Stony Brook University

Molecular Dynamics Modeling of a Biotoxin in Various Temperature and pH Environments and Another Protein Interacting with Small Molecules

A Dissertation Presented

By

Xin Chen

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Xin Chen

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Professor Yuefan Deng – Dissertation Advisor
Department of Applied Mathematics and Statistics

Professor Robert Shrock – Chairman of Defense
Department of Physics and Astronomy

Professor Peter Paul
Department of Physics and Astronomy

Professor Thomas T.S. Kuo
Department of Physics and Astronomy

Dr. James W. Davenport – Outside Member
Computational Science Center
Brookhaven National Laboratory

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstraction of the Dissertation

Molecular Dynamics Modeling of a Biotoxin in Various Temperature and pH Environments and Another Protein Interacting with Small Molecules

by

Xin Chen

Doctor of Philosophy

in

Physics

Stony Brook University

2007

There has been a long history to use classical molecular dynamics in investigating the properties of biomolecular systems. As a supplement to experimental methods, it has provided invaluable insight into the properties and mechanism of proteins and other macro-molecular systems which are hard or impossible to investigate by regular in vitro experiments. In the thesis, we used classical molecular dynamics to study botulinum and PDZ domains (a common domain shared by post synaptic density protein, Drosophila disc large tumor suppressor, and zonula occuldens-1 protein) in detail. Botulinum neurotoxins type A (BoNT/A) are highly potent toxins, but are also useful in the treatment of illnesses. We studied the properties of BoNT/A at various temperatures and pH

values in order to understand its toxicity and structure variations. The pH values of the environment of BoNT/A were modeled by changing the protonation states of certain titratable residue groups. Our results show that certain parts of the protein are active at acidic pH environments or at high temperatures. The protein is more stable in neutral environments at normal human body temperature; whereas, at high temperature, the protein is more stable in acidic environments. The Dishevelled (Dvl) PDZ domain is believed to play an essential role in the canonical and noncanonical Wnt signaling pathways, which are involved in embryo development as well as in tumorigenesis. Also, it binds directly to Frizzled (Fz) receptors. An organic molecule (NSC668036) from the National Cancer Institute small-molecule library has been identified to be able to bind to the Dvl PDZ domain. Molecular dynamic simulation was used to analyze the binding between them in detail.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to thank my advisor, Professor Yuefan Deng, for his deep insights and ideas, lots of invaluable advices, many helpful discussions, endless encouragement and continuous support in my Ph. D. program. Thank you so much for introducing me to this great research area.

I also thank my co-advisor, Professor Robert Shrock, for his generous help and many priceless advices and suggestions for my thesis. Also I want to thank my Ph. D committee members, Professor Peter Paul, Professor Thomas T.S. Kuo and Professor James Davenport, for their time and valuable comments.

I will never forget all members in the molecular dynamics research group, especially, Yongzhi Chen, for his contribution in the constant pH simulations with AMBER 8, Guowen Hang, Bin Fang, Yuxiang Gao and Peter Rissland. I have enjoyed very much working with all of you.

I also thank the Institute of Scientific Computing, Nankai University and Nankai Stars team. Numerical experiments were performed in part using Nankai Stars super-computer.

Finally, I would like to express my deepest love and thanks to my wife, Yao Li, for her supports and sacrifices, and my parents who have been keeping encouraging me every moment.

This work was supported by J. Davenport in the Brookhaven National

Laboratory by a grant under contract #84054 and the Army Research Office grants #W911NF0510413.

Chapter 1 Introduction

There has been a long history[1] to use classical molecular dynamics (MD)[2, 3] in investigating the properties of biomolecular systems. As a supplement to experimental methods, it has provided invaluable insight into the properties and mechanism of proteins and other macro-molecular systems which are hard or impossible to investigate by regular in vitro experiments. However, due to the large number of the particles in the system and the complicated interactions between them, MD simulations of proteins are usually computationally intensive. For example, the folding processes of proteins are of great interest and well studied by MD simulations. Various simulations of protein folding were performed[4-6] and the time scales needed for simulations are generally based on their molecular weights. For small proteins with about 100 residues, a simulation period of about 10 μs is needed. With a regular time step of 1 fs, i.e. computing the position and speed of every particle every 10^{-15}s, 10^{10} time steps are needed for this simulation.

The simulation parameters of a typical MD run with physical time 10^{-4}s for a small protein are listed in Table 1-1.

Physical time for simulation	10^{-4} s
Typical time-step size	10^{-15} s
Number of MD time steps	10^{11}
Atoms in a typical protein and water simulation	32000
Approximate number of interactions in force calculation	10^{9}
Machine instructions per force calculation	1000
Total number of machine instructions	10^{23}

Table 1-1 Simulation parameters of a typical MD run

In this thesis, MD was used to simulate protein botulinum and PDZ domain. Botulinum neurotoxins (BoNTs) are among the most potent toxins to human beings. But their properties in an acidic environment are still unclear at this time. A series of experiments in 1-unit pH increment from 3.0 to 8.0 were implemented expecting to get the conformational information in each case. And there was an interesting observation that the spectra at pH 3.0, 6.0, 7.0 and 8.0 were identical[7]. Unfortunately, this method did not work at pH 4.0 and 5.0, because the protein was insoluble under these conditions. It is reasonable that the protein would become more insoluble as it rearranges itself to interact with

the membrane, but the problem is that the interesting conformational change just happens in this pH 4.0 -5.0 range. For this reason, the property of BoNT/A in this range has not been addressed seriously so far and a constant pH MD simulation is motivated.

The dishevelled (Dvl) PDZ domain (a common domain shared by post synaptic density protein, Drosophila disc large tumor suppressor, and zonula occuldens-1 protein) is believed to play an essential role in the canonical and noncanonical Wnt signaling pathways, which describes a complex network of proteins most well known for their roles in embryogenesis and cancer. The name Wnt came from a combination of Wg (wingless) and Int[8] gene. They are involved in embryo development as well as in tumorigenesis. An organic molecule (NSC668036) from the National Cancer Institute small-molecule library has been identified to be able to bind to the Dvl PDZ domain. MD simulation was used to analyze the binding between them in detail.

In this dissertation, the properties of the BoNT/A and the PDZ domain are studied in detail. In Chapter 2, we give a brief discussion for the MD basics. The principles of MD are introduced, and different force fields are discussed. In Chapter 3, the modeling method of proteins is discussed in general. The

common structure of proteins is introduced, and also the different methods used to model proteins computationally. Chapter 4 introduces the platforms we used in our computational simulations of proteins, including the software package and hardware platform.

Chapter 5 gives a brief introduction of botulinum, the proteins we studied in detail. In Chapter 6, the simulations of BoNT/A are discussed in detail[9, 10]. Simulation results are shown and analyzed in detail for different pH values at various temperatures.

Chapter 7 discusses another simulation of protein: the simulation of PDZ domain. Simulation procedures are introduced and results and analysis are shown in detail. Chapter 8 outlines the main contributions of the present thesis work and future outlook of the thesis topic.

Chapter 2 Molecular Dynamics

2.1 Introduction

MD is widely used by physicists, chemists and biologists in different fields. It is a kind of computational simulation where atoms and molecules of the systems are allowed to interact for a period of time under the laws of physics. It was firstly used in theoretical physics, then in the materials sciences. Since the 1970s, it has also been used in biochemistry and biophysics in the study of properties of various biomolecular systems.

First reported work of MD simulation for macromolecule was published in 1977 in Nature magazine[1]. The protein they simulated, Bovine Pancreatic Trypsine Inhibitor, is one of the best studied proteins in terms of folding and kinetics. The simulation time was 9.2 ps with a size of 500 atoms.

In the following years, with the fast development of high performance computing, the computing power we can acquire has increased dramatically. With parallel computing, MD simulations can be performed using many processors in parallel with continuous communication between them. A recent simulation on satellite tobacco mosaic virus (STMV) [11] would take a single

desktop computer around 35 years to complete. The simulation time was 50 ns with a size of 1 million atoms using NAMD(NAnoscale Molecular Dynamics)[12, 13], a software developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. As an example of a small protein whose folding properties have been studied by means of NAMD, Villin Headpiece[14] was simulated. It contained 20,000 atoms and had a simulation time of 500 μs. The simulation was run in about 200,000 CPUs of the participating personal computers of folding@home project[15, 16] around the world.

In the classical MD, classical dynamics equation is used, while in the quantum MD, Schrödinger equation is used instead to take the quantum character of particles into consideration. Quantum MD is widely used to understand the properties of ions and other sub-atom particles[17, 18]. In some cases, quantum-classical MD (QCMD) is also used. In QCMD, quantum mechanics is used to generate parameters, and these parameters are used in classical MD simulations. It is used as an approximation to full quantum MD[19].

The biomolecular system we studied contains a large number of particles and is characterized by a large number of degrees of freedom. From first principles, it is believed that a reliable prediction of biomolecular processes should be fully based on a quantum dynamical description of all these degrees of freedom. However, even with the biggest and fastest computers, now and for the near future, we can predict that it is impossible for large system simulation to be based on such quantum model. Therefore, our simulations of biomolecular systems are based on classical MD assuming that the system obeys a classical Hamiltonian dynamics equation.

2.2 Dynamics Equation

In classical molecular dynamics, Hamilton's equations of motion are solved for a set of generalized coordinates q_i and momenta p_i

$$\dot{\vec{p}}_i = -\frac{\partial H}{\partial \vec{q}_i} \tag{2.1}$$

$$\dot{\vec{q}}_i = \frac{\partial H}{\partial \vec{p}_i} \tag{2.2}$$

where H is the Hamiltonian, or Hamiltonian function

$$H\left(\{\vec{p}_i, \vec{q}_i\}\right) = \sum_i |\vec{p}_i|^2 / 2m_i + E\left(\{\vec{q}_i\}\right) \tag{2.3}$$

where m_i is the mass of the ith particle and E represents the potential energy of the whole system.

In the cases we studied and in many other cases, the potential energy is conservative and the Cartesian coordinates $\vec{x}_i$ and velocities $\vec{v}_i$ are used.

With $\vec{p}_i = m_i \vec{v}_i$ and (2.1) we have

$$\begin{aligned} m_i \dot{\vec{v}}_i &= -\frac{\partial H}{\partial \vec{q}_i} \\ &= -\frac{\partial}{\partial \vec{x}_i} E\left(\{\vec{x}_i\}\right) \end{aligned} \tag{2.4}$$

Let $\vec{F}_i = -\frac{\partial}{\partial \vec{x}_i} E\left(\{\vec{x}_i\}\right)$, we now obtain Newton's equation of motion

$$m_i \dot{\vec{v}}_i = \vec{F}_i \tag{2.5}$$

Here $\vec{F}_i$ is a function of coordinates only in the conservative force field.

Equation (2.5) can also be written as:

$$m_i \ddot{\vec{x}}_i = \vec{F}_i \tag{2.6}$$

which is a set of second-order differential equations.

The approximate solutions of the equations can be written as:

$$\vec{x}_i(t+\Delta t) = 2\vec{x}_i(t) - \vec{x}_i(t-\Delta t) + \vec{F}_i / m_i \Delta t^2 \tag{2.7}$$

Thus, with the initial coordinates and velocities information of the interacting particles, the trajectories of the system can be calculated through the

approximate solution of the Hamiltonian equations with an empirical potential function or force field, which we will discuss in detail next.

2.3 Force Field and Computation

The basic functional form of a force field includes both bonded terms relating to atoms that are linked by covalent bonds, and nonbonded (also called "noncovalent") terms describing long-range electrostatic and van der Waals forces. For the bonded terms of the force field, they consist of a summation of bonded forces associated with chemical bonds, bond angles, and bond dihedrals. So the total potential can be written as:

$$\begin{aligned} E_{total} &= E_{bonded} + E_{nonbonded} \\ &= E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{vanderWaals} \end{aligned} \tag{2.8}$$

The first three terms in (2.8) describe the stretching, bending, and torsional bonded interactions,

$$E_{bond} = \sum_{bonds} k_i^{bond} (r_i - r_{oi})^2 \tag{2.9}$$

$$E_{angle} = \sum_{angles} k_i^{angle} (\theta_i - \theta_{0i})^2 \tag{2.10}$$

$$E_{dihedral} = \sum_{dihedral} k_i^{dihedral} [1 + \cos(n_i \phi_i - \gamma_i)] \tag{2.11}$$

where *bonds* account for each covalent bond in the system, which is a harmonic potential between bonded atoms when the bond length r_i deviates

from the equilibrium value r_{0i}, *angles* are the angles between each pair of covalent bonds sharing a single atom at the vertex, which is a harmonic potential in the valence angles of the molecules, and *dihedrals* describe atom pairs separated by exactly three covalent bonds with the central bond subject to the torsion angle ϕ, which produces a torsion potential describing the periodic variation in energy due to bond rotations. They are illustrated in Figure 2-1.

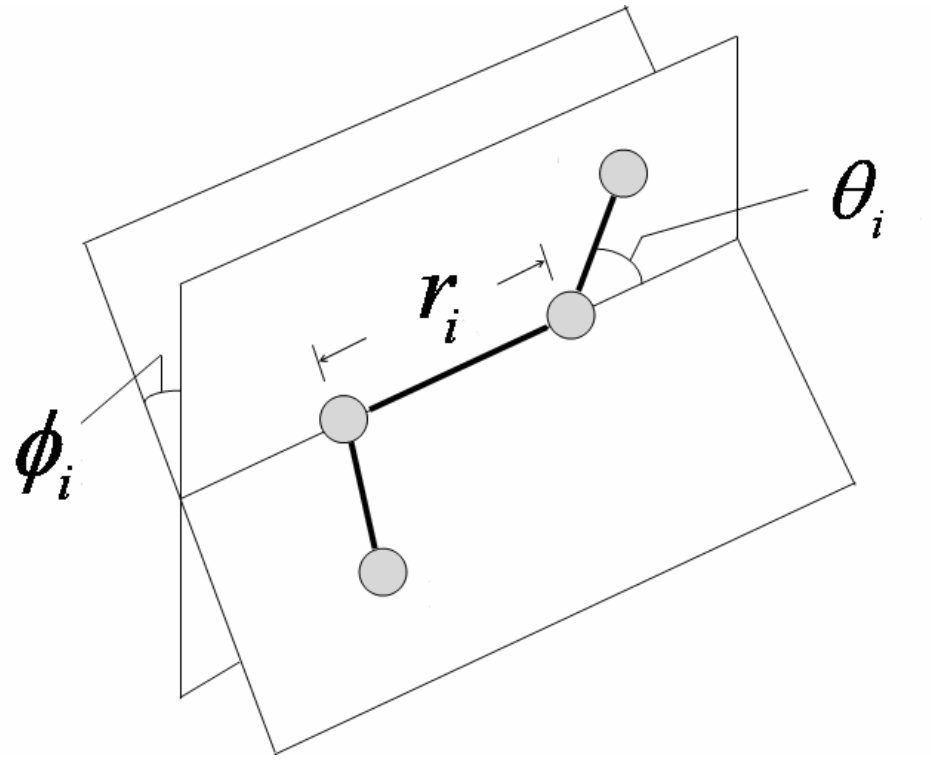

Figure 2-1 Geometry of a simple chain molecule, illustrating the bond length r_i, bend angle θ_i and torsion angle ϕ_i

The final two terms in (2.8) describe the interactions between nonbonded atom pairs,

$$E_{electrostatic} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.12}$$

$$E_{vanderWaals} = \sum_i \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.13}$$

which correspond to the electrostatic interactions (Coulomb's potential) and van der Waal's forces (approximated by a Lennard-Jones 6-12 potential[20]), respectively.

All the parameters are determined through a combination of empirical techniques and quantum mechanical calculations. They are based on numerous approximations and are derived from different types of experimental data. Besides, a force field also defines a set of parameters for each type of atom, like atomic mass, van der Waals radius, charge, the equilibrium values for bond lengths, bond angles, and dihedral angles for bonded atoms, etc. Popular classical force fields developed primarily for molecular dynamics of macromolecules include AMBER(Assisted Model Building with Energy Refinement)[21], CHARMM(Chemistry at HARvard Molecular Mechanics)[22, 23], GROMOS(GROningen MOlecular Simulation)[24],

GROMACS(Groningen Machine for Chemical Simulations)[25], etc.

Force calculation is the heart of MD simulations. To avoid the surface effects at the boundary of the simulated system, periodic boundary conditions are often used in MD simulations. In periodic boundary conditions, all the particles are enclosed in a cell that is replicated to infinity by periodic translations. Any particle that leaves the cell on one side is replaced by a copy of it entering the cell on the opposite side, and each particle is subject to the potential from all other particles including images in the surrounding cells. However, because the van der Waals and electrostatic interactions are non-local forces, they exist between every nonbonded pair of the atoms in the system including those in the neighboring cells. These nonbonded interactions are nonlocal and involve at least weak interactions between every pair of particles in the system, and are normally the bottleneck in the speed of MD simulations[26]. This actually makes computing the long-range interactions exactly unfeasible. To solve this problem, for van der Waals interactions, it is spatially truncated at a specified cutoff distance. For electrostatic interactions, Particle Mesh Ewald (PME) summation[27, 28], or the newer Particle-Particle Particle Mesh Ewald (P3ME)[29] method is used.

In Ewald summation, the summation of interacting energies in real space is replaced by an equivalent summation in Fourier space. It scales poorly with increasing number of atoms. To make Ewald summation practical in actual simulation, mesh based Ewald summation methods have been investigated. All these methods employ an interpolation scheme to generate an approximation to the atomic charge density that can be calculated in the complexity of $O(N \log N)$ at constant cutoff by using a three dimensional real-to-complex fast Fourier transform (3D-FFT).

Chapter 3 Modeling of Proteins in General

3.1 Protein Structure

Proteins are a class of biomolecules made of amino acids arranged in a linear chain and joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. The sequence of amino acids in a protein is defined by a gene and encoded in the genetic code.

There are 20 different standard amino acids used by cells in protein biosynthesis. All amino acids share a common structure including an α carbon bonded with an amino group, a carboxyl group and a variable side group. Different types of side groups define different types of amino acids. The side groups have different chemical properties that produce protein's three-dimensional structure and are therefore critical to protein function.

The details of the 20 standard amino acids are listed in Table 3-1.

Amino Acids	Abbreviation	Side Chains

Alanine	A	ALA	nonpolar
Cysteine	C	CYS	nonpolar
Aspartic acid	D	ASP	negative
Glutamic acid	E	GLU	negative
Phenylalanine	F	PHE	nonpolar
Glycine	G	GLY	nonpolar
Histidine	H	HIS	positive
Isoleucine	I	ILE	nonpolar
Lysine	K	LYS	positive
Leucine	L	LEU	nonpolar
Methionine	M	MET	nonpolar
Asparagine	N	ASN	uncharged polar
Proline	P	PRO	nonpolar
Glutamine	Q	GLN	uncharged polar
Arginine	R	ARG	positive
Serine	S	SER	uncharged polar
Threonine	T	THR	uncharged polar
Valine	V	VAL	nonpolar

Tryptophan	W	TRP	nonpolar
Tyrosine	Y	TYR	uncharged polar

Table 3-1 List of standard amino acids

Two amino acids can be combined in a condensation reaction. By repeating this reaction, long chains of residues (amino acids in a peptide bond) can be generated. A peptide bond is a chemical bond formed between two molecules when the carboxyl group of one molecule reacts with the amino group of the other molecule, releasing a molecule of water (H_2O).

Each protein has a unique sequence of amino acid residues as its primary structure. The secondary structure of proteins refers to the highly regular sub-structure, including alpha helix and strands of beta sheet. Secondary structures are locally defined and stabilized by hydrogen bonds, meaning that there can be many different secondary motifs present in one protein molecule. The tertiary structure of a protein is its overall shape, also known as its fold. It is a spatial arrangement of the secondary structures. Many proteins are actually assemblies of more than one polypeptide chain, which in the context of the larger assemblage are known as protein subunits. The quaternary structure is the arrangement of multiple folded protein molecules in a multi-subunit complex.

