)}{r} \\ \frac{(y_{i+1}-y_i)}{r} \\ \frac{(z_{i+1}-z_i)}{r} \end{bmatrix} \\ &= -\frac{\vec{r}_i}{r} \end{aligned}$$ (B.7)

Then,

$$\begin{aligned} \vec{F}_{\mathsf{bond}}(r_i) &= -2K_{\mathsf{bond}}(r - r_0)\nabla r_i \\ &= -2K_{\mathsf{bond}}(r - r_0)\frac{-\vec{r}_i}{r} \end{aligned}$$ (B.8)

and analogously

$$\vec{F}_{\text{bond}}(r_{i+1}) = -\vec{F}_{\text{bond}}(r_i). \tag{B.9}$$

Virtual bond-angle bending

Virtual bond-angle bending forces are three-body forces between three successive bonded particles. The associated angle potential is characterized by an angle θ between atoms $i-1$, i, and $i+1$. For the potential we can use the form of trigonometric harmonic potential as the following equation:

$$U_{\text{angle}} = K_\theta(\cos\theta - \cos\theta_0)^2. \tag{B.10}$$

The potential depends on angle determined by three successive particles. The associated angle can be computed from the cartesian coordinates using the relation

$$\theta = \arccos(\frac{-\vec{r}_{i-1} \cdot \vec{r}_i}{r_{i-1}r_i}). \tag{B.11}$$

The force can be simply determined by the relation:

$$\vec{F}_{\text{angle}}\Big|_i = -\left.\vec{F}_{\text{angle}}(r)\right|_{i-1} - \left.\vec{F}_{\text{angle}}(r)\right|_{i+1}. \tag{B.12}$$

Moreover, it holds that

$$
\begin{aligned}
\vec{F}_{\text{angle}}(r)\Big|_{i-1} &= -2K_\theta(\cos\theta_i - \cos\theta_{0i}) \cdot (-\sin\theta_i)\nabla\theta_i \\
&= -2K_\theta(\cos\theta_i - \cos\theta_{0i}) \cdot (-\sin\theta_i)\frac{\partial\theta_i}{\partial\cos\theta_i}\nabla(\cos\theta_i) \\
&= -2K_\theta(\cos\theta_i - \cos\theta_{0i}) \cdot (-\sin\theta_i)(-\frac{1}{\sin\theta_i})\nabla(\cos\theta_i) \\
&= -2K_\theta(\cos\theta_i - \cos\theta_{0i})\nabla(\cos\theta_i).
\end{aligned}
\tag{B.13}
$$

For the derivation part, we define $S = -\vec{r}_{i-1} \cdot \vec{r}_i$ and $D = r_{i-1}r_i$, so $\cos\theta_i = S/D$. For the x direction, we obtain

$$\frac{\partial\cos\theta_i}{\partial x_{i-1}} = \frac{S'}{D} - \frac{SD'}{D^2}. \tag{B.14}$$

$$S' = \frac{\partial}{\partial x_{i-1}}(-\vec{r}_{i-1} \cdot \vec{r}_i)$$

$$= x_{i+1} - x_i \text{ , and} \tag{B.15}$$

$$
\begin{aligned}
D' &= \frac{\partial}{\partial x_{i-1}}(r_{i-1}r_i) \\
&= r_i \frac{\partial}{\partial x_{i-1}}(r_{i-1}) \\
&= -\frac{(x_i - x_{i-1})}{r_{i-1}}r_i \\
&= -\frac{(x_i - x_{i-1})}{D}r_i^2.
\end{aligned} \tag{B.16}
$$

In the same way for other directions, we have

$$S' = \vec{r}_i \text{ , and} \tag{B.17}$$

$$D' = -\frac{\vec{r}_{i-1}}{D}r_i^2. \tag{B.18}$$

Then,

$$\vec{F}_{\text{angle}}\Big|_{i-1} = -2K_\theta(\cos\theta_i - \cos\theta_{0i})\frac{1}{D}\left(\vec{r}_i + \frac{S}{D^2}\vec{r}_{i-1}r_i\right). \tag{B.19}$$