Most proteins fold into a unique 3-dimensional structure. The shape into which a protein naturally folds is known as its native state. The process by which the higher structures form is often called protein folding. A protein may have more than one stable folded conformation, but usually only one conformation is considered to be the active one. The time scale of the folding process varies dramatically depending on the protein. For proteins with lengths of a hundred or so amino acids, they typically fold on time scales of milliseconds[30-32]. The fastest known protein folding reactions are completed within a few microseconds[33]. In addition to the size of proteins, the folding and unfolding rates also depend on environmental conditions like temperature, solvent viscosity, pH value and more.

3.2 Protein Modeling

In order to study the properties of proteins numerically, we must set up models that reveal the interactions within the atoms and between the atoms and the environment. The most ideal solution is to solve the many-body Schrödinger equations for the potential function of the system. However, it is practically impossible to do so for relatively large systems and is only used to study

interactions with a very short time frame. Hence, several simplified models are employed in the studies.

One of the simplest model of proteins is the hydrophobic-polar (HP) model[34-36], which was first proposed by Dill in 1985. In this model, only two kinds of amino acids are considered, which are either hydrophobic or polar (hydrophilic) monomers, and they are labeled as H or P. The whole amino acid sequence of a protein is modeled as a binary sequence connected by a string. The residues can only occupy the vertices of a two or three dimensional square lattice. One vertex allows the occupation of one residue or none at all; and the adjacent amino acids in a real protein will occupy adjacent vertices too.

In spite of the simplicity of the HP model, there have been many successful applications[37] of the model predicting the native conformations of proteins without using much computing resources.

Other than the HP model, there are some coarse-grained beaded-string models in which proteins are represented as a chain of one-bead amino acids. There are also two-bead model in which the C_α atom is conveniently chosen as one bead and the side chain group is represented by another bead.

A more detailed model is the united atom model. This model is used in

simulations in vacuo or in implicit solvent[38-40]. As the name implies, this model treats a group of atoms in the protein as one particle. Implicit solvent means that the water around the protein is simulated as a continuous media instead of discrete molecules. As these kinds of models are more precise, they require more computing resources. One of the widely used implicit-solvent model is the Generalized Born(GB) model[41]. It's used in the simulations of different proteins [42, 43]. The GB model was originally developed for small compounds where it was found to work quite well; however, its performance on larger molecules was worse than expectations[44, 45]. And modifications had been proposed that improved the performance of GB model in various ways[44, 46, 47].

The model we used is the all-atom model[48] which is more precise than the united atom model. In this model, the water molecules are simulated in the atom level as explicit solvent. There are many models in simulating liquid water[49, 50], including TIPS[51], TIP3P[50], TIP4P, etc. The water model we used is TIP3P, which uses atom-centered point charges to represent the electrostatic interactions. It has three interaction sites, corresponding to the three atoms of the water molecule.

As most of the computing power is used to calculate the interactions between the water molecules, this model requires a much larger computing power.

Chapter 4 Computational Simulation of Proteins

4.1 Hardware Platform

Our simulation was performed on the Beowulf cluster NankaiStars[52] at Nankai University in China with 800 Intel Xeon processors running at 3.06 GHz. It was ranked 42nd in the Top500 list as of June 2004 after it was built, which has a real performance of 3.23 Tflops in sustained High Performance Linpack (HPL) test with overall peak performance of 4.7 Tflops.

The system couples 800 Pentium 4 Xeon processors and a great 54 TB amount of disk drive, hosted in 19 standard rack mount cabinets. There are 384 IBM xSeries 335 compute nodes with 2 GB of DDR266 memory, internal 40GB IDE disk and dual Intel Pentium 4 Xeon processors clocked at 3.06 GHz, located in the 12 computing cabinets.

All supporting services are centralized in four service cabinets including storage servers, user access servers, Gigabit Ethernet cluster local area backbone, remote control Ethernet network and RAID storage. All the computing nodes are

interconnected by a Myrinet-2000 switch and adapters to ensure a full bisection network bandwidth and a point-to-point, unidirectional (bidirectional) bandwidth of 250MB/sec (500MB/sec).

A typical run of our protein with a typical configuration takes three months to achieve 100 ns physical time at a rate of about 1 ns/day on a 128-nodes portion of the machine. Table 4-1 shows the simulation running speed on the Nankai Stars.

Number of CPUs	Running Speed (ns/day)	Speedup
32	0.27	32
64	0.63	74
128	1.16	137
256	2.01	238
512	2.95	350

Table 4-1 Running speed on the Nankai Stars Beowulf computer

Figure 4-1 reports the speedup of simulations of the Botulinum neurotoxin with NAMD v2.6 package using up to 512 processors on Nankai Stars. The seemingly super speedup for the cases of 64 and 128 CPUs may results from heavy virtual memory accesses for 32-CPU case. For our protein, it fits well on 64 CPUs. Of course, for systems with more CPUs (256 and 512),

communication slows the system down.

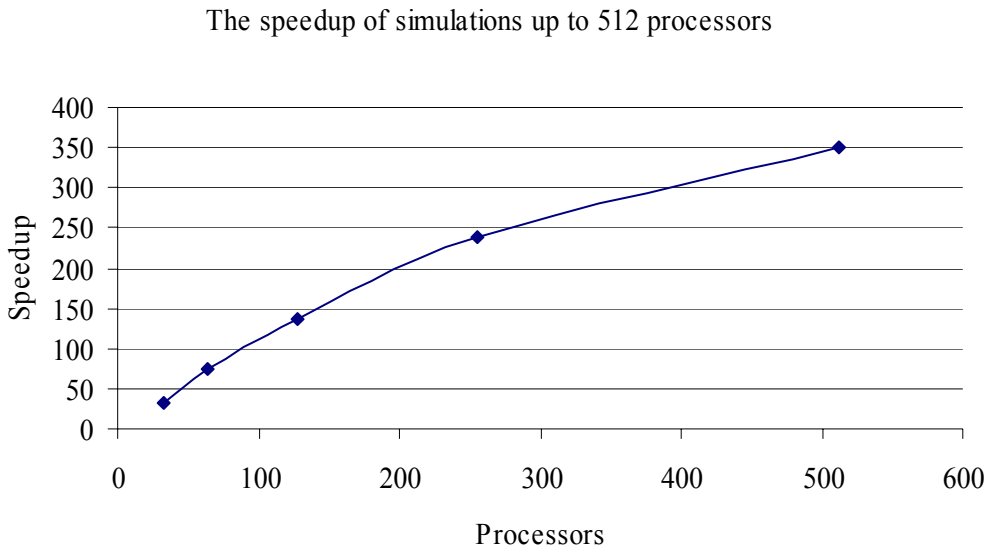

Figure 4-1 The speedup of simulations up to 512 processors

4.2 Software Package

Molecular dynamics simulations of proteins can be performed on various software platforms. The softwares we used for MD simulations include NAMD[12, 13] version 2.6 and AMBER[53, 54] version 8.

NAMD (NAnoscale Molecular Dynamics) was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at

Urbana-Champaign. It is a parallel molecular dynamics software designed for high-performance simulation of large biomolecular systems. NAMD uses the popular graphic program VMD[55] for simulation setup and trajectory analysis, but it is also file-compatible with AMBER, CHARMM, and X-PLOR.

NAMD claims to scale up to hundreds of processors on high-end parallel platforms and tens of processors on commodity clusters using gigabit Ethernet. From our experiences, it scales well up to 512 processors on the Nankai Stars clusters.

AMBER (Assisted Model Building with Energy Refinement) is a package of molecular simulation programs. It also refers to a set of molecular mechanical force fields for the simulation of biomolecules, which are used in a variety of simulation programs including AMBER and NAMD.

The latest version of AMBER is version 9 released on March 2006. The version we primarily used is version 8 released on March 2004.

Chapter 5 Introduction to Botulinum

Botulinum neurotoxins (BoNT) are among the most toxic proteins to humans. They are about 100 billion times [56] more toxic than cyanide. They have seven different serotypes (BoNT/A-BoNT/G), each produced by different strains of Clostridium botulinum.

The BoNT we studied is of type A (BoNT/A). We studied the properties, and the mechanism of the toxicity of BoNT/A in a low pH environment by computational simulations. Such properties and mechanism, elusive at the present time, are very important to understand because a low pH environment plays a very important role in the reaction of BoNT/A with cells.

Like most known poisonous substances, Botulinum toxin can be used as a biological weapon [57, 58]. Only about 10^{-7} g BoNT/A can kill a person, which means that a single gram of crystalline toxin, evenly dispersed and inhaled, can kill more than one million people. Although BoNT/A is among the most toxic proteins to humans, purified botulism toxin is the first bacterial toxin used as a medicine. The FDA licensed botulinum toxin as Oculinum in December 1989 for treating two eye conditions, blepharospasm [59] and strabismus [60],

characterized by excessive muscle contractions. It is now marketed under the trade name Botox, and is widely used for cosmetic purposes to remove wrinkles [61]. It is also used to treat blepharospasm, strabismus, and cervical dystonia [62]. Other applications for Botox are currently under investigation. It has been reported that spasmodic dysphonia, a neurological disorder that affects the muscles of the larynx, responds well to Botox treatment [63]. It has also been used to treat other dystonias, such as writer's cramp, as well as facial spasms, head and neck tremors and hyperhidrosis. A recent study [64] has even been conducted to observe its use in treating chronic neck and back pain.

BoNT/A is synthesized in Clostridium botulinum as a ~150 kDa single chain protein with 1,295 amino acids, which are cleaved endogenously or exogenously resulting in a 100 kDa heavy chain (HC) and a 50 kDa light chain (LC) linked through a disulfide bond(CYS429-CYS453). It is composed of three ~50 kDa functional domains [65, 66]: the catalytic domain which is confined to the ~50 kDa LC, the translocation domain which is confined to the N-terminal half of the ~100 kDa HC, and the receptor binding domain which is confined to the C-terminal half of HC. They are indicated in different colors in Figure 5-1. From the figure we can see that there is a belt part which belongs to the

translocation domain and is a long loop that wraps around the catalytic domain, as though to keep the catalytic domain in position. It plays an important role in shielding the active site[66] of the LC. The locations of the disulfide bond and the Zinc ion are shown in Figure 5-2. The red part shows the residues in the LC and HC linked by the disulfide bond. The blue part shows the location of the Zinc ion. Another representation of BoNT/A is shown in Figure 5-3 indicating the locations of alpha-helixes and beta-sheets. Figure 5-4 and Figure 5-5 also show the structure of BoNT/A in the top view and the left view. The dimensions of BoNT/A molecules [67] are approximately 45×105×130 $Å^3$. The overall dimensions for binding domain are 32×37×76 $Å^3$, for the translocation domain are 28×32×105 $Å^3$, while for the catalytic domain are 55×55×62 $Å^3$.

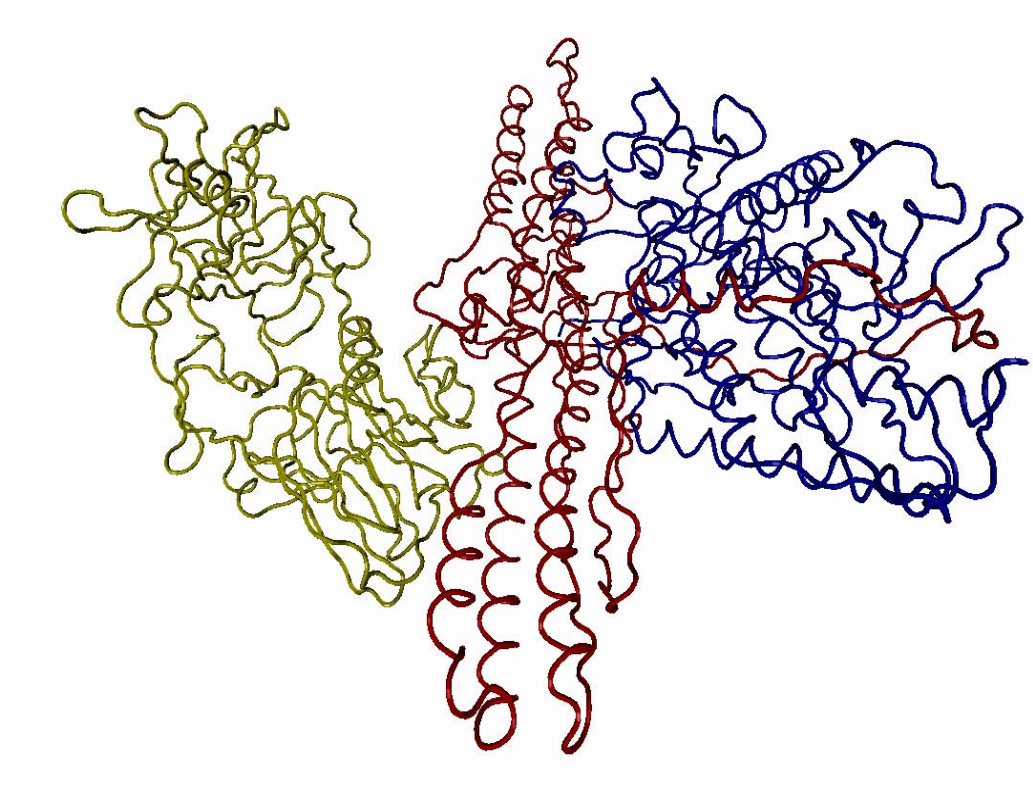

Figure 5-1 Domain structures of BoNT/A. yellow: binding domain; red: translocation domain; blue: catalytic domain