$\vec{F}_{\text{angle}}\Big|_{i+1}$ can be computed in an analogous way.

$$\vec{F}_{\text{angle}}(r)\Big|_{i+1} = -2K_\theta(\cos\theta_i - \cos\theta_{0i})\frac{1}{D}\left(-\vec{r}_{i-1} + \frac{S}{D^2}\vec{r}_ir_{i-1}\right). \tag{B.20}$$

Virtual bond-torsional term

We define ϕ as the virtual bond-torsional angle which is formed by for subsequent residues $i-1$, i, $i+1$, and $i+2$.

$$U_{\text{dihedral}} = K_\phi[1 - \cos(\phi - \phi_0)] + \frac{K_\phi}{2}[1 - \cos(3 \times (\phi - \phi_0))] \tag{B.21}$$

$\vec{n}_1$ is defined as normal vector of the plane of atoms $i-1$, i, and $i+1$, and $\vec{n}_2$ is defined as normal vector of the plane of atoms i, $i+1$, and $i+2$. So

we have,

$$\vec{n}_1 = -\vec{r}_{i-1} \times \vec{r}_i \tag{B.22}$$

$$\vec{n}_2 = \vec{r}_i \times -\vec{r}_{i+1} \tag{B.23}$$

The classical definition of the dihedral angle ϕ is given by the relation

$$\phi = sign(\phi) \arccos \left(\frac{\vec{n}_1 \cdot \vec{n}_2}{|\vec{n}_1||\vec{n}_2|} \right). \tag{B.24}$$

Mostly the gradient $\nabla U_{\text{dihedral}}$ is obtained by the following chain rule factorization

$$\left. \vec{F}_{\text{dihedral}} \right|_x = -(\mathrm{d}U/\mathrm{d}\phi)(\mathrm{d}\phi/\mathrm{d}\cos\phi)\nabla(\cos\phi), \tag{B.25}$$

but it would give us singularity because it contains division by $\sin\phi$. To avoid it, we use

$$
\begin{aligned}
\left. \vec{F}_{\text{dihedral}} \right|_x &= -(\mathrm{d}U/\mathrm{d}\phi)\nabla\phi \\
&= - \left[K_\phi \sin(\phi - \phi_0)\nabla\phi + -\frac{3K_\phi}{2}\sin(3 \times (\phi - \phi_0))\nabla\phi \right]
\end{aligned}
\tag{B.26}
$$

First, we construct the expression of $\left. \vec{F}_{\text{dihedral}} \right|_{i-1}$ by considering the direction and the length of $\left. \vec{F}_{\text{dihedral}} \right|_{i-1}$. Notes that $\left. \vec{F}_{\text{dihedral}} \right|_{i-1}$ must be normal to the equipotential plane in which particle $i-1$ can move without changing ϕ. Since ϕ does not change when $i-1$ is moved in the plane $i-1$, i, $i+1$, the equipotential plane is obviously the plane $i-1$, i, $i+1$. Hence $\left. \vec{F}_{\text{dihedral}} \right|_{i-1}$ is in the direction $\vec{n}_1$. Therefore $\left. \vec{F}_{\text{dihedral}} \right|_{i-1} = \left| \left. \vec{F}_{\text{dihedral}} \right|_{i-1} \right| \hat{n}_1$. When particle $i-1$ is given small displacement Δr_i in the direction $\vec{n}_1$ then

$$
\begin{aligned}
\Delta\phi &= \frac{\Delta r_{i-1}}{\text{distance of } i-1 \text{ to line } i, i+1} \\
&= \frac{\Delta r_{i-1}}{\left| -\vec{r}_{i-1} - \langle -\vec{r}_{i-1}, \frac{\vec{r}_i}{r_i} \rangle \frac{\vec{r}_i}{r_i} \right|} \\
&= \frac{\Delta r_{i-1}}{\left| -\vec{r}_{i-1} \times \vec{r}_i \right|/r_i} \\
&= \Delta r_{i-1} \frac{r_i}{|\vec{n}_1|}
\end{aligned}
\tag{B.27}
$$

Then we obtain

$$\left| \vec{F}_{\text{dihedral}} \right|_{i-1} \Bigg| = \left| -\frac{\mathrm{d}U}{\mathrm{d}\phi} \right| \frac{\Delta\phi}{\Delta r_{i-1}}$$
$$= \left| \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} \right| \frac{r_i}{|\vec{n}_1|}, \tag{B.28}$$

and

$$\vec{F}_{\text{dihedral}} \Big|_{i-1} = -\frac{\mathrm{d}U}{\mathrm{d}\phi} r_i \frac{\vec{n}_1}{|\vec{n}_1|^2}. \tag{B.29}$$

Analogously we obtain,

$$\vec{F}_{\text{dihedral}} \Big|_{i+2} = \frac{\mathrm{d}U}{d\phi} r_i \frac{\vec{n}_2}{|\vec{n}_2|^2}. \tag{B.30}$$

We have $\vec{F}_{i-1} + \vec{F}_i + \vec{F}_{i+1} + \vec{F}_{i+2} = 0$. Therefore,

$$\vec{F}_i = -\vec{F}_{i-1} + \vec{B}, \tag{B.31}$$

$$\vec{F}_{i+1} = -\vec{F}_{i+2} - \vec{B}, \tag{B.32}$$

with $\vec{B}$ is an unknown vector and perpendicular to $\vec{r}_i$. The total torque vanishes:

$$\vec{q}_{i-1} \times \vec{F}_{i-1} + \vec{q}_i \times (-\vec{F}_i + \vec{B}) + \vec{q}_{i+1} \times (-\vec{F}_{i+2} - \vec{B}) + \vec{q}_{i+2}\vec{F}_{i+2} = 0, \tag{B.33}$$

where $\vec{q}$ is the cartesian coordinate of the particle. It implies

$$(-\vec{r}_{i-1} \times \vec{F}_{i-1} + \vec{r}_{i+1} \times \vec{F}_{i+2}) - (\vec{r}_i \times \vec{B}) = 0. \tag{B.34}$$

Defining $\vec{A} = -\vec{r}_{i-1} \times \vec{F}_{i-1} + \vec{r}_{i+1} \times \vec{F}_{i+2}$, we have $\vec{r}_i \times \vec{B} = \vec{A}$. Since $\vec{B} \perp \vec{r}_i$, so $\vec{B}$ has direction $\vec{A} \times \vec{r}_i$ and the size $|\vec{A}|/r_i$. Hence,

$$\vec{B} = \frac{\vec{A} \times \vec{r}_i}{r_i^2} = \frac{\vec{r}_i \times (\vec{r}_{i-1} \times \vec{F}_{i-1}) - r_i \times \vec{r}_{i+1} \times \vec{F}_{i+2}}{r_i^2}. \tag{B.35}$$

Using the vector identity $\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{C} \cdot \vec{A}) - \vec{C}(\vec{B} \cdot \vec{A})$ and the fact that

66

$\vec{F}_{i-1}, F_{i+2} \perp \vec{r}_i$ we find

$$\vec{B} = \frac{1}{r_i^2}(-\vec{F}_{i-1}(\vec{r}_{i-1} \cdot \vec{r}_i) + \vec{F}_{i+2}(\vec{r}_{i+1} \cdot \vec{r}_i)). \qquad (B.36)$$

So, we have

$$\vec{F}_i = -\vec{F}_{i-1} - \left(\frac{\vec{r}_{i-1} \cdot \vec{r}_i}{r_i^2}\right)\vec{F}_{i-1} + \left(\frac{\vec{r}_{i+1} \cdot \vec{r}_i}{r_i^2}\right)\vec{F}_{i+2} \qquad (B.37)$$

and

$$\vec{F}_{i+1} = -\vec{F}_{i+2} + \left(\frac{\vec{r}_{i-1} \cdot \vec{r}_i}{r_i^2}\right)\vec{F}_{i-1} - \left(\frac{\vec{r}_{i+1} \cdot \vec{r}_i}{r_i^2}\right)\vec{F}_{i+2}. \qquad (B.38)$$

Finally we consider the sign of dihedral angle ϕ. A simpler definition of $sign(\phi)$ is $sign(\phi) = signum(-\vec{r}_{i-1} \cdot \vec{n}_2)$.