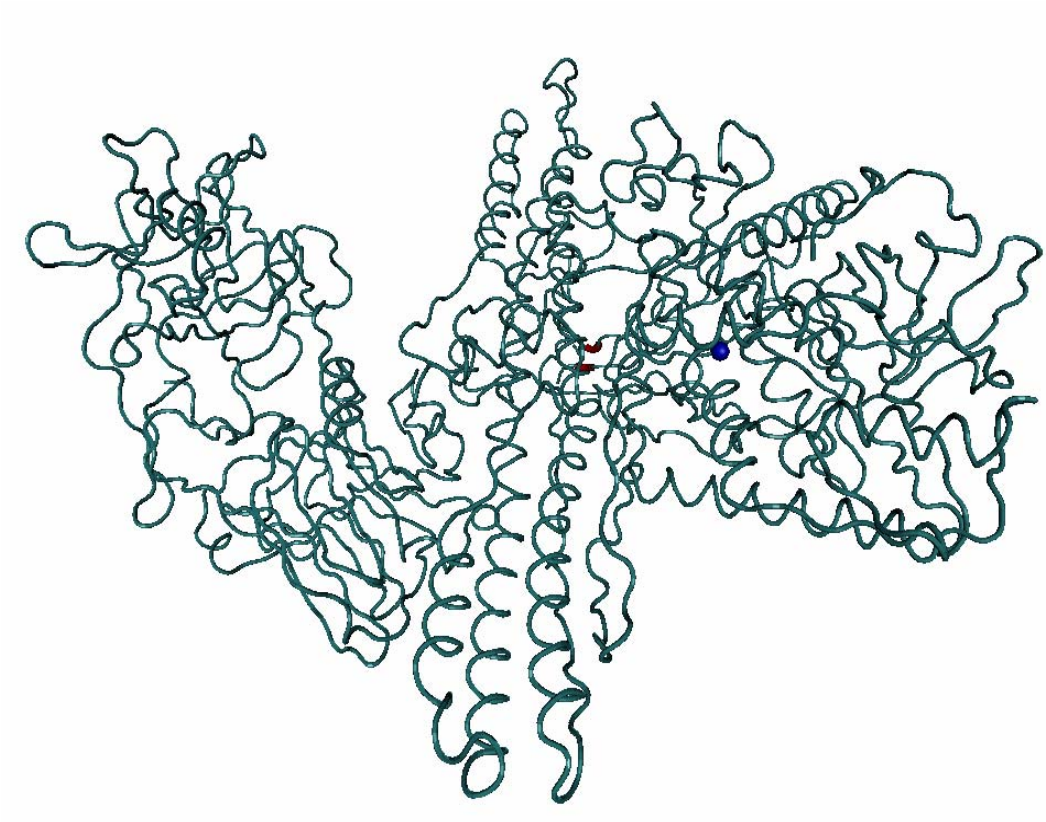

Figure 5-2 The locations of the disulfide bond and Zinc ion

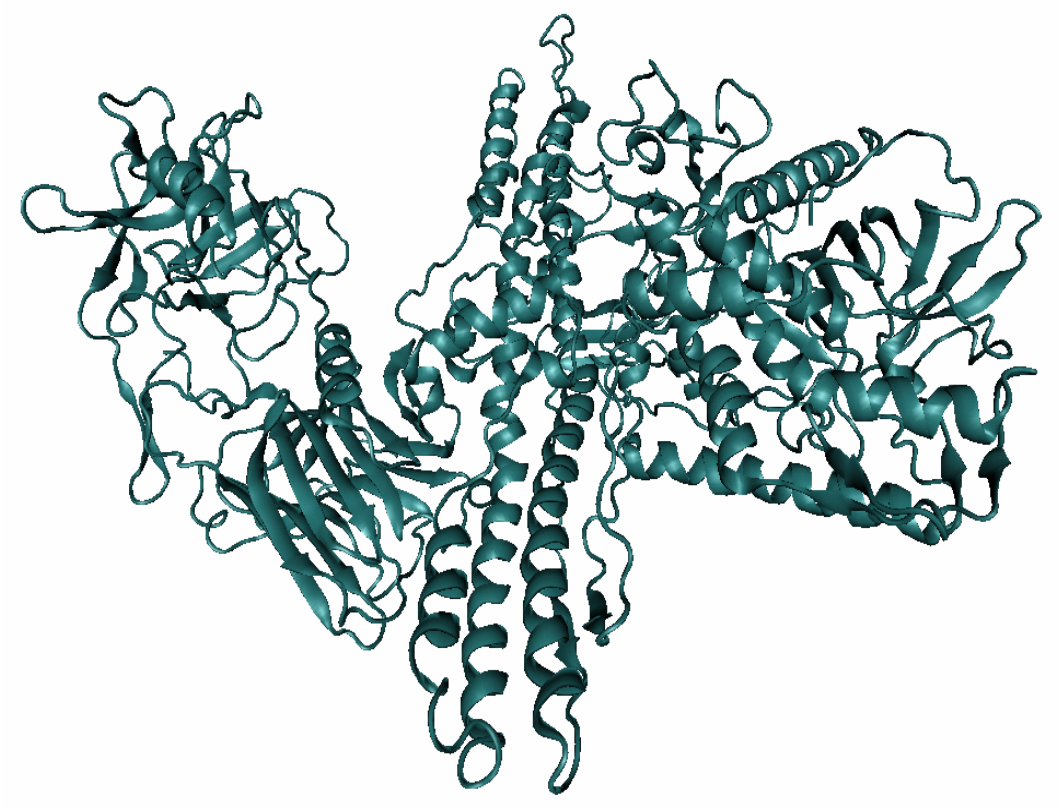

Figure 5-3 Alpha-helixes and beta-sheets in BoNT/A

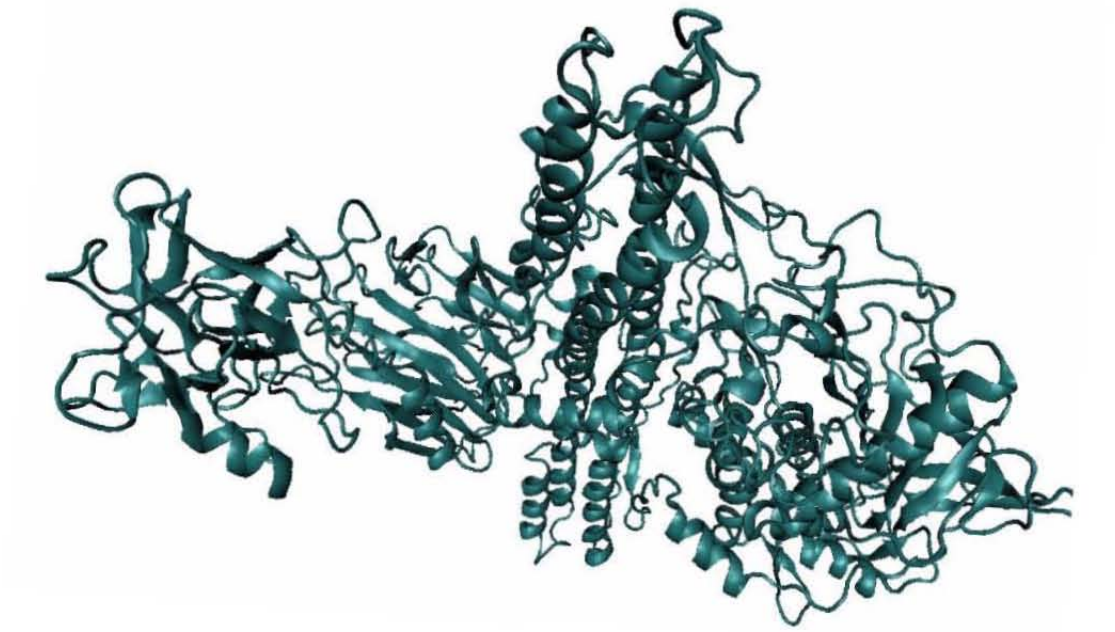

Figure 5-4 Alpha-helixes and beta-sheets in BoNT/A, top view

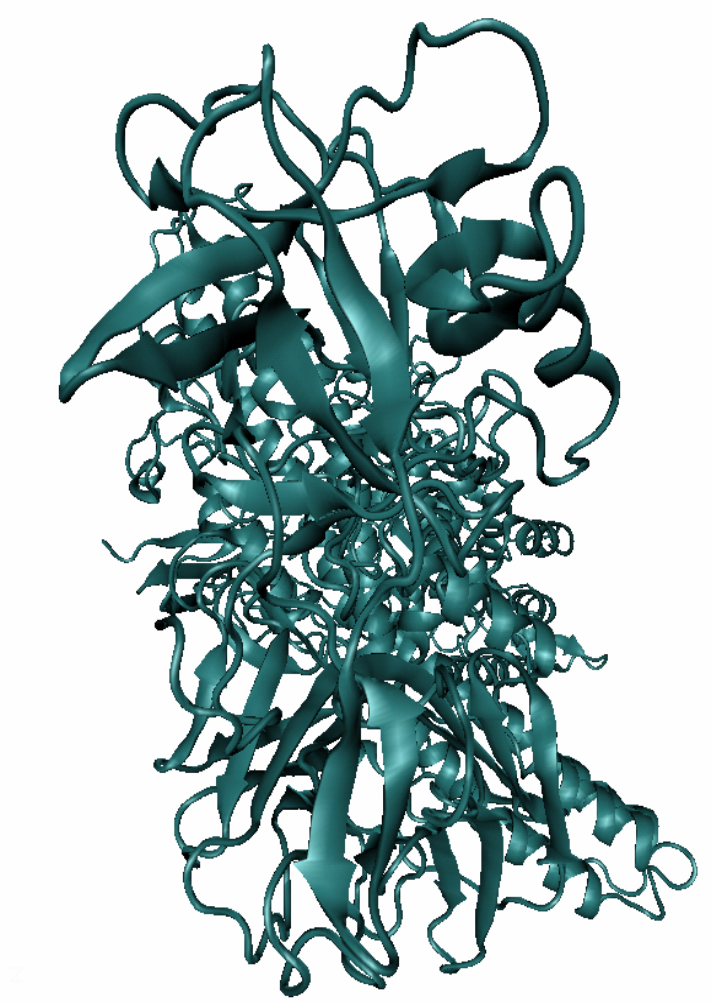

Figure 5-5 Alpha-helixes and beta-sheets in BoNT/A, left view

The toxicity of BoNT/A is proposed as a result of a four-step mechanism [56], as illustrated in Figure 5-6: extracellular binding, internalization, membrane translocation, and intracellular blockage of acetylcholine release.

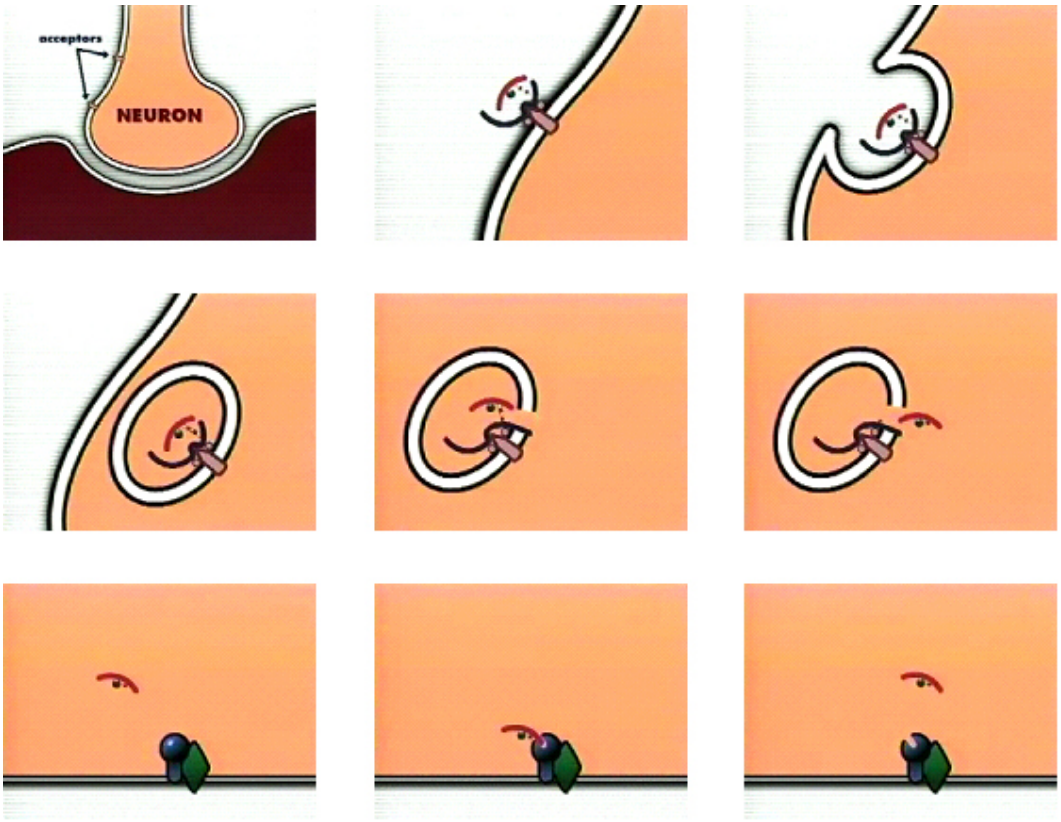

Figure 5-6 Botulinum Reaction Pathway

In the first step, the BoNT/A binds to the presynaptic nerve endings of cholinergic neurons through protein receptors on unmyelinated presynaptic membrane. This process is mediated by the binding domain of BoNT/A. In the second step, the BoNT/A is internalized by receptor-mediated endocytosis. After that, the low pH inside the endosome is believed[67] to trigger a membrane pore formation by the HC by introducing exposure of the hydrophobic polypeptide segment in the translocation domain. In this process, HC forms a channel, which

allows translocation of at least the LC across the endosomal membrane into the cytosol. Low pH in endosome is very important in this process; it is believed to cause conformation changes in the translocation domain to form the channel. The third step is the translocation of LC across the endosome membrane from lumen of endosome to cytosol. In this process, the HC and LC become separated after reduction of the disulfide bond. However, the size of the pore formed by HC has been estimated to be about 8 Å, which is too small to accommodate the 50 kDa LC. It has been speculated that the low pH could cause conformational change of the LC so that it can be accommodated in the pore formed by HC[56, 67]. When it reaches cytosol, a more neutral pH restores LC structure. In the final step, the LC specifically cleaves the synaptosomal protein of 25 kDa, SNAP-25(synaptosome-associated protein of 25,000 daltons), through its Zn^{2+}.

As one of the most toxic proteins, it is difficult and dangerous to handle BoNT/A for conducting traditional laboratory experiments, making computational simulation a necessity. However, simulating large proteins like BoNT/A is extremely challenging even for very large supercomputers due to the need to simulate it for long time scales. Until now, there is no known report for simulating BoNT/A for long-time scales.

Protonation states of titratable residues were changed to simulate an acidic pH environment for BoNT/A in this thesis. This method was used to verify another method proposed by Mongan et al. [68] to simulate the constant pH value in AMBER [53] version 8. BoNT/A was simulated at two different temperatures: the normal human body temperature of 37 °C and a higher temperature of 55 °C. The reason we chose 55 °C is that at this temperature spectacular structural changes were observed [65], the BoNT/A seems to be more stable at this temperature than at 37 °C when put in an acidic environment. And it was also simulated at two different pH values, the neutral one of pH 7.0 and an acidic one of pH 4.7, to study the influence of acidic environment on BoNT/A at two temperatures. We also simulated the LC of BoNT/A at 37 °C at a pH of 4.7 to study its conformational changes at this environment after being cleaved from the HC.

In our experiments, the structural variations of each domain were studied at different pH values and temperatures. It is organized as follows: modeling methods; simulation results and analysis; further discussions and conclusions.

Chapter 6 Simulation of BoNT/A at Different Temperatures and pH Values

6.1 Modeling Method

The structure data file of BoNT/A was obtained from the Protein Data Bank (PDB) [69] with the PDB code 3BTA. The parameter files, the coordinate file (crd file) and the topology file (top file) needed by NAMD [12, 13] were generated in AMBER version 7. The force field used was the AMBER force field ff99 [21]. The NAMD performed the molecular dynamics simulations, generating the results in dcd files, with which the Root Mean Square Deviation, or RMSD, values for BoNT/A were calculated using VMD [55]. All simulations were performed in the presence of explicit solvent, using TIP3PBOX water model, at different temperatures. Since Zn^{2+} plays an important role in its toxicity, it was included in all simulations.

Two classes of simulations were run, modeling conditions of neutral (pH 7.0) and low pH (pH 4.7) respectively. The method used to model the pH values was to change the protonation state of the residues according to its pKa values.

In biochemistry, the acid-ionization constant (Ka), or the acidity constant, is a specific type of equilibrium constant that indicates the extent of dissociation of hydronium ions from an acid. The equilibrium is of that a proton transfer from an acid, HA, to water, H_2O. The term for the concentration of water, $[H_2O]$, is omitted from the general equilibrium constant expression because it is almost equal to 1.0 in a dilute solution and it remains unchanged throughout dissociation.

The Ka is defined as follows in the interaction:

$$HA(aq) + H_2O \rightleftharpoons H_3O^+(aq) + A^-(aq)$$
$$K_a = \frac{[H_3O^+][A^-]}{[HA]} \tag{6.1}$$

Since the constant differs for each acid and varies over many orders of magnitude, the acidity constant is often represented by the additive inverse of its common logarithm, represented by the symbol pKa using the same mathematical relationship as $[H^+]$ to pH:

$$pK_a = -\log_{10} K_a \tag{6.2}$$

When the pH value is far smaller than the pKa value, the reaction moves to its left hand side and almost all the titratable group are in the protonated states. When the pH value is far larger than the pKa value, the situation is the contrary,

the reaction moves to its right hand side and almost all the titratable group are in the deprotonated states. In general, the fraction protonated F_{HA} can be calculated using the formula:

$$F_{HA} = \frac{1}{1+10^{pH-pK_a}} \tag{6.3}$$

We used Root Mean Square Deviation (RMSD) to analyze the extent to which the system has moved from equilibrium. It characterizes the amount by which a given selection of the molecule deviates from a defined position. It is defined as follows:

$$RMSD_{\alpha}(t_j) = \sqrt{\frac{\sum_{\alpha=1}^{N_\alpha}\left(\vec{r}_\alpha(t_j) - \vec{r}_{\alpha 0}\right)^2}{N_\alpha}} \tag{6.4}$$

where N_α is the number of atoms whose positions are being compared, $\vec{r}_\alpha(t_j)$ is the position of atom α at time t_j, and $\vec{r}_{\alpha 0}$ is the initial value of the position of atom α.

For the analysis of BoNT/A structure at various pH values, we obtained the RMSD by calculating the deviation of the molecule structure at a certain time compared to the initial structure. RMSD values were calculated for all atoms of the protein backbone (without hydrogen) for the entire protein. In the analysis, RMSD for the belt part was also included besides the RMSD for the

whole protein and its three domains.

The method used to model the pH values was to change the protonation states of the residues before the simulation started and kept it constant.

According to the pKa values of all the titratable residues, varying the pH values from pH 7.0 to pH 4.7 requires protonation states of Histidine (HIS) residues to change while the protonation states of other residues mostly remain unaffected as we can see from Figure 6-1. The BoNT/A contains 12 HIS residues; we changed protonation states of them in low pH simulations. They are represented in red in Figure 6-2.

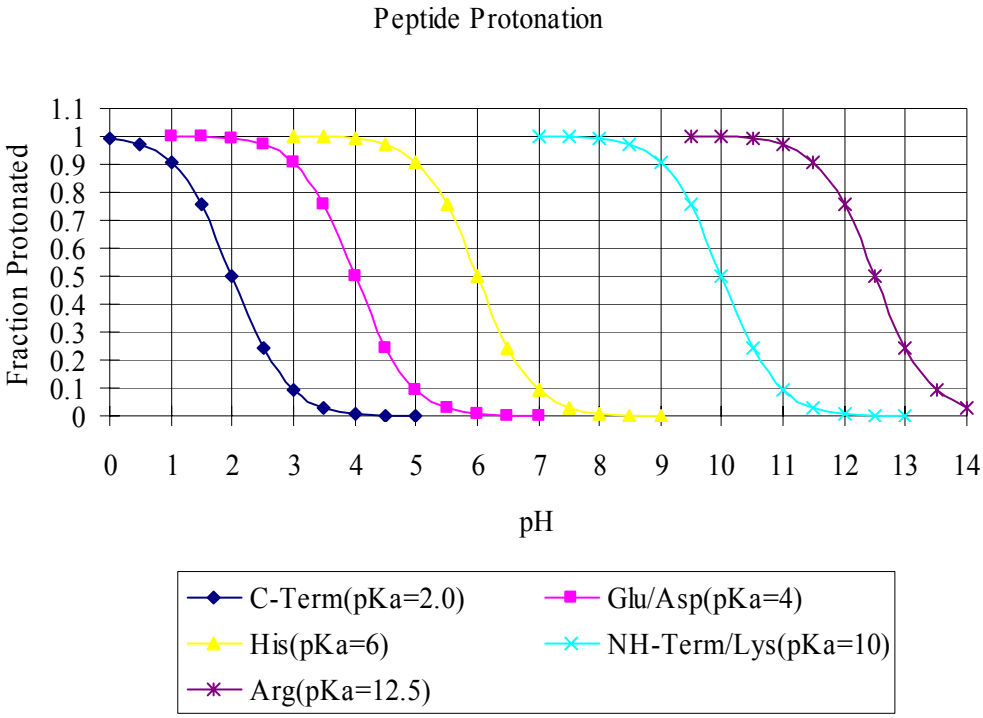

Figure 6-1 Titration curve

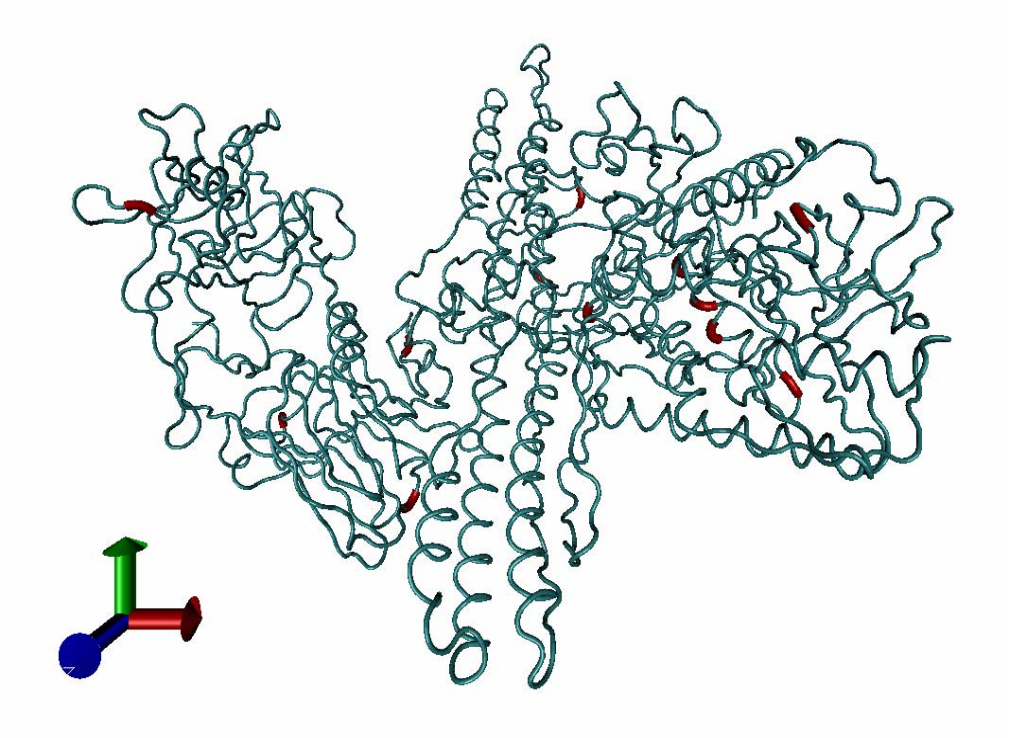

Figure 6-2 HIS residues in BoNT/A

6.2 Results and Analysis

The summary of all computational experiments we conducted is recorded in Table 6-1; simulations were performed at two temperatures: 37 °C and 55 °C. Two pH values were chosen for our simulations, a neutral one of pH 7.0, and an acidic one of pH 4.7, which is the pH value at which many biological experiments were conducted with speculation that interesting phenomena may

occur. For the pH 4.7 at 37 °C experiments, we also made a special simulation by cutting off the LC of BoNT/A and simulated it separately to help understanding its mechanism of toxicity. The Table 6-2 shows the details of and the difference between the systems used for the simulations of BoNT/A at neutral and low pH values.

Table 6-3 shows some of the previous simulations of proteins before our simulation [4-6] to understand the time scales needed to simulate a protein based on its molecular weight. The number of residues simulated for BoNT/A is 1277 instead of 1295, the total number of residues of BoNT/A, because there is 18 missing residues whose location cannot be determined. From this table, we can see that for small proteins with ~100 residues, a simulation period of about ~10 μs is needed. For large proteins like BoNT/A, no similar simulations were performed. However, it can be logically inferred that a simulation timescale on the order of microseconds is needed.

Temperature	**pH Values**	
	pH 4.7	**pH 7.0**
37 °C	Whole protein (~64 ns)	Whole protein (~63 ns)

	LC only (200 ns)	LC only (200 ns)
55 °C	Whole protein (~57 ns)	Whole protein (~67 ns)
	LC only (200 ns)	LC only (200 ns)

Table 6-1 Summary of parameters for all numerical experiments

	Neutral pH (pH 7)	**Acid pH (pH 4.7)**
Histidine side chains	Not protonated	Protonated
Overall protein charge	-9	+3
Zinc ion	Present	Present
Bounding box (Å)	106×120×158	106×120×158
Volume (Å$^{3)}$	2,023,840	2,023,840
Total number of atoms	173,549	173,561
Protein atoms	20,698 + Zn^{2+}	20,710 + Zn^{2+}
Water molecules	50,950	50,950
Total mass	1,065,388 Da	1,065,400 Da
Total density	0.874 g/cc	0.874 g/cc

Table 6-2 The difference between the systems used for the simulations of BoNT/A at neutral and low pH values

Name	# of residues	Simulation time
Villin headpiece (alpha helical protein)	36	~10 μs
Trp-cage	20	~100ns at 315K
Beta hairpin	54	38μs at 300K
BoNT/A	1277	Unknown
BoNT/A LC only	431	Unknown

Table 6-3 Simulation time needed

6.2.1 Constant pH Simulations by Constant Protonation State

6.2.1.1 Whole Protein Runs

Figure 6-3 and Figure 6-4 show the RMSDs of BoNT/A and its three domains as functions of time at pH 4.7 and pH 7.0 at a temperature of 37 °C. The RMSD of the belt part is also shown. From these two figures, we find that the RMSD of the belt part oscillates the most, which suggests that it is the most flexible part of the translocation domain. Figure 6-5 compares of the RMSDs of the whole protein at two pH values. From this, we can see that the BoNT/A have

larger RMSD at the low pH environment, which indicates that the BoNT/A may have some conformation changes at pH 4.7.

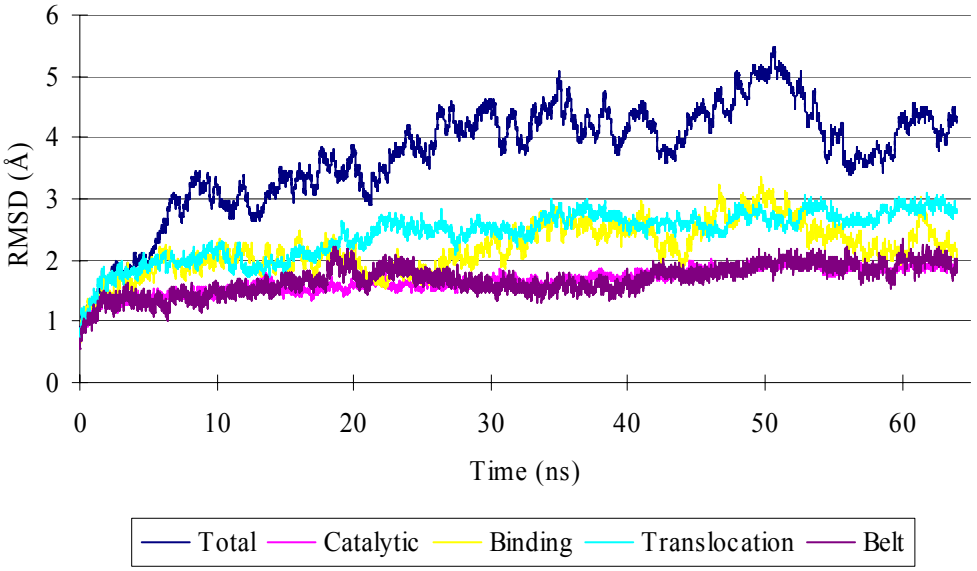

Figure 6-3 RMSD of BoNT/A for pH 4.7 at 37 °C

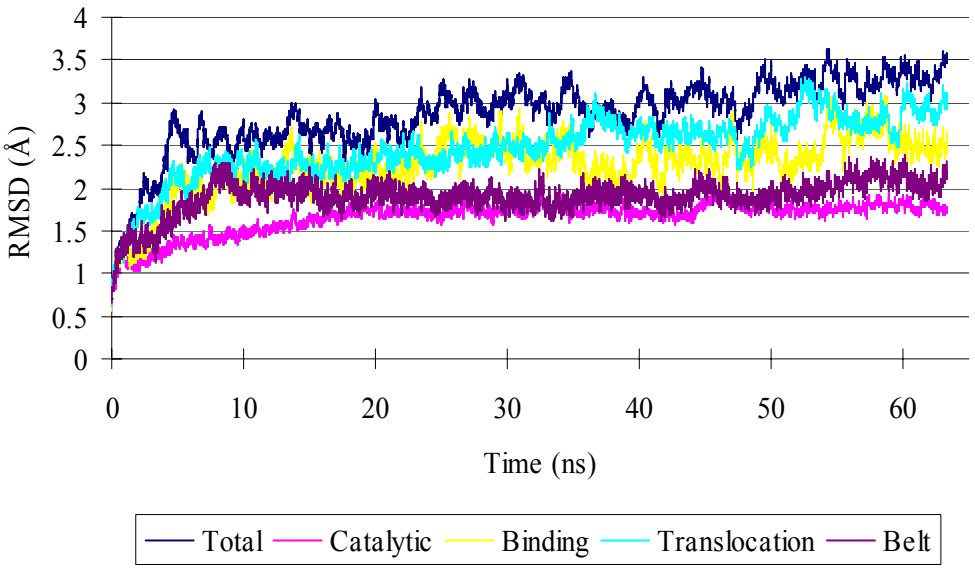

Figure 6-4 RMSD of BoNT/A for pH 7.0 at 37 °C

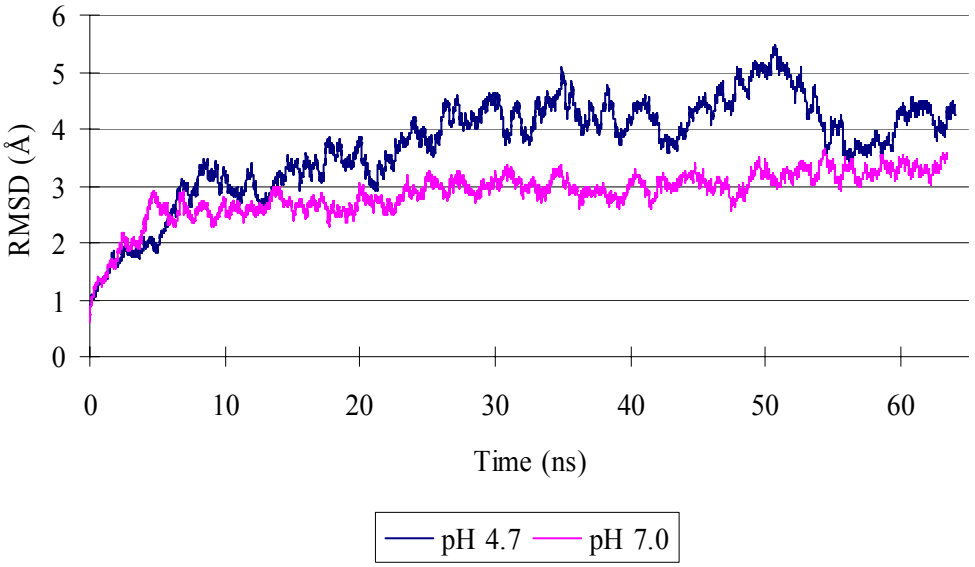

Figure 6-5 Comparison of the RMSD for the whole protein at 37 °C

To see it more clearly, Figure 6-6 was made as follows; the points indicate the average RMSD and the deviation of it for the protein at pH 4.7 and pH 7.0 at 37 °C and 55 °C. From it, we can see that BoNT/A behaves differently at different pH values. In an acidic environment, the RMSD of the total protein decreases as temperature increases, while in the neutral environment, the RMSD of the total protein increases with increasing temperature. It may imply that BoNT/A is more stable and temperature resistant in an acidic environment. Figure 6-7 shows the comparison of the RMSD for the belt part.

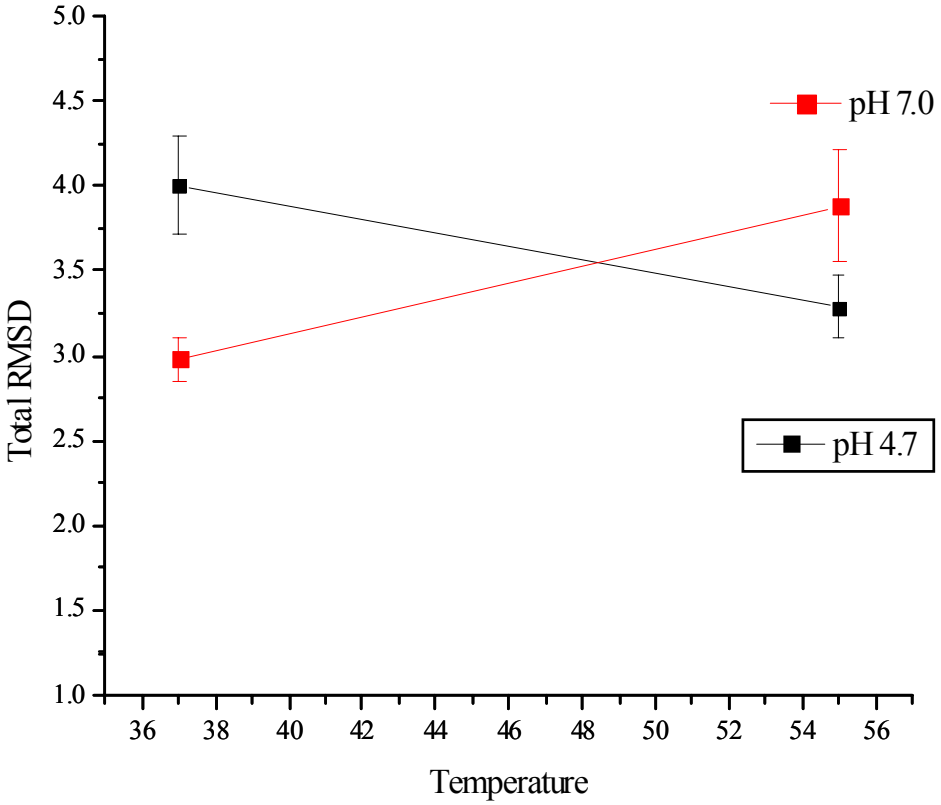

Figure 6-6 Comparison of the RMSD for the whole protein

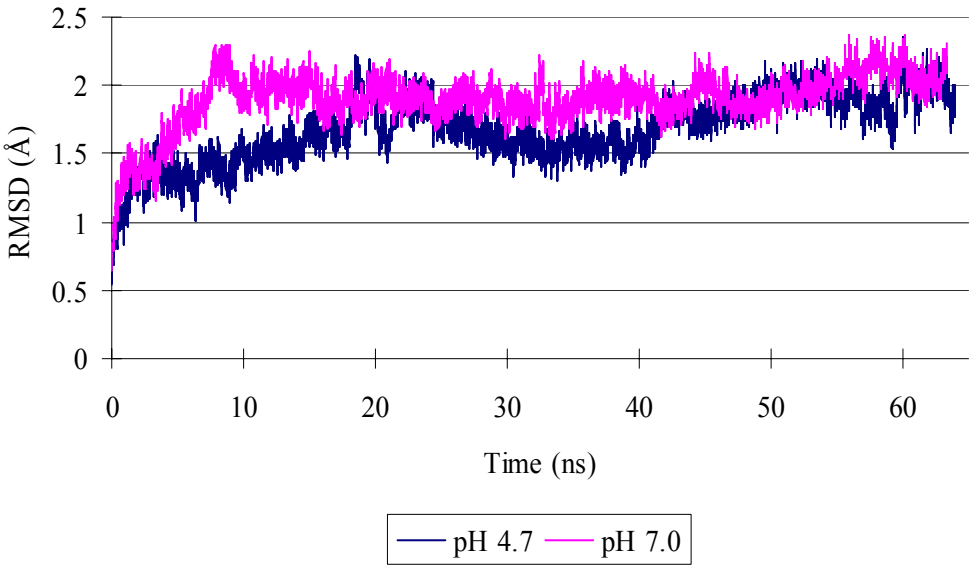

Figure 6-7 Comparison of RMSD for the belt part

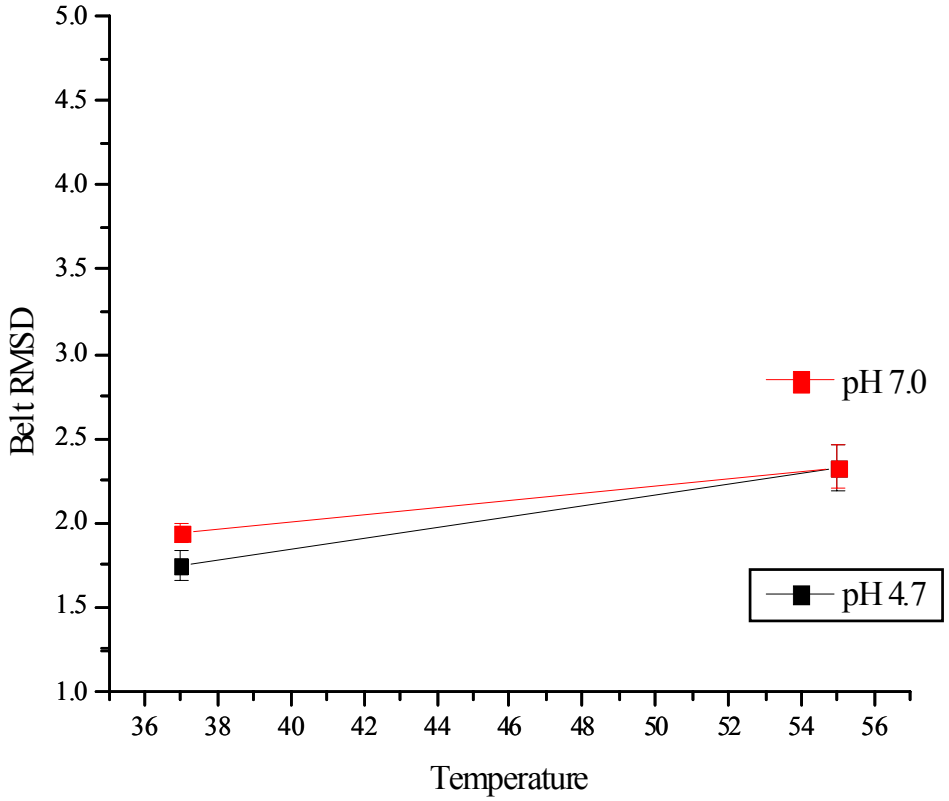

Figure 6-8 Comparison of Belt RMSD

Using the same method as we did on Figure 6-6, we made Figure 6-8. It may imply that the conformation of belt part changes larger with increasing temperature at low pH environment. Also, the RMSD of the belt part has a different value at 37 °C since their error bars do not cross, which implies that at this temperature, different pH values greatly affect the structure of belt part.

Li et al. [65] found that at a temperature of about 55 °C, the LC denatures at pH 7.0 while remaining stable at low pH 4.7. Their results of temperature-dependent unfolding of LC at pH 7.0 and pH 4.7 as monitored by the near-UV circular dichroism band at 280 nm were shown in Figure 5 of their paper. Their results indicated that at pH 4.7, much of the tertiary structure of the light chain is still retained at 100 °C, while at pH 7.0 it is completely lost. The ANS (1-anilinonaphthalene-8-sulfonate) binding experiment indicated that the acidic pH value leads to the conformation change of the LC, and this change is reversible. For the secondary structure of the light chain, it is almost completely lost at 55 °C when put in a pH 7.0 environment, while it is retained when put in a pH 4.7 environment.

Simulations of BoNT/A at 55 °C were conducted to examine this properties at this temperature. The RMSD results are shown in Figure 6-9 and Figure 6-10. From Figure 6-11, comparing the RMSD for the whole protein at 55 °C, we find that the RMSD at pH 4.7 is lower than at pH 7.0, which supports the idea that the whole protein is more stable in an acidic environment than in a neutral environment. Figure 6-12 shows the comparing of RMSD for the LC at 55 °C. From it, we can see that although RMSD for the LC at pH 4.7 is lower

than that at pH 7.0 before 15 ns, they cross after that time. We expect that longer simulations would verify the claims by Li et al. of LC at 55 °C.

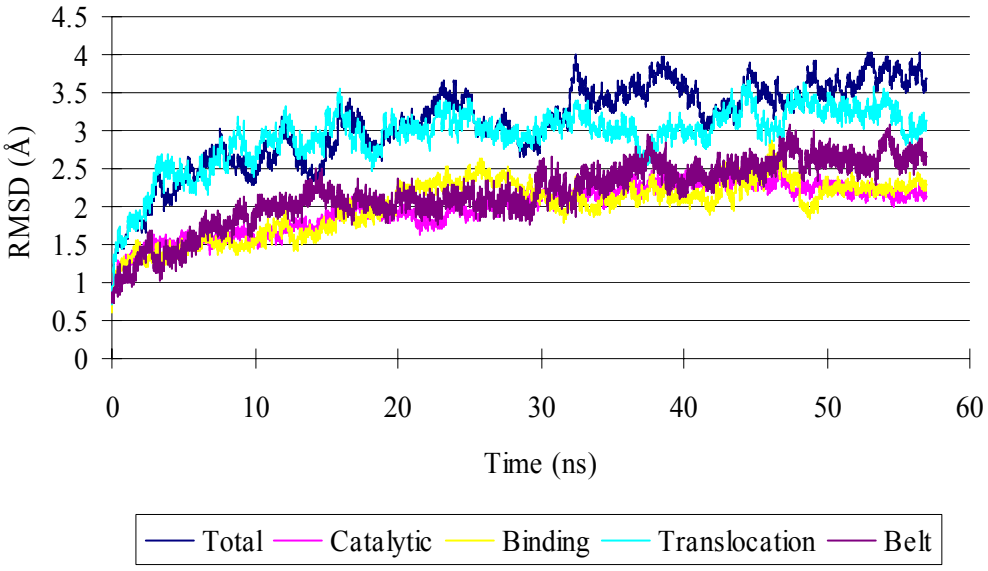

Figure 6-9 RMSD of BoNT/A for pH 4.7 at 55 °C

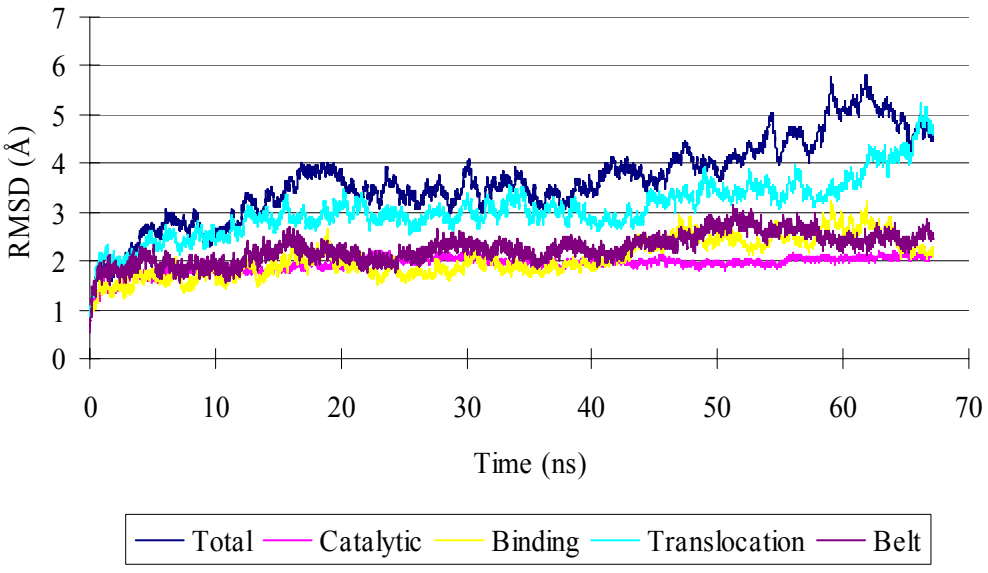

Figure 6-10 RMSD of BoNT/A for pH 7.0 at 55 °C

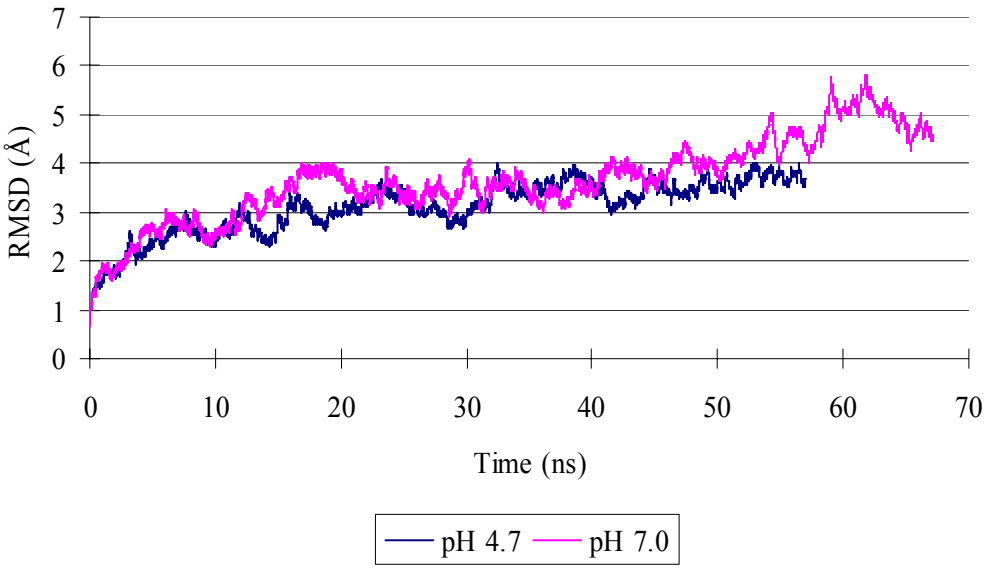

Figure 6-11 Comparing of RMSD for the whole protein at 55 °C

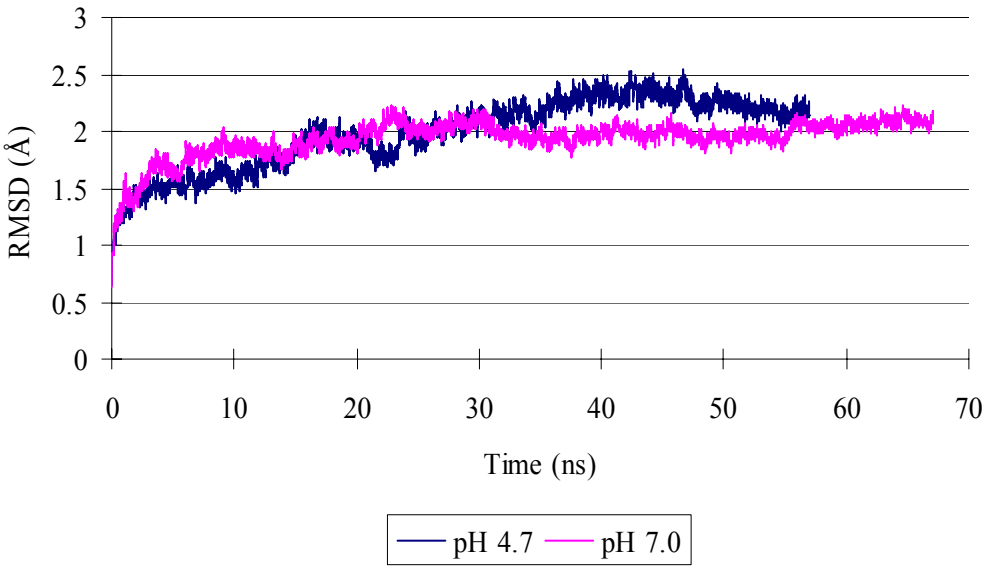

Figure 6-12 Comparing of RMSD for the LC at 55 °C

6.2.1.2 Light Chain Only Runs

It is believed that the LC is only active after reduction of the disulfide bonds [56]. The LC of BoNT/A neurotoxin will undergo autocatalytic fragmentation that is accelerated by the presence of the metal cofactor, zinc. [70] It is possible that the low pH triggers a conformational change that forms the LC transporter channel, which must fits within the membrane bilayer (~30 Å). The whole process is still not clear. Some suggest that the translocation of the BoNT/A LC may occur through a mechanism other than the membrane channel formed by the BoNT/A HC. In a "cleft model", interaction of the BoNT/A LC directly with the lipid bilayer has been proposed [71]. To understand this process, we simulated the LC after it had been cut off from the other part of the protein with a Zn ion.