Nonbonded: Native interaction

For non-bonded interaction, we define r_{ij} as the distance between non-bonded particles. In Gō model, σ_{ij} is represented by r_{ij} in the native structure.

$$U_{\text{nat}}(r_{ij}) = \varepsilon_{\text{nat}}\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] \qquad (B.39)$$

$$\begin{aligned}
\vec{F}_{\text{nat}}(r_{ij})\Big|_i &= -\varepsilon_{\text{nat}}\left[60\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{11} - 60\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9}\right]\nabla_i\left(\frac{\sigma_{ij}}{r_{ij}}\right) \\
&= -60\varepsilon_{\text{nat}}\sigma_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{11} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9}\right]\nabla_i\left(\frac{1}{r_{ij}}\right) \\
&= -60\varepsilon_{\text{nat}}\sigma_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{11} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9}\right]\frac{\vec{r}_{ij}}{r_{ij}^3} \\
&= -60\varepsilon_{\text{nat}}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right]\frac{\vec{r}_{ij}}{r_{ij}^2}.
\end{aligned} \qquad (B.40)$$

Nonbonded: Non-native interaction

Similarly to the native interaction, the force can be derived as follow:

$$U_{\text{non-nat}}(r_{ij}) = \varepsilon_{\text{non-nat}} \left(\frac{C}{r_{ij}}\right)^{12} \tag{B.41}$$

$$
\begin{aligned}
\vec{F}_{\text{non-nat}}\Big|_i &= -12\varepsilon_{\text{non-nat}} \left(\frac{C}{r_{ij}}\right)^{11} \nabla_i \left(\frac{C}{r_{ij}}\right) \\
&= -12\varepsilon_{\text{non-nat}}C \left(\frac{C}{r_{ij}}\right)^{11} \nabla_i \left(\frac{1}{r_{ij}}\right) \\
&= -12\varepsilon_{\text{non-nat}}C \left(\frac{C}{r_{ij}}\right)^{11} \frac{\vec{r}_{ij}}{r_{ij}^3} \\
&= -12\varepsilon_{\text{non-nat}} \left(\frac{C}{r_{ij}}\right)^{12} \frac{\vec{r}_{ij}}{r_{ij}^2}
\end{aligned}
\tag{B.42}
$$

B.2 6-12 Lennard-Jones potential

For the 6-12 Lennard-Jones potential as in Equation 5.1, the force can be derived into:

$$
\begin{aligned}
\vec{F}_{LJ}\Big|_i &= -4\varepsilon_{LJ} \left[12\left(\frac{\sigma}{r_{ij}}\right)^{11} - 6\left(\frac{\sigma}{r_{ij}}\right)^{5}\right] \nabla_i \left(\frac{\sigma}{r_{ij}}\right), \\
&= -4\varepsilon_{LJ} \left[12\left(\frac{\sigma}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \sigma\frac{\vec{r}_{ij}}{r_{ij}^2}
\end{aligned}
\tag{B.43}
$$

B.3 Modified Lennard-Jones potential

For the modified Lennard-Jones potential we used in Appendix A (see Equation A.1), the force can be derived into:

$$
\begin{aligned}
\vec{F}_{ij}^{\alpha\beta}\Big|_i &= -4\varepsilon_{\text{inter}} \left[12\left(\frac{\sigma}{r_{ij}}\right)^{11} - 6a_{ij}^{\alpha\beta}\left(\frac{\sigma}{r_{ij}}\right)^{5}\right] \nabla_i \left(\frac{\sigma}{r_{ij}}\right), \\
&= -4\varepsilon_{\text{inter}} \left[12\left(\frac{\sigma}{r_{ij}}\right)^{12} - 6a_{ij}^{\alpha\beta}\left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \sigma\frac{\vec{r}_{ij}}{r_{ij}^2}
\end{aligned}
\tag{B.44}
$$

List of publications

Papers as first author

1. **M. Rusmerryani**, M. T. Pakpahan, M. Nishimura, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Transition state analysis of azurin via Gō-like model*, AIP Conf. Proc., **1518**, 641-644 (2013).

2. **M. Rusmerryani**, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Coarse-grained simulation of azurin crystal complex system: Protein–protein interactions*, ISCS 2013 Selected Papers, **4** (2013).

3. **M. Rusmerryani**, M . Takasu, K. Kawaguchi, H. Saito, and H. Nagao. *Protein–protein interactions of azurin complex by coarse-grained simulations with a Gō-like model*, JPS Conf. Proc., **1**, 012054 (2014).

Other papers

1. **M. T. Pakpahan**, M. Rusmerryani, K. Kawaguchi, H. Saito, and H. Nagao. *Evaluation of scoring functions for protein-ligand docking*, AIP Conf. Proc., **1518**, 645-648 (2013).

2. **H. Nagao**, S. Kawamoto, M. Rusmerryani, A. Purqon, K. Kawaguchi, and H. Saito. *Molecular dynamics study on entrainment phenomenon in model molecular systems*, AIP Conf. Proc., **1518**, 729-732 (2013).

References

[1] Feynman, R.P. 1992. "There's plenty of room at the bottom". *Journal of Microelectromechanical Systems* **1**:60-66. A reprint of the talk.

[2] Ambler, R. P. and Brown, L. H. 1967. The amino acid sequence of Pseudomonas fluorescens azurin. *Biochemical Journal* **104**(3):784-825.

[3] Norris, G. E. 1982. *The Three Dimensional Structure of Azurin, A Blue Copper Protein, At 3Å Resolution*. Massey University, New Zealand.

[4] Fuentes, L., et. al. 2004. Conformational Changes in Azurin from Pseudomonas aeruginosa Induced through Chemical and Physical Protocols. *Biophysical Journal* **87**(3):1873-1880.

[5] Antonini, E., et. al. 1970. Kinetics of electron transfer between azurin and cytochrome 551 from Pseudomonas. *The Journal of Biological Chemistry* **245**:4847-4856.

[6] Brunori, M., Parr, SR., Greenwood, C., Wilson, MT. 1975. A temperature-jump study of the reaction between azurin and cytochrome c oxidase from Pseudomonas aeruginosa. *Biochemical Journal* **151**(1):185-188.

[7] Keyhanian, K., Mansoori, G. A., and Rahimpour, M. 2010. Prospects for Cancer Nanotechnology Treatment by Azurin. *Dynamic Biochemistry, Process Biotechnology and Molecular Biology* **4**(1):48-66.

[8] Punj, V., Das Gupta, TK., and Chakrabarty, AM. 2003. Bacterial cupre-doxin azurin and its interactions with the tumor suppressor protein p53. *Biochemical and Biophysical Research Communications* **312**(1):109-114.

[9] Yamada, T., et. al. 2002. The bacterial redox protein azurin induces apoptosis in J774 macrophages through complex formation and stabilization of the tumor suppressor protein p53. *Infection and Immunity* **70**(12):7054-7062.

[10] Punj, V., et. al. 2004. Bacterial cupredoxin azurin as an inducer of apoptosis and regression in human breast cancer. *Oncogene* **23**(13):2367-2378.

[11] Mahfouz, M., Hashimoto, W., Das Gupta, TK., and Chakrabarty, AM. 2007. Bacterial proteins and CpG-rich extrachromosomal DNA in potential cancer therapy. *Plasmid* **57**(1):4-17.

[12] Yang, DS., et. al. 2005. Bacterial redox protein azurin induce apoptosis in human osteosarcoma U2OS cells. *Pharmacological Research* **52**(5):413-421.

[13] Chaudhari, A., et. al. 2007. Cupredoxin-cancer interrelationship: azurin binding with EphB2, interference in EphB2 tyrosine phosphorylation, and inhibition of cancer growth. *Biochemistry*, **46**(7):1799-1810.

[14] Kwan, JM., et. al. 2009. Bacterial proteins as potential drugs in the treatment of leukemia. *Leukemia Research*, **33**(10):1392-1399.

[15] Nar, H., et. al. 1991. Crystal structure analysis of oxidized *Pseudomonas Aeruginosa* azurin at pH 5.5 and pH 9.0. *Journal of Molecular Biology* **221**(3):765 - 772.