The LC contains six HIS residues which changed the protonation states during the pH change. The net charge of the LC is -1 not including the Zn^{2+} at pH 7.0 (Histidine not protonated), and 5 not including the Zn^{2+} at pH 4.7 (Histidine protonated). Table 6-4 shows the differences between the systems used for the simulations of LC for pH 4.7 at 37 °C. For LC only runs, the number of atoms is about 1/3 of the whole protein run, which greatly increased

the simulation speed. With 64 processors at NankaiStars, we can get about 1.2 ns physical time per day, compared to 0.5 ns per day for the whole protein run. This is very important for us, since the possible conformational change occurs over a long time scale.

	Cut-off run	**Whole protein run**
Total atoms	59,672	173,561
Protein atoms	6,958 + Zn^{2+}	20,710 + Zn^{2+}
Water molecules	17,571	50,950

Table 6-4 Differences between the systems used for the simulations of LC for pH 4.7 at 37 °C

The result for pH 4.7 at 37 °C is shown in Figure 6-13. From it, we can see that the RMSD of cut-off LC is significantly higher than the normal run. It is very important since for the mechanism of BoNT/A's toxicity, the LC is released from the HC belt after disulfide reduction, thereby making the active site accessible to the substrate. Our results imply that the LC will have a conformational change after it being released, which supports the experimental expectation.

We also have results for LC only for pH 7.0 at 37 °C in Figure 6-14. We can see that the RMSD of the LC is significantly higher after being cut off. This also clarifies that the LC's activity depends on its being cut-off from the other part of the protein.

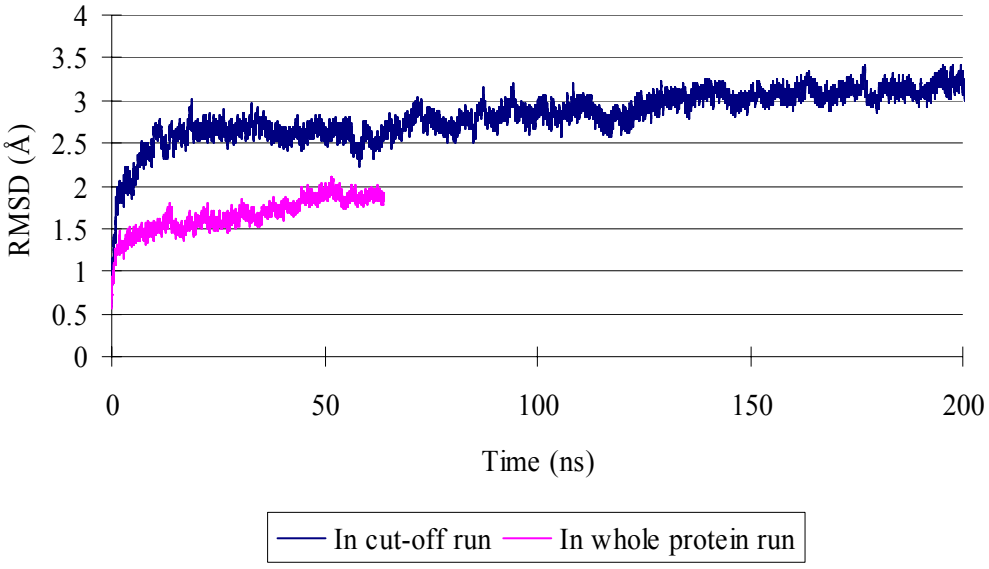

Figure 6-13 Cut-off simulation for LC for pH 4.7 at 37 °C

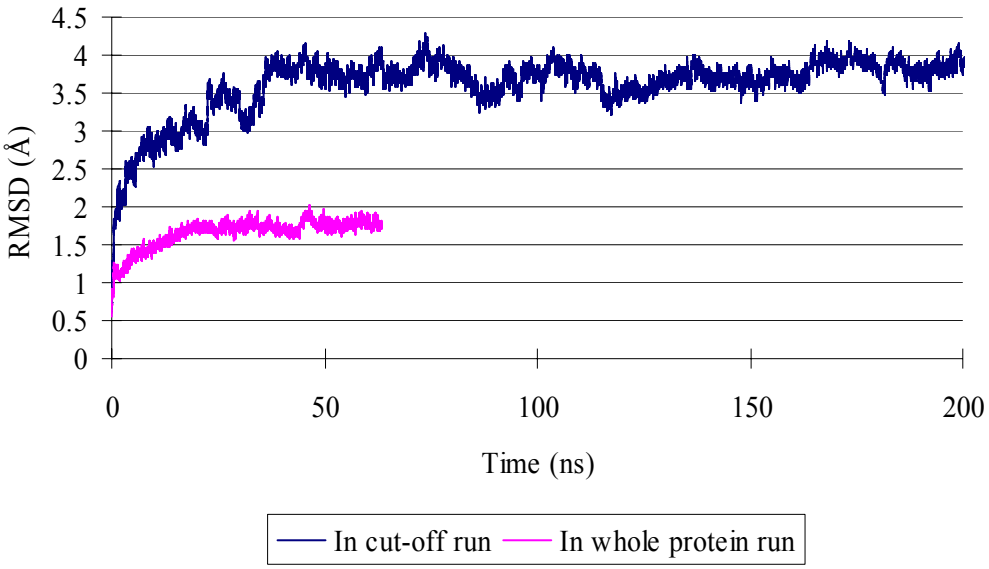

Figure 6-14 Cut-off simulation for LC for pH 7.0 at 37 °C

Figure 6-15 and Figure 6-16 show the LC-only results compared with the normal runs at a higher temperature, 55 °C. In Figure 6-17, we compare the RMSD for LC after being cut off at different pH and temperatures. From this graph, we can see that at 55 °C , the pH 4.7 run (the red one) has a lower RMSD than the pH 7.0 run (the light blue one), which is what we expected [65]: the acidic environment makes the protein more stable at higher temperature. While at 37 °C, the pH 4.7 run (the blue one) also have a much lower RMSD than the pH 7.0 run (the yellow one). It may indicate that at these two temperatures, the LCs are more stable in an acidic environment than in a neutral environment.

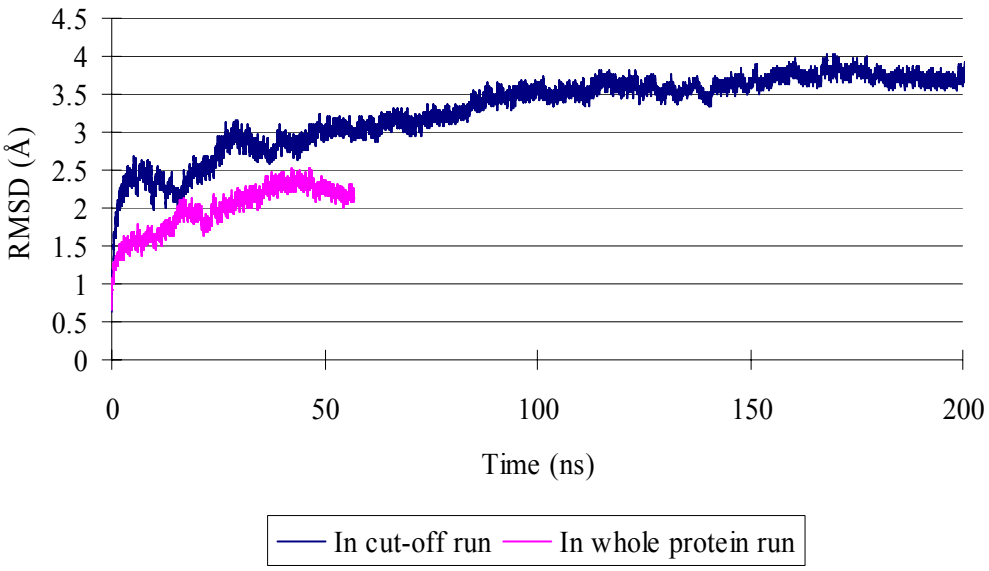

Figure 6-15 Cut-off simulation for LC for pH 4.7 at 55 °C

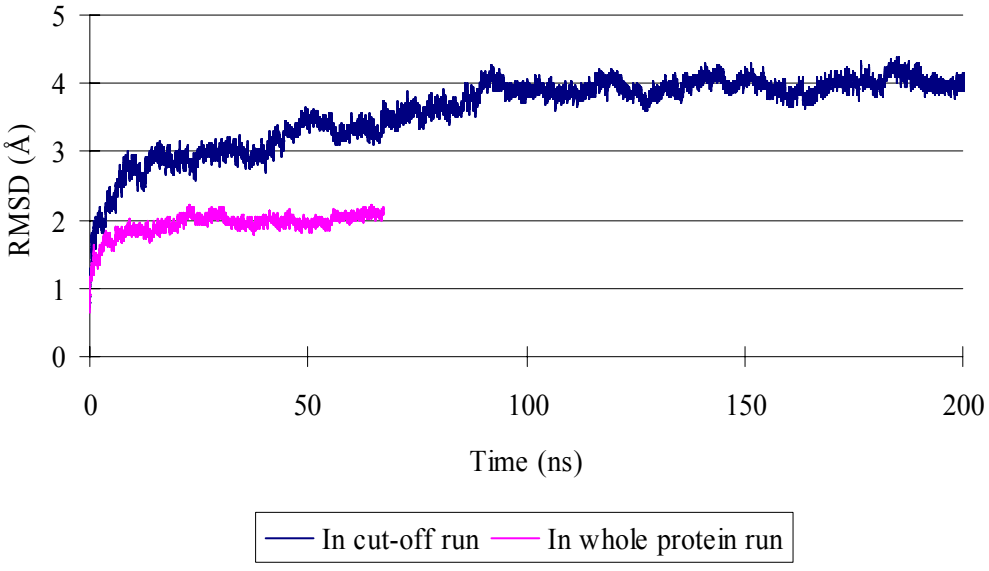

Figure 6-16 Cut-off simulation for LC for pH 7.0 at 55 °C

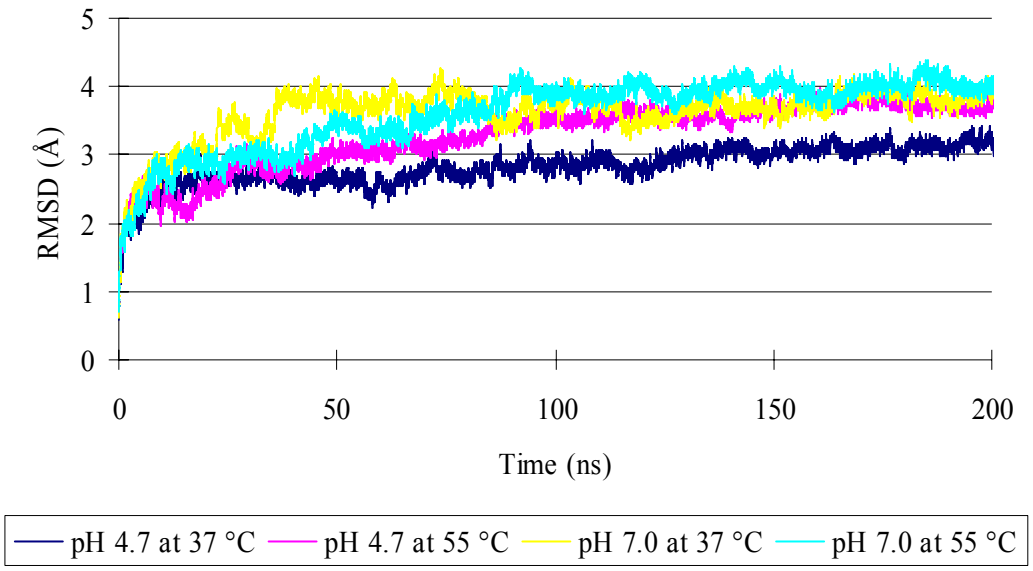

Figure 6-17 Cut-off simulation for LC at different pH and temperatures

With the same method used in drawing Figure 6-6, we obtained Figure 6-18 for RMSD of LC only runs and Figure 6-19 for RMSD of LC in whole proteins runs. In these two figures, we see significant difference. Generally, they have larger RMSD values when running alone than running as part of the protein. Also, as temperature increases, they become more stable, while in the whole protein run, the situation is contrary. Their RMSDs become larger with increasing temperatures. There must be some conformational change for the LC after it has been cut off and run separately.

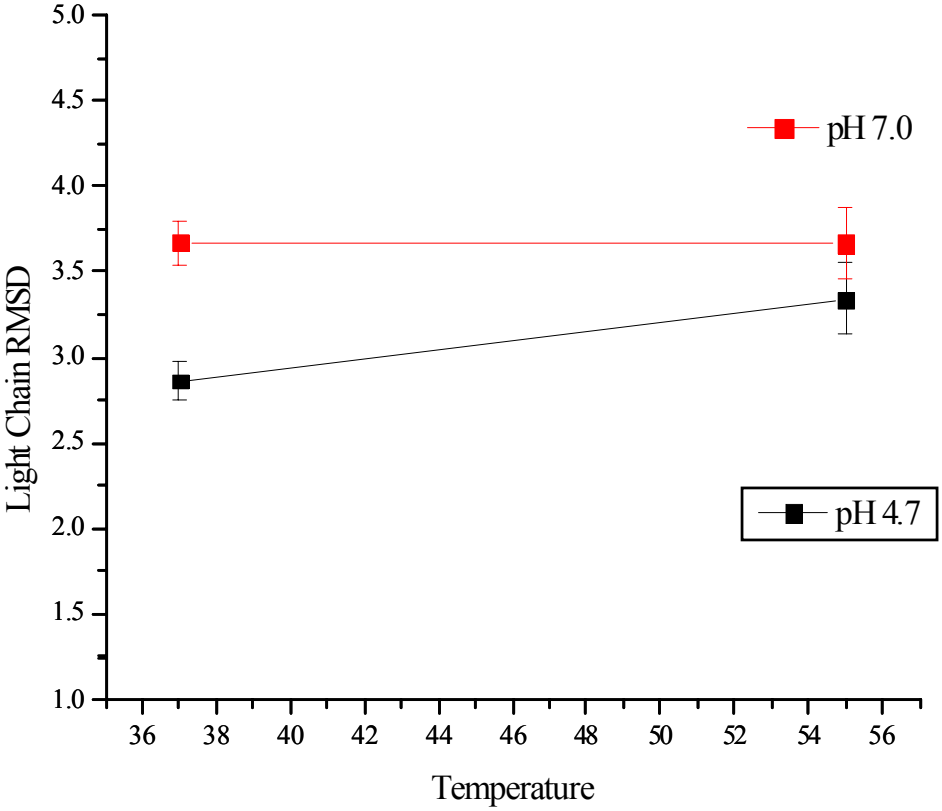

Figure 6-18 Comparison of RMSD of LC only runs

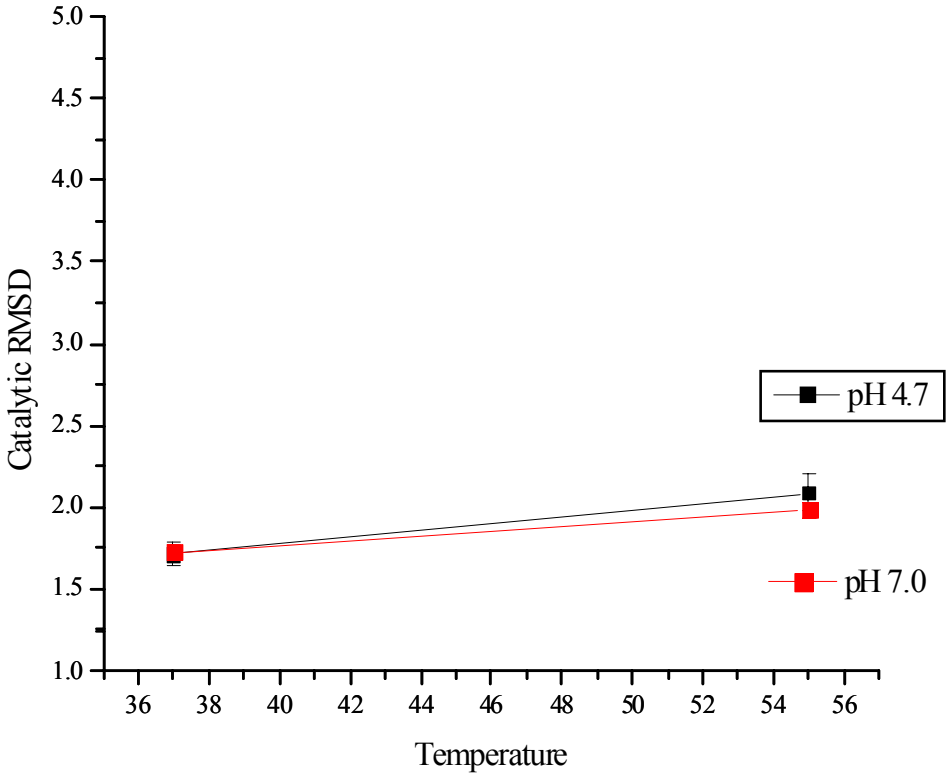

Figure 6-19 Comparison of RMSD of Catalytic domain in whole proteins runs

To find out the difference, we compared the configuration of two structures in LC only run and whole protein run at the same time. Figure 6-20 shows the deviation of these two structures. The red part means the largest conformational change. From the graph, we can see that the edge part of the LC has a higher deviation. It can be inferred from the graph that after the LC is cut-off from the protein, its conformational change starts from its edge part, and as time goes on,

it eventually causes the denaturation of the whole LC.