[16] Branden, C. I. and Tooze, J. 1999. *Introduction to protein structure.* 2nd ed. Garland Publishing, USA.

[17] Smit, B., et. al. 1990. Computer simulations of water/oil interface in the presence of micelles. *Nature* **348**:624-625.

[18] Marrink, S.J., de Vries, A.H., and Mark, A.E. 2004. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B* **108**(2):750-760.

[19] Lu, L. and Voth, GA. 2009. Systematic coarse-graining of a multicomponent lipid bilayer. *The Journal of Physical Chemistry B* **113**(5):1501-1510.

[20] Kawamoto, S., et. al. 2011. Inverted Micelle Formation on Cell-Penetrating Peptide Studied by Coarse-Grained Simulation: Importance on Attractive Force Between Cell-Penetrating Peptides and Lipid Head Group. *The Journal of Chemical Physics* **134**(9):095103.

[21] Gō , N. 1983. Theoretical studies of protein folding. *Annual Review of Biophysics and Bioengineering* **12**:183-210.

[22] Clementi, C., Nymeyer, H., and Onuchic, J.N. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* **298**(5):937-953.

[23] Miyazawa, S. and Jernigan, R. L. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of Molecular Biology* **256**:623-644.

[24] J. I. Sulkowska, M. Cieplak, *Biohys. J.* **95**, 3174-3191 (2008).

[25] H. Kenzaki, et al. (2011). CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comp.*, **7**, 1979 - 1989.

[26] W. Li, P. G. Wolynes, and S. Takada (2011). Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proceedings of the National Academy Sciences of United States of America* **108**, 3504 - 3509.

[27] Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. 2007. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **111**(27):7812-7824.

[28] Monticelli, L., et. al. 2008. The MARTINI coarse grained forcefield: extension to proteins. *Journal of Chemical Theory and Computation* **4**(5):819-834.

[29] Tirion, M. M. 1996. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis *Physical Review Letters* **77**(9):1905-1908.

[30] Bahar, I., Atilgan, A. R., and Erman, B. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* **2**(3):173-181.

[31] Kim, Y. C. and Hummer, G. 2008. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *Journal of Molecular Biology* **375**(5),1416-1433.

[32] Miyazawa, S. and Jernigan, R. L. 1999. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins* **34**(1):49-68.

[33] Reith, D., Putz, M., and Muller-Plathe, F. 2003. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24**(13):1624-1636.

[34] Lyubartsev, A. P. and Laaksonen, A. 1995. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52**(4):3730-3737.

[35] Ercolessi, F. and Adams, J. B. 1994. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhysisics Letters* **26**:583.

[36] Izvekov, S. and Voth, G. A. 2005. A Multiscale Coarse-Graining Method for Biomolecular Systems. *The Journal of Physical Chemistry B* **109**(7):2469-2473.

[37] Anfinsen, C. B. 1973. Principles that Govern the Folding of Protein Chains. *Science* **181**:223-230.

[38] Onuchic, J. N., Schulten, Z. L., and Wolynes, P.G. 1977. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry* **48**:545-600.

[39] Bryngelson, J. D. and Wolynes, P. G. 1987. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy Sciences of United States of America* **84**:7524-7528.

[40] Leopold, P. E., Montal, M., and Onuchic, J. N. 1992. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy Sciences of United States of America* **89**:8721-8725.

[41] Matouschek, A., Kellis, JT. Jr, Serrano, L., and Fersht, AR. 1989. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**(6229):122-126.

[42] Onuchic, J.N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. 1996. Protein folding funnels: the nature of the transition state ensemble. *Folding and Design* **1**(6):441-450.

[43] Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. 2002. Studies of Pseudomonas aeruginosa azurin mutants: cavities in beta-barrel do not affect refolding speed. *Biophysical Journal* **82**(5):2645-2651.

[44] Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. 2001. Probing copper ligands in denatured *Pseudomonas aeruginosa* azurin: unfolding His117Gly and His46Gly mutants. *Journal of Biological Inorganic Chemistry* **6**:182-188.