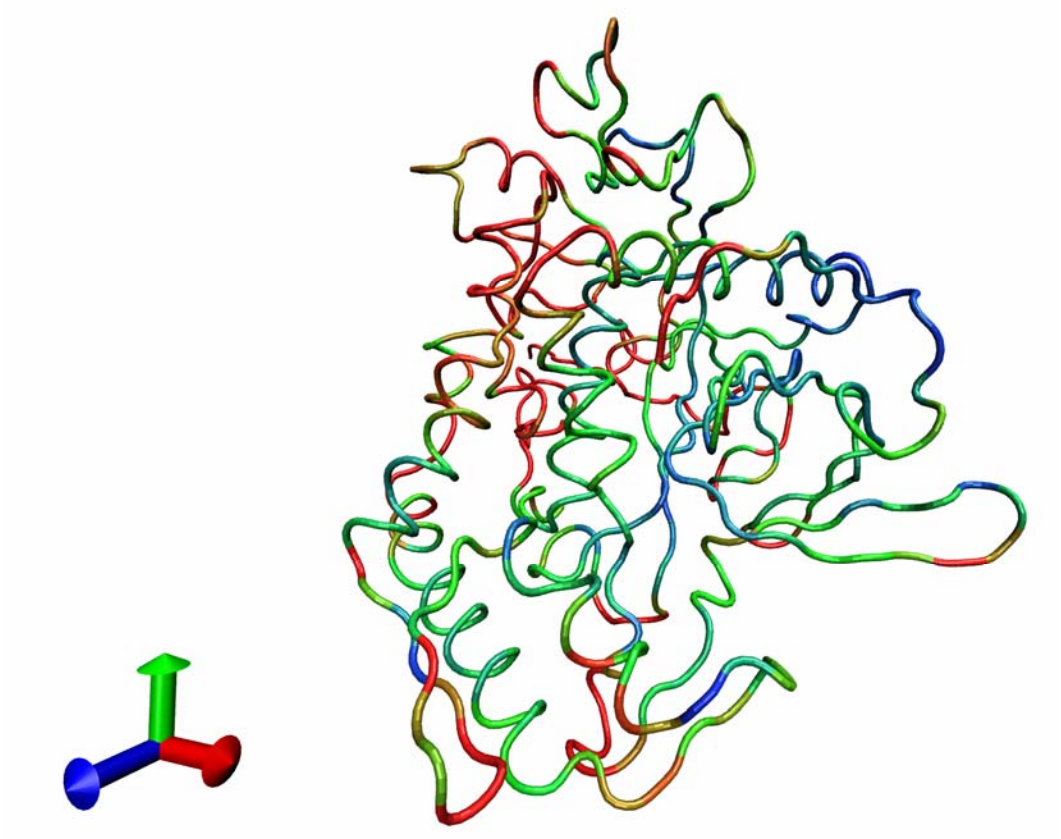

Figure 6-20 Comparing of LC only run and whole protein run

6.2.1.3 Low Frequency Phenomena

Oscillations of the binding and the catalytic domain during the simulation were observed in our simulations, especially in the Y (the direction of the two

alpha-helix in the translocation domain) and Z (the direction perpendicular to the surface of Figure 5-1) directions. Figure 6-21 shows the distances between the center of mass of each domain to that of the whole protein. This also confirms the conservation of momentum. In addition, this explains why there are some peaks on the RMSD of total protein, while the RMSD of the three domains are stable, like the peak at 59.05 ns in Figure 6-10.

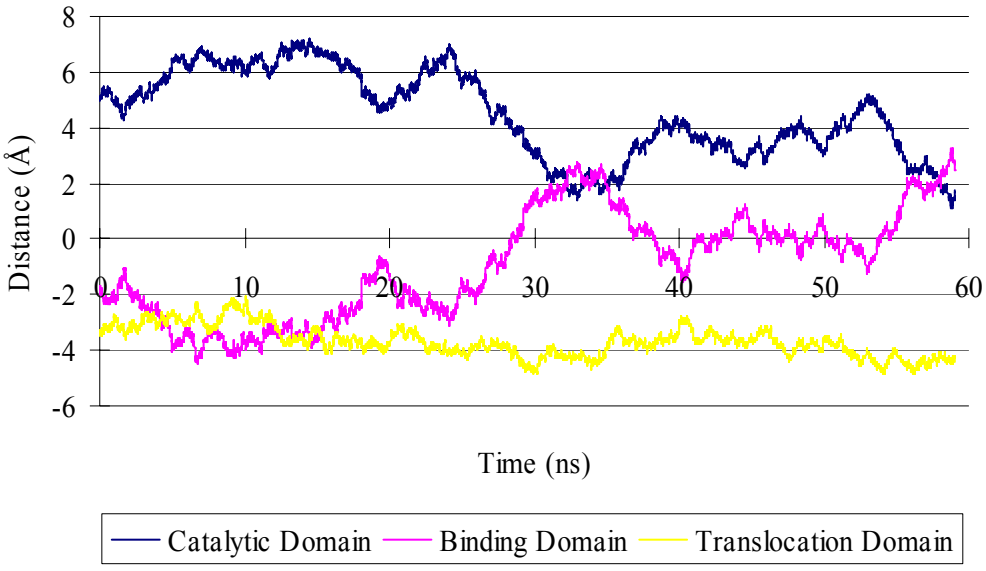

Figure 6-21 Low frequency oscillation in Z direction for pH 7.0 at 55 °C

6.2.1.4 Structure Analysis

To investigate deeply inside what happens during the temperature and pH change, we need to take a close look of the protein structural change.

Firstly we compared the structure of the whole protein at 37 °C at two pH values after simulation of 50 ns. The protein structures after alignment are shown in Figure 6-22. Figure 6-23 shows the structures after alignment of the whole protein at 55 °C at two pH values after 50 ns.

The structures after alignment for the LC only runs after 190ns are shown in Figure 6-24 for 37 °C and Figure 6-25 for 55 °C. Because the pH change has a significant effect on the energy and structure of the LC, we need an in-depth analysis for the individual HIS residues whose protonation states change with pH value at the atom-level.

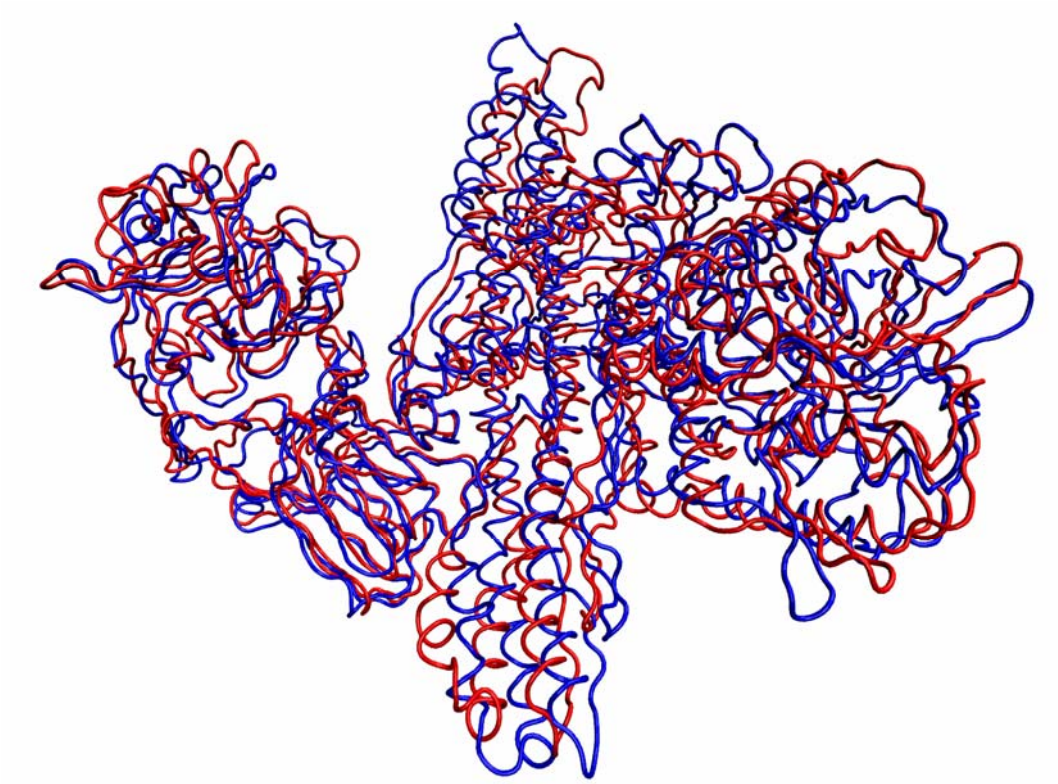

Figure 6-22 Structures of BoNT/A at different pH and 37 °C, Red: pH=4.7; Blue: pH=7.0

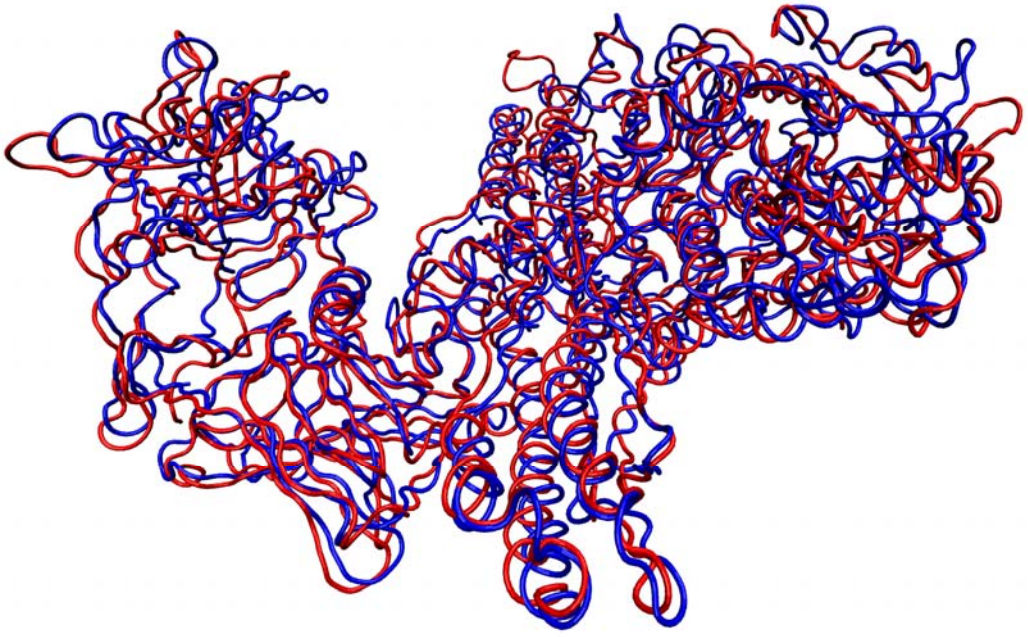

Figure 6-23 Structures of BoNT/A at different pH and 55 °C, Red: pH=4.7; Blue: pH=7.0

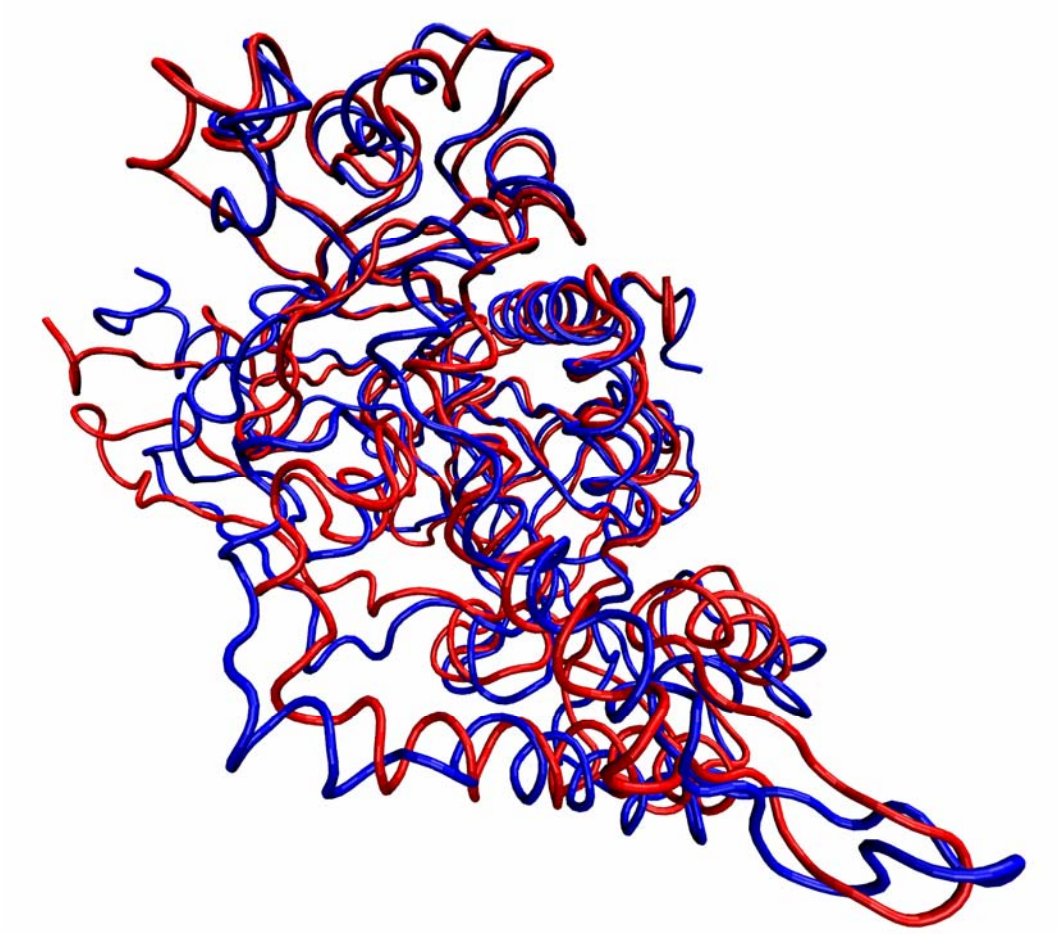

Figure 6-24 Structures of LC at different pH and 37 °C, Red: pH=4.7; Blue: pH=7.0

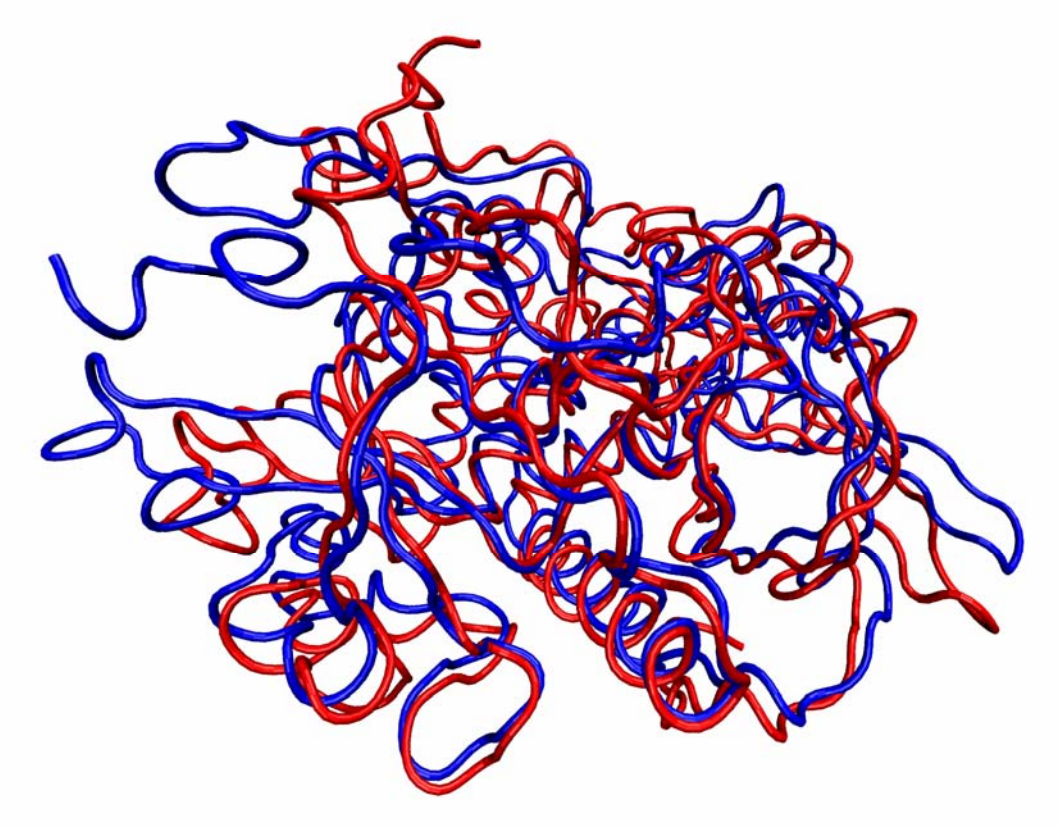

Figure 6-25 Structures of LC at different pH and 55 °C, Red: pH=4.7; Blue: pH=7.0

BoNT/A is believed to be a zinc endopeptidase [72] that contains the consensus sequence HEXXH (residues 222-226) in the LC. Its crystal structure supports a model in which the HIS222, HIS226, and GLU261 of the HEXXH motif directly coordinate the zinc, and GLU223 coordinates a water molecule as the fourth ligand[67, 73]. Since the zinc atom is coordinated with three amino acid residues and an activated water molecule (nucleophilic water), its role is

thought to be catalytic[74, 75]. Results show the critical importance of the presence of zinc ion in the cleaving progress of SNAP-25[72].

Other results also show the structure of BoNT/B at various pH values ranging from 4 to 7[76]. Their research suggested that at a low pH environment, the coordination may be lost, but the zinc ion retains its catalytic function. Figure 6-26 and Figure 6-27 shows the coordination between the zinc ion and HIS222 and HIS226 at pH 4.7 and 7.0 and 37 °C. Figure 6-28 and Figure 6-29 show the coordination between the zinc ion and HIS222 and HIS226 at pH 4.7 and 7.0 and 55 °C. From these figures, we find that at both temperatures the protonated histidine residue with +1 charge repels the zinc ion, which confirms S. Eswaramoorthy's experimental results[66], but the zinc ion remains at its location during the simulation.

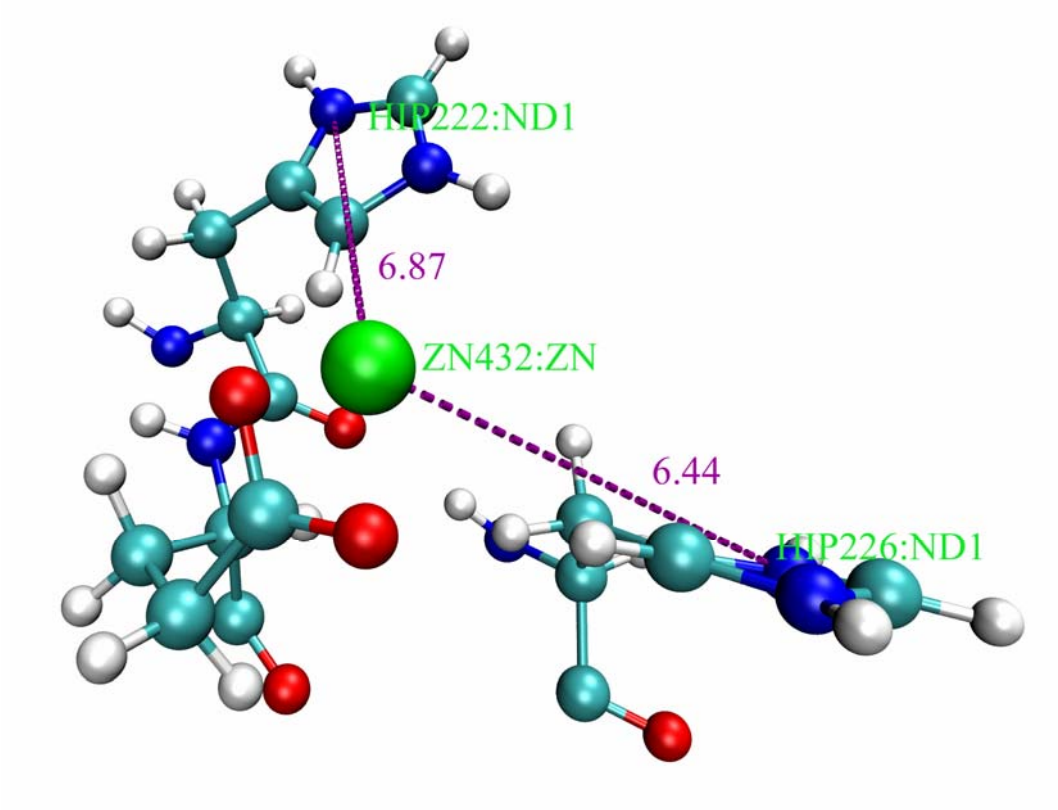

Figure 6-26 Coordination between zinc ion and HIS222 and HIS226 at pH 4.7, 37 °C

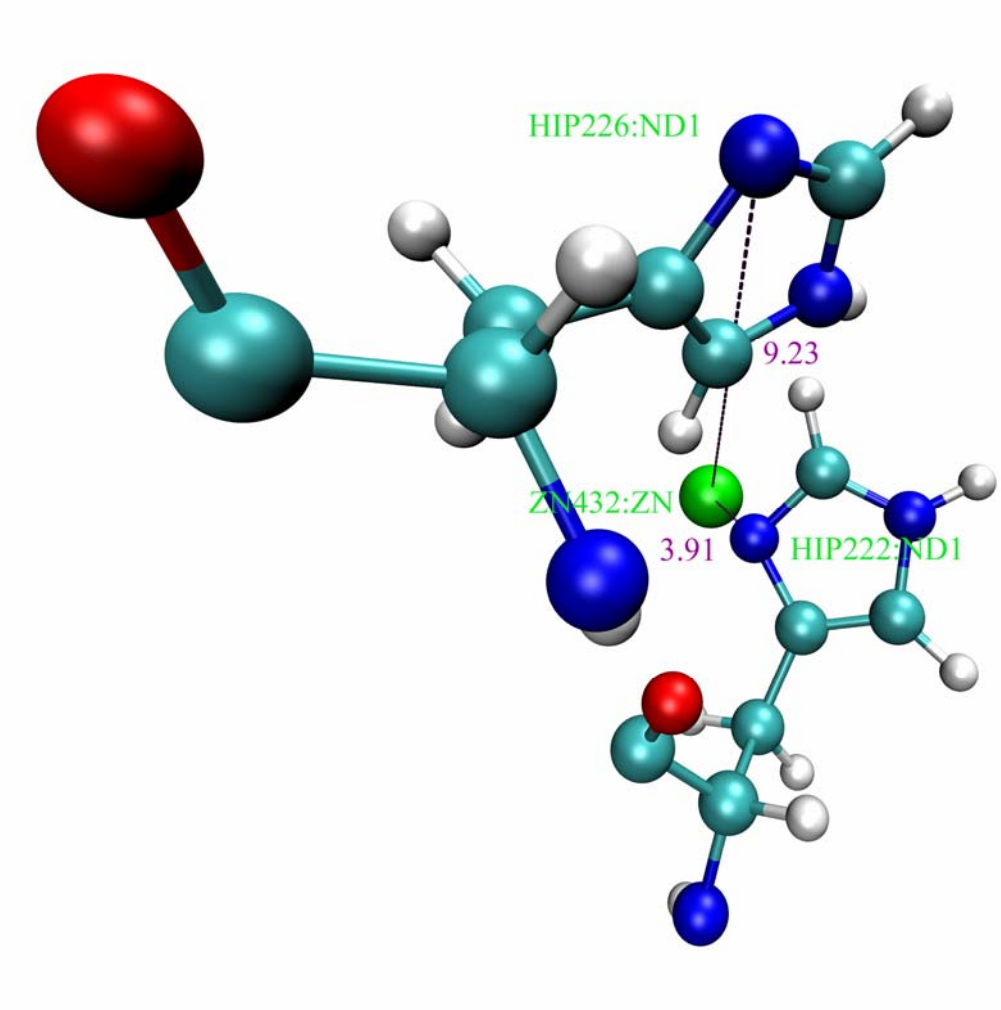

Figure 6-27 Coordination between zinc ion and HIS222 and HIS226 at pH 7.0, 37 °C

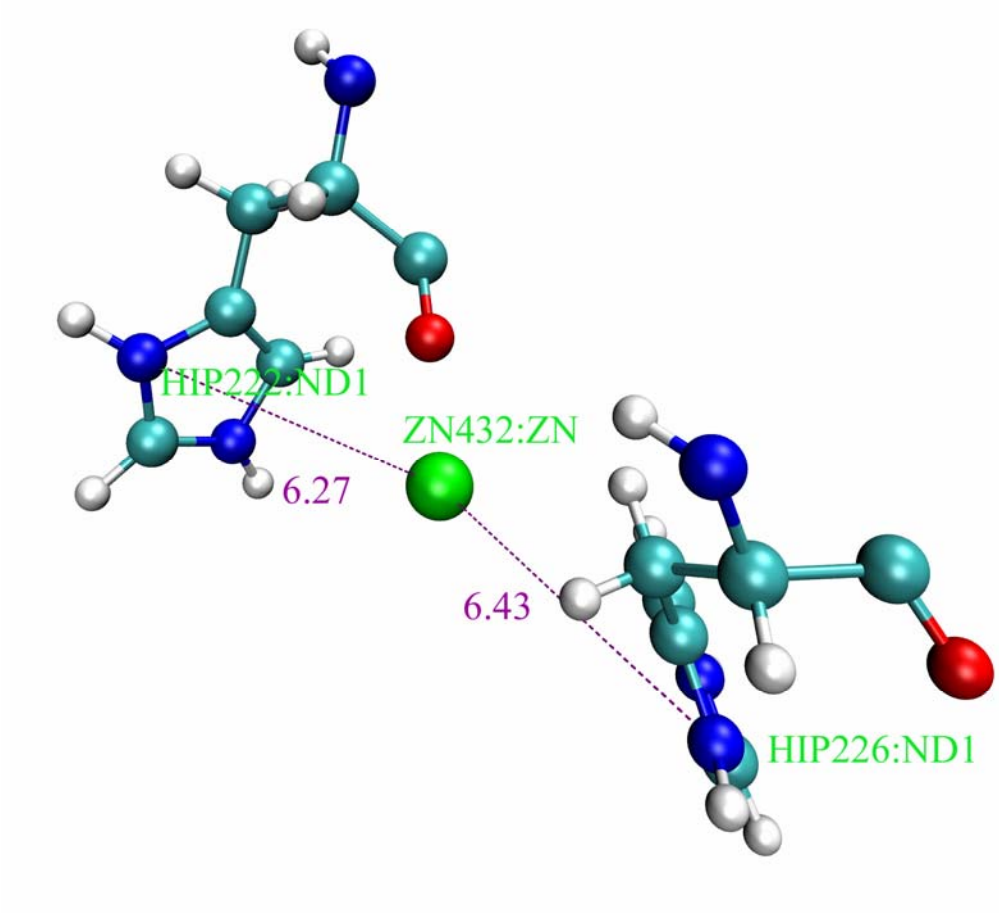

Figure 6-28 Coordination between zinc ion and HIS222 and HIS226 at pH 4.7, 55 °C

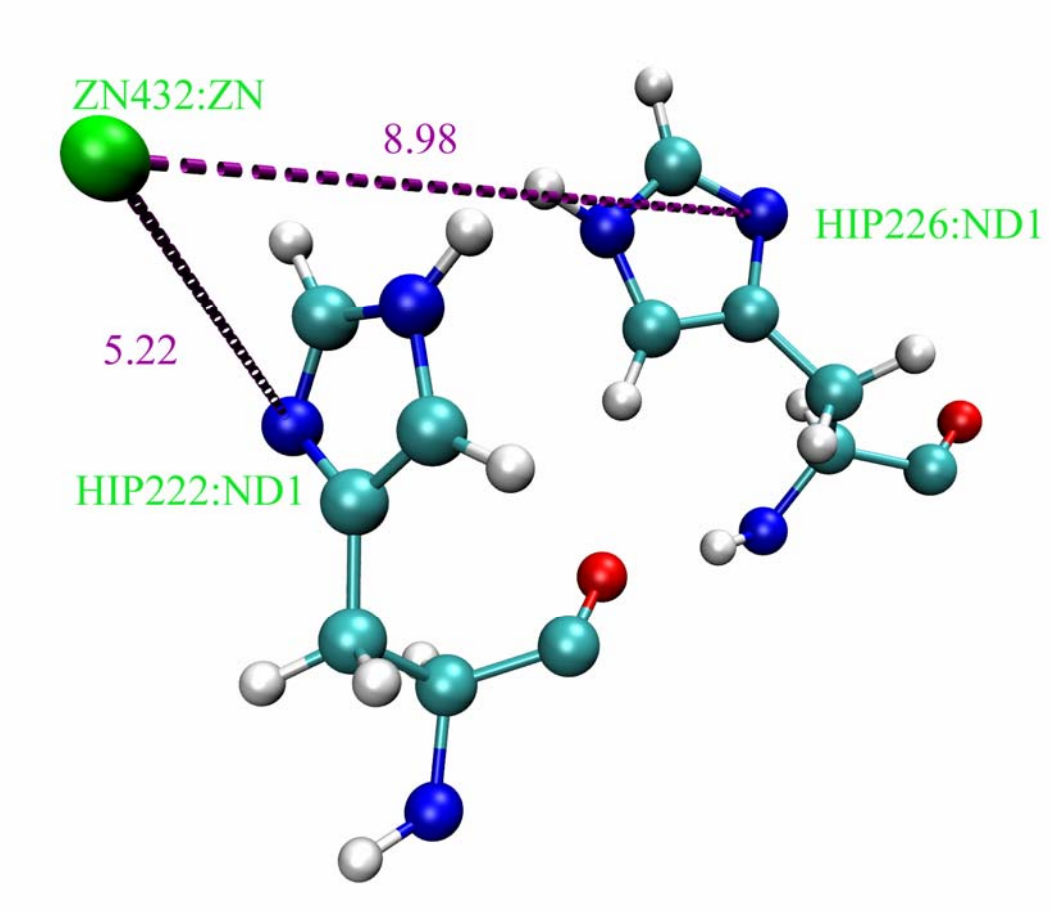

Figure 6-29 Coordination between zinc ion and HIS222 and HIS226 at pH 7.0, 55 °C

Chapter 7 Simulation of PDZ Domains

7.1 Introduction

The PDZ domain is a common structural domain containing about 80 to 90 amino acids which is found in the signaling proteins of bacteria, yeast, plants and animals[77]. PDZ is an acronym combining the first letters of the three proteins which were first discovered to share the domain: post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (DlgA), and zonula occuldens-1 protein (zo-1). PDZ domains are also referred to as DHR (Dlg homologous region) or GLGF (glycine-leucine-glycine-phenylalanine) domains.

Generally, signal transduction in biology means any process by which a cell converts one kind of signal or stimulus into another. Signaling transduction pathways provide critical cell-cell communications which are required to coordinate the activities of vast numbers of cells in every animal's life. Regulations of signaling is crucial, inappropriate activity from a given signal transduction pathway can cause devastating results. Many disease processes such as diabetes, heart disease, autoimmunity and cancer arise from defects in signal transduction pathways, which further indicate the critical importance of

signal transduction pathways.

The Wnt signaling pathway is a major route by which the cell conveys information from its exterior to the nucleus[78]. It describes a complex network of proteins most well known for their roles in embryogenesis and cancer. The name Wnt came from a combination of Wg (wingless) and Int[8]. The wingless gene had originally been identified as a segment polarity gene in Drosophila melanogaster[79]. The INT genes were originally identified as vertebrate genes near several integration sites of mouse mammary tumor virus[80]. The Wnt signaling pathway is believed to be involved in embryonic and postembryonic development as well as in tumorigenesis[81-83].

The canonical Wnt pathway describes a series of events that occur when Wnt proteins bind to cell-surface receptor of the Frizzled family, causing the receptor to activate the Dishevelled (Dvl) family protein and ultimately resulting in a change in the amount of beta-catenin (a subunit of the cadherin protein complex) that reaches the nucleus. It is associated with cancers, body axis specification and morphogenic signaling, etc. Non-canonical Wnt signaling is associated with other activities, such as planar cell polarity, axon guidance, stem cells, etc. Dishevelled is a key component of a membrane-associated Wnt

receptor complex. It relays the Wnt signals from a membrane-bound receptor to downstream components and thereby plays an essential role in this signaling. The Dvl proteins are composed of an N-terminal DIX domain, a central PDZ motif, and a C-terminal DEP domain[84]. Of the three, the PDZ domain is believed to play an essential role in both the canonical and the noncanonical Wnt signaling pathways.

The PDZ domain we investigated is of mouse Dvl-1 (mDvl1) residues 247-341[85]. It is shown in Figure 7-1

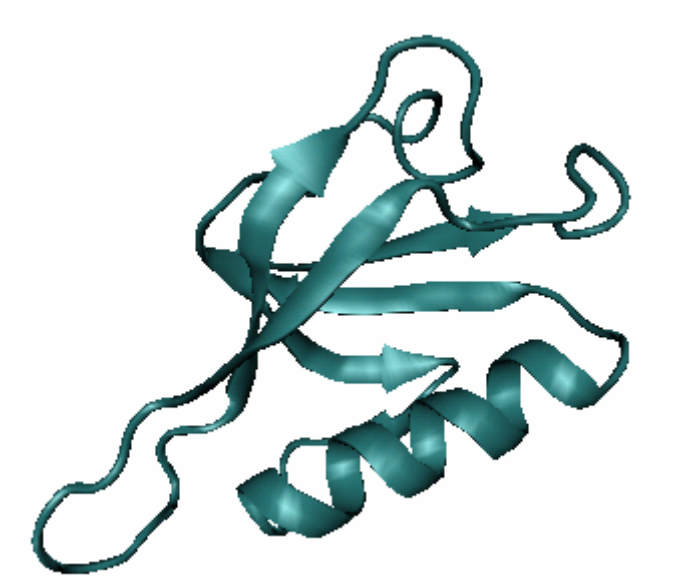

Figure 7-1 The PDZ domain

Because the structure of the Dvl PDZ domain is well studied, it's possible

to use structure-based virtual ligand screening to access potential ligands. Jufang Shan et al. have found an organic compound, NSC668036, which binds to the Dvl PDZ domain[85]. They have used MD simulation to study the binding between them in detail. But their simulations were limited to 5 ns, and our extended simulation of 20 ns has revealed some interesting phenomena after the initial 5 ns period.

The reason we study the PDZ domain is that an inhibitor of the Dvl PDZ domain is likely to effectively block the Wnt signaling pathway at the Dvl level, thus making it an ideal pharmaceutical target. Small organic inhibitors of the PDZ domain in the Dvl might be useful in dissecting molecular mechanisms and formulating agents that target cancers or other disease in which Wnt signaling in involved. We simulated the interaction between it and the PDZ domain with molecular dynamics.

7.2 Simulation Procedure

Charge assignment for the ligand was done using the Gaussian program[86]. The MD simulations and minimizations were done using the sander program from in Amber version 8 using the e16 modified parm99 force

field. The systems were minimized using an eight step approach: (1) Steepest descent minimization restraining the complex with a weight (force constant) of 5 kcal/mol $Å^{-2}$. (2) NPT (constant number of molecules, pressure and temperature) MD equilibration of 15 picoseconds at 1bar and 300K while restraining the complex with a weight of 5 kcal/mol $Å^{-2}$. (3) Steepest descent minimization restraining the complex with a weight of 2 kcal/mol $Å^{-2}$. (4) An MD equilibration of NPT of 1.5 picoseconds at 1bar and 300K while restraining the complex with a weight of 5 kcal/mol $Å^{-2}$. (5) NPT MD equilibration of 5 picoseconds at 1bar and 300K while restraining the complex with a weight of 1 kcal/mol $Å^{-2}$. (6) NPT MD equilibration of 5 picoseconds at 1bar and 300K while restraining the complex with a weight of 0.5 kcal/mol $Å^{-2}$. (7) NPT MD equilibration of 10 picoseconds at 1bar and 300K while restraining all the carbon (including backbone alpha carbons) and nitrogen atoms with a weight of 0.5 kcal/mol $Å^{-2}$. (8) NPT MD equilibration of 5 picoseconds at 1 bar and 300K with no restraints.

All MD simulations were done with a time step of 1 femtosecond and a cutoff of 8 Å for non-bonded force. After minimization, the actual MD simulations were performed by NAMD with a time step of 2 femtoseconds and

the nonbonded cutoff set to 9.0 Å at a temperature of 300K, 1 atm of constant pressure with PME.

7.3 Results and Analysis

The simulation results were analyzed by calculating the root-mean-square deviations (RMSDs) first. Only the backbone of PDZ domain and the NSC668036 were used to calculate the RMSD. The back bone of NSC668036 was defined as the 13 atoms in the main chain between and including the carbonyl carbon of the carboxylate group and the carbonyl carbon at the other end of NSC668036, as shown in Figure 7-2.

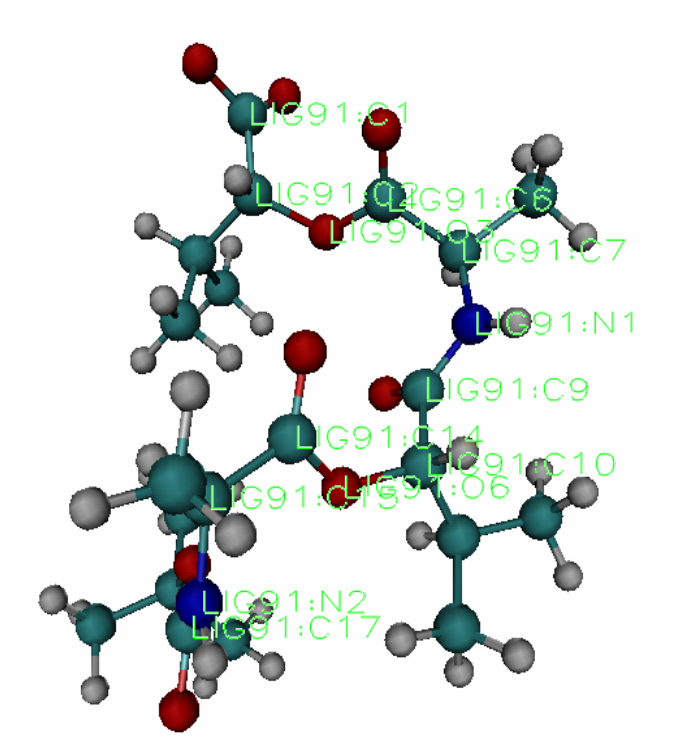

Figure 7-2 Backbone of NSC668036

We simulated all the 30 conformational structures of NSC668036 provided by Shan et al after we docked them to the PDZ domain. Figure 7-3 shows the structure of the PDZ domain with one NSC668036 molecule. The RMSD of the PDZ domain and NSC668036 for the 20 ns simulations are shown in the supplementary material.

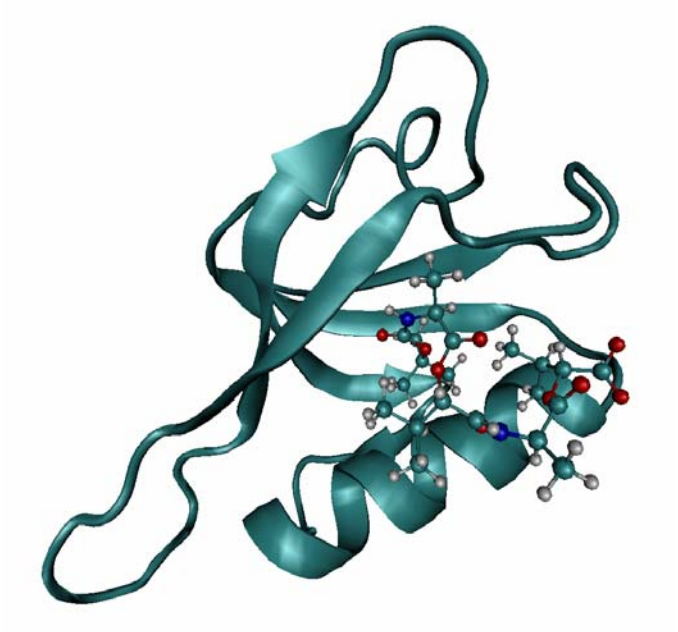

Figure 7-3 The PDZ domain with NSC668036

From these graphs, we can figure out that some of the structures are stable after the docking process, while some others are not. For example, if we take a close look at structure #10, which is shown in Figure 7-4, we can find that the ligand remains rather stable after the docking, which is indicated by a low and stable RMSD of it.