[45] Alagaratnam, S., et. al. 2011. Probing the reactivity of different forms of azurin by flavin photoreduction. *FEBS Journal* **278**:1506-1521.

[46] Cheluvaraja, S. and Meirovitch, H. 2006. Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method. *Journal of Chemical Physics*, **125**:024905.

[47] Fersht, A.R. 1995. Characterizing transition states in protein folding: an essential step in the puzzle. *Current Opinion in Structural Biology* **5**(1):79-84.

[48] Fersht, A.R and Sato, S. 2004. Φ-Value analysis and the nature of protein-folding transition states. *Proceedings of the National Academy Sciences of United States of America* **101**(21):7976-7981.

[49] Koga, N. and Takada, S. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *Journal of Molecular Biology* **313**(1):171-180.

[50] Plaxco, K., Simons, K. T., and Baker, D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* **277**(4):985-994.

[51] Zong, C., Wilson, C. J., Shen, T., Wolynes, P. G., and Wittung-Stafshede, P. 2006. Φ-Value Analysis of Apo-Azurin Folding: Comparison between Experiment and Theory. *Biochemistry* **45**(20):6458-6466.

[52] Chen, M., Wilson, C.J., Wu, Y., Wittung-Stafshede, P., and Ma, J. 2006. Correlation between Protein Stability Cores and Protein Folding Kinetics: A Case Study on Pseudomonas aeruginosa Apo-Azurin. *Structure* **14**:1-10.

[53] Minton, A. P. 2006. How can biochemical reactions within cells differ from those in test tubes? *Journal of Cell Science* **119**:2863-2869.

[54] Homouz, D., Stagg, L., Wittung-Stafshede, P., and Cheng, M. S. 2009. Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. *Biophysical Journal* **96**:671-680.

[55] Pincus, D. L. and Thirumalai, D. 2009. Crowding effects on the mechanical stability and unfolding pathways of ubiquitin. *The Journal of Physical Chemistry B* **113**:359-368.

[56] Predeus, A. V., et. al. 2012. Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. *The Journal of Physical Chemistry B* **116**:8610-8620.

[57] Fulton, A. B. 1982. How crowded is the cytoplasm? *Cell* **30**:345-347.

[58] Maldonado, G. A. M., et al. 2012. On the centre of mass velocity in molecular dynamics simulations. *Revista Mexicana de Fisica* **58**:55-60.

[59] Guenza, M. 2002. Intermolecular effects in the center-of-mass dynamics of unentangled polymer fluids. *Macromolecules* **35**:2714-2722.

[60] Gummadi, S. N. 2003. What is the role of thermodynamics in protein stability? *Biotechnology and Bioprocess Engineering* **8**:9-18.

[61] O. Keskin, et. al. (2005). Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. *Phys. Biol.*, **2**, S24-S35.

[62] Jones, J. E. 1924. On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature. *Proceedings of The Royal Society A* **106**(738):441-462.

[63] Israelachvili, J. N. 2011. *Intermolecular and Surface Forces*. 3rd ed. Elsevier, California.

[64] Lee, B. and Richards, F. M. 1971. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology* **55**:379-400.

[65] Wodak, S. J. and Janin, J. 1980. Analytical approximation to the accessible surface area of proteins. *Proceedings of the National Academy Sciences of United States of America* **77**(4):1736-1740.

[66] Cavallo, L., Kleinjung, J. and Fraternali, F. 2003. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research* **31**:3364-3366.

[67] Elcock, A.H., Sept, D. and McCammon, J. A. 2001. Computer Simulation of Protein-Protein Interactions. *The Journal of Physical Chemistry B* **105**(8):1504-1518.

[68] Creighton, TE. 1993. *Proteins: structures and molecular properties.* 2nd Ed. W.H. Freeman and Company, New York.

[69] Phillips, J. C., et. al. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**:1781-1802.

[70] MacKerell, Jr., et. al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**:3586-3616.

[71] Bekker, H. 1996. *Molecular dynamics simulation methods revised*. Proefschrift, Groningen.

[72] Griebel, M., et al. 2007. *Numerical simulation in molecular dynamics*. Springer, Germany.