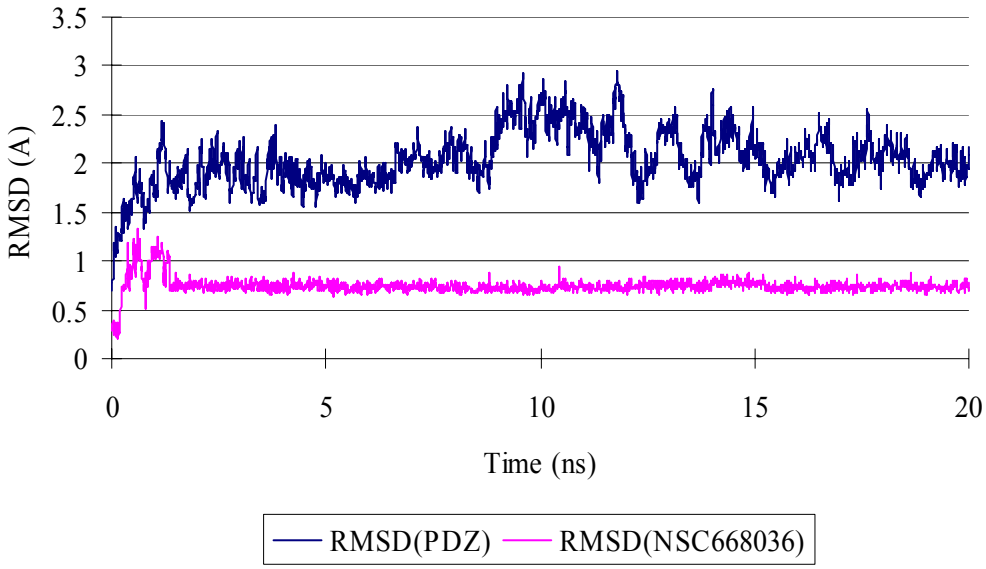

Figure 7-4 RMSD of PDZ domain and ligand #10

To study the convergence of final states of the PDZ domain after the 20 ns simulation, we calculated the RMSD between each structure, and matrix was made according to the results in Figure 7-5. In this figure, the X and Y axes stand for the 30 different structures, and the color represents the value of the RMSD between the 2 structures labeled in X and Y axes.

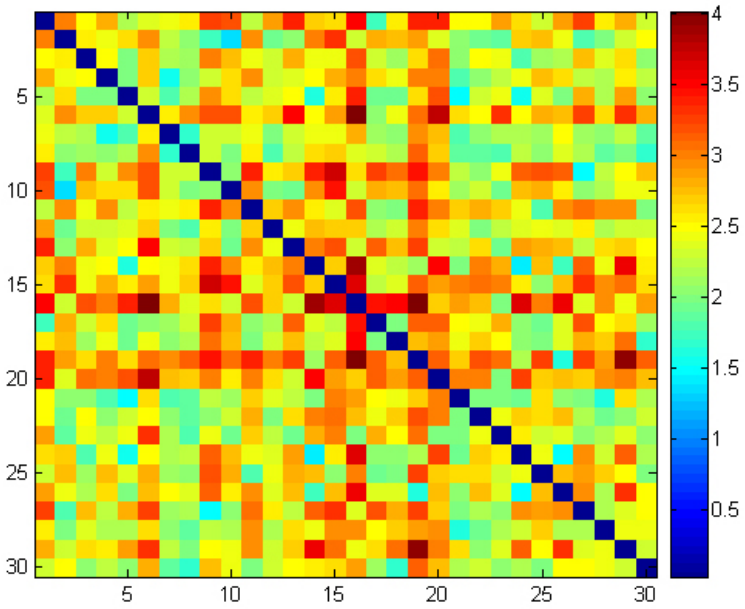

Figure 7-5 Final state of PDZ domain after 20 ns simulation

Since the structures have zero RMSD when comparing with itself, the diagonal line of the matrix is zero. An island with small numbers in the matrix means that the relevant structures have similar conformation after the 20 ns simulation, and tend to converge to a similar structure, like the structure #8, #22, #27. On the contrary, an island with large numbers in the matrix means that this structure does not converge well to other structures after the simulation, like structure #16, #19, #28. A more accurate numerical analysis shows that the following group structures tend to converge after the simulation: group of

structure #3, 4, 5, 7; group of structure #2, 9 ,10, 12; group of structure #5, 14, 17, 18, 21, group of structure #7, 8, 30.

To understand the time evolvement of the convergence, we also made the matrix at time 10ns. A series of the matrices clearly showed the forming of small number islands in the matrix. From Figure 7-6, the matrix at 10 ns, we can see that the number of “red-island” is significant less than that at 20 ns, which indicates that these 30 structures tends to evolve into several groups, the differences between the groups become large as time increases. On the other hand, the area of yellow part in Figure 7-6 are larger than Figure 7-5, which suggests that those similar members within the groups evolve into more similar structure as time increases.

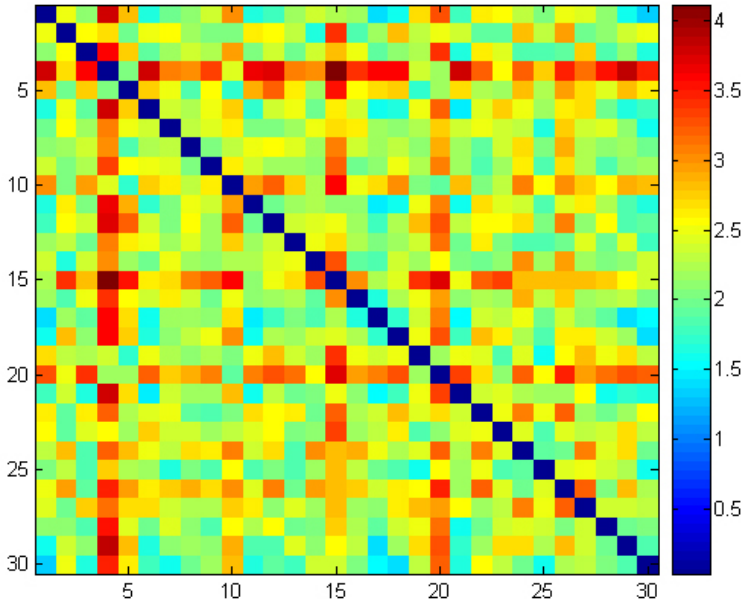

Figure 7-6 Final state of PDZ domain after 10 ns simulation

7.4 Conclusion

We extended the simulation by J. Shan et al [85] from 5 ns to 20 ns, and found many interesting properties of docking process between the PDZ domain and the NSC668036 molecule. Some RMSD graphs of the PDZ domain and NSC668036 for the 20 ns simulations, like structure #15, #25 and #30, shows that there may be dramatic changes for the RMSD value after the 5 ns period. It indicates that the docking process may needs a longer time to achieve the

equilibrium state, thus, in the study of docking problem using molecular dynamics, a longer simulation time may be necessary.

Chapter 8 Summary and Future Work

8.1 Summary and Contributions

Using MD to simulate proteins has a long history, but most of the all-atom simulations with explicit solvent were performed on relatively small proteins with less than 100 residues. All-atom computational simulations with explicit solvent of large proteins with ~1,000 residues like BoNT/A for more than 100 ns are very rare, and we are the first group simulating BoNT/A at this space and time scale. These explicit-solvent simulations are complement to the implicit-solvent simulations, which have been carried out for times of order 10s' of ns and do not require so much computer time, but also do not include so much of the details of the molecular interactions.

In this thesis, we have focused mainly on the simulation of BoNT/A at different pH values and temperatures. Two different methods were applied to modeling the BoNT/A at various pH values. Although large conformational changes usually take times of the order of microsecond to occur, our simulations of 200 nanoseconds appear to have shown some interesting results that are consistent with laboratory experiments. Certainly, more structural changes will

emerge if and when one can simulate such proteins for a longer time scale.

In the simulations of the PDZ domain with the small molecule NSC668036, we found many interesting properties of the docking process between them. Comparing to regular Monte Carlo method, Molecular Dynamics also may provide useful information in the study of docking problems.

8.2 Future Work

In following up the present work, there are several future directions worth investigating:

1. Folding and unfolding processes for large system like BoNT/A take time scales much larger than nanoseconds or even microseconds. But due to the limitation of computing power we can acquire, we can only simulate it with explicit solvent for a relatively short time. With the rapidly increasing computing power available, we can simulate it for a much longer time in the future to understand its folding and unfolding processes.
2. As we discussed in Chapter 5, how the light chain of BoNT/A passes across the endosomal membrane and enters into the cytosol

remains mysterious. In future studies, simulations of BoNT/A with the endosomal membrane may be considered, which may help to understand the mechanism from the first principles.

3. All-atom force fields can give the best descriptions of the system we studied. However, the computation time it takes to calculate the interactions with all atoms included is by no means an easy task. For some systems, a coarse-grained model can also provide a good approximation. In future studies on relatively large systems like BoNT/A, an implicit-solvent simulation or hybrid method may also be considered.

References

1. J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**: p. 585-590.
2. J. M. Haile, *Molecular dynamics simulation: elementary methods*. 1992, New York: Wiley.
3. R. Zhou S. J. Stuart, and B. J. Berne, *Molecular dynamics with multiple time scales: The selection of efficient reference system propagators.* J. Chem. Phys., 1996. **105**(4): p. 1426-1436.
4. Bojan Zagrovic, Eric J. Sorin, and Vijay Pande, *beta-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model* Journal of Molecular Biology, 2001. **313**: p. 151-169.
5. Jed W. Pitera and William Swope, *Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins.* Proceedings of the National Academy of Sciences, 2003. **100**(13): p. 7587-7592.
6. Bojan Zagrovic, et al., *Simulation of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwidedistributed Computing.* Journal of Molecular Biology, 2002. **323**: p. 927-937.
7. D.B. Lacy and R.C. Steven, *Recombinant Expression and Purification of the Botulinum Neurotoxin Type A Translocation Domain.* Protein expression and purification, 1997. **11**: p. 195-200.
8. F Rijsewijk, et al., *The Drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless.* Cell, 1987. **50**(4): p. 649-657.
9. Xin Chen and Yuefan Deng, *Long-time molecular dynamics simulations of botulinum biotoxin type-A at different pH values and temperatures.* J. of Molecular Modeling, 2007. **13**(5): p. 559-572.
10. Yongzhi Chen, Xin Chen, and Yuefan Deng, *Simulating botulinum neurotoxin with constant pH molecular dynamics in Generalized Born implicit solvent* Computer Physics Communications, 2007. **177**: p. 210-213.
11. Peter L. Freddolino, et al., *Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus.* Structure, 2006. **14**(3): p. 437-449.

12. L Kalé, et al., *NAMD2: Greater scalability for parallel molecular dynamics.* J. of Computational Physics, 1999. **151**: p. 283-312.

13. James C. Phillips, et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**: p. 1781-1802.

14. Guha Jayachandran, V. Vishal, and V. S. Pande, *Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece.* Journal of Chemical Physics, 2006. **124**: p. 164902.

15. Michael R. Shirts and Vijay Pande, *Screen savers of the world, Unite!* . Science, 2000. **290**: p. 1903-1904.

16. Stefan M. Larson, et al., *Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology.* Computational Genomics, 2002.

17. J. Aichelin, et al., *Quantum molecular dynamics approach to heavy ion collisions: Description of the model, comparison with fragmentation data, and the mechanism of fragment formation.* Phys. Rev. C, 1988. **37**(6): p. 2451-2468.

18. J. Aichelin and H. Stöcker, *Quantum molecular dynamics — A novel approach to N-body correlations in heavy ion collisions.* Physics Letters B, 1986. **176**(1-2): p. 14-19.

19. Folkmar A. Bornemann, Peter Nettesheim, and Christof Schütte, *Quantum-classical molecular dynamics as an approximation to full quantum dynamics.* The Journal of Chemical Physics, 1996. **105**(3): p. 1074-1083.

20. J. E. Lennard-Jones, *Cohesion.* Proceedings of the Physical Society, 1931. **43**(5): p. 461-482.

21. J.W. Ponder and D.A. Case, *Force fields for protein simulations.* Adv. Prot. Chem., 2003. **66**: p. 27-85.

22. BR Brooks, et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* J Comp Chem, 1983. **4**: p. 187-217.

23. A. D. Mackerell, et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J. Phys. Chem. B, 1998. **102**: p. 3586-3616.

24. Walter R. P. Scott, et al., *The GROMOS Biomolecular Simulation Program Package.* J. Phys. Chem. A, 1999. **103**(19): p. 3596-3607.

25. Erik Lindahl, Berk Hess, and David van der Spoel, *GROMACS 3.0: a package for*

molecular simulation and trajectory analysis Journal of Molecular Modeling, 2001. **7**(8): p. 306-317.

26. R. S. Germain, et al., *Early performance data on the Blue Matter molecular simulation framework.* IBM J. Res. & Dev, 2005. **49**(2/3): p. 447-455.

27. T Darden, D York, and L Pedersen, *Particle mesh Ewald: an N log (N) method for Ewald sums in large systems.* J. Chem. Phys, 1993. **98**(12): p. 10089-10092.

28. Henrik G. Petersen, *Accuracy and efficiency of the particle mesh Ewald method.* J. Chem. Phys., 1995. **103**(9): p. 3668-3679.

29. R.W. Hockney and J. W. Eastwoo, *Computer simulation using particles*. 1988, Bristol, PA: Taylor & Francis, Inc.

30. A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, *Chain Length Scaling of Protein Folding Time.* Phys. Rev. Lett., 1996. **77**(27): p. 5433-5436.

31. Ji-Tao Huang and Jing Tian, *Amino acid sequence predicts folding rate for middle-size two-state proteins.* Proteins: Structure, Function, and Bioinformatics, 2006. **63**(3): p. 551-554.

32. Blake Gillespie and Kevin W. Plaxco, *Using protein folding rates to test protein folding theories.* Annual Review of Biochemistry, 2004. **73**(1): p. 837-859.

33. Jan Kubelka, James Hofrichter, and William A Eaton, *The protein folding 'speed limit'.* Current Opinion in Structural Biology, 2004. **14**(1): p. 76-88.

34. KA Dill, *Theory for the folding and stability of globular proteins.* Biochemistry, 1985. **24**(6): p. 1501-1509.

35. KF Lau and KA Dill, *A lattice statistical mechanics model of the conformational and sequence spaces of proteins.* Macromolecules, 1989. **22**(10): p. 3986-3997.

36. Ken A. Dill, *Dominant forces in protein folding.* Biochemistry, 1990. **29**(31): p. 7133-7155.

37. Ron Unger and John Moult, *Genetic Algorithms for Protein Folding Simulations.* Journal of Molecular Biology 1993. **231**(1): p. 75-81.

38. Michael Feig and Charles L Brooks III, *Recent advances in the development and application of implicit solvent models in biomolecule simulations.* Current Opinion in Structural Biology, 2004. **14**(2): p. 217-224.

39. Nathan A Baker, *Improving implicit solvent simulations: a Poisson-centric view.* Current Opinion in Structural Biology, 2005. **15**(2): p. 137-143.

40. Michael S. Lee, Freddie R. Salsbury Jr., and Mark A. Olson, *An efficient hybrid explicit/implicit solvent method for biomolecular simulations.* Journal of Computational Chemistry, 2004. **25**(16): p. 1967-1978.

41. D. Bashford and D.A. Case, *Generalized Born models of macromolecular solvation effects.* Annu. Rev. Phys. Chem., 2000. **51**(1): p. 129-152.

42. Vickie Tsui and David A. Case, *Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model.* J. Am. Chem. Soc., 2000. **122**(11): p. 2489-2498.

43. Vickie Tsui and David A. Case, *Theory and applications of the generalized born solvation model in macromolecular simulations.* Biopolymers, 2001. **56**(4): p. 275-291.

44. Alexey Onufriev, David A. Case, and Donald Bashford, *Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect.* J. of Computational Chemistry, 2002. **23**: p. 1297-1304.

45. W.C. Still, et al., *A General Treatment of Solvation for Molecular Mechanics.* J. Am. Chem. Soc., 1990. **112**: p. 6127.

46. Alexey Onufriev, Donald Bashford, and David A. Case, *Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model.* Proteins: Structure, Function, and Bioinformatics, 2004. **55**: p. 383-394.

47. Avijit Ghosh, Chaya Sendrovic Rapp, and Richard A. Friesner, *Generalized Born Model Based on a Surface Integral Formulation.* Journal of Physical Chemistry, 1998. **102**(52): p. 10983-10990.

48. Leonid Mirny and Eugene Shakhnovich, *Protein folding theory: from lattice to all-atom models.* Annual Review of Biophysics and Biomolecular Structure, 2001. **30**(1): p. 361-396.

49. William L. Jorgensen and Julian Tirado-Rives, *Potential energy functions for atomic-level simulations of water and organic and biomolecular systems.* Proceedings of the National Academy of Science, 2005. **102**(19): p. 6665-6670.

50. William L. Jorgensen, et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

51. William L. Jorgensen, *Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water.* J. Am. Chem. Soc., 1981. **103**(2): p. 335-340.

52. Y Deng, A Korobka, and B Xiang, *Nankai Stars: an example of designing, constructing, evaluating, and applying a 5-Tflops Beowulf supercomputer.* Int. J. High Perf. Comp.

App., Accepted.

53. D.A. Pearlman, et al., *AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules.* Comp. Phys. Commun., 1995. **91**: p. 1-41.

54. D. A.; Case, et al., *Amber 8 Users' Manual.* 2004.

55. W. Humphrey, A. Dalke, and K. Schulten, *VMD - Visual Molecular Dynamics.* J. Molec. Graphics, 1996. **14**: p. 33-38.

56. B R Singh, *Intimate details of the most poisonous poison.* Nature Structural Biology, 2000. **7**: p. 617-619.

57. SS Arnon, et al., *Botulinum toxin as a biological weapon - Medical and public health management.* J. of the American Medical Association, 2001. **285**(8): p. 1059-1070.

58. HD Shukla and SK Sharma, *Clostridium botulinum: A bug with beauty and weapon.* Critical Reviews in Microbiology, 2005. **31**(1): p. 11-18.

59. L Silveira-Moriyama, et al., *Long-term efficacy of botulinum toxin A in the treatment of blepharospasm over a 10-year period.* Movement Disorders, 2004. **19**: p. S98-S99.

60. SH Han, et al., *Effect of botulinum toxin A chemodenervation in sensory strabismus.* J. of Pediatric Ophthalmology & Strabismus, 2001. **38**(2): p. 68-71.

61. AV Benedetto, *The cosmetic uses of Botulinum toxin type A.* International J. of Dermatology, 1999. **38**(9): p. 641-655.

62. A Brashear, et al., *Longitudinal assessment of the dose consistency of botulinum toxin type A (BOTOX (R)) for cervical dystonia.* Advances in Therapy, 2005. **22**(1): p. 49-55.

63. JF Damrose, et al., *The impact of long-term botulinum toxin injections on symptom severity in patients with spasmodic dysphonia.* J. of Voice, 2004. **18**(3): p. 415-422.

64. M Porta and G Maggioni, *Botulinum toxin (BoNT) and back pain.* J. of Neurology, 2004. **251**: p. 15-18.

65. L. Li and B R Singh, *Spectroscopic Analysis of pH-Induced Changes in the Molecular Features of TypeA Botulinum Neurotoxin Light Chain.* Biochemistry, 2000. **39**: p. 6466-6474.

66. S Swaminathan and S Eswaramoorthy, Nature Structural Biology, 2000. **7**: p. 693-699.

67. D. Borden Lacy, et al., *Crystal structure of botulinum neurotoxin type A and implications for toxicity.* Nature Structural Biology, 1998. **5**: p. 898-902.

68. John Mongan, David A. Case, and J. Andrew McCammon, *Constant pH Molecular Dynamics in Generalized Born Implicit Solvent.* J Comput Chem, 2004. **25**: p. 2038-2048.

69. H.M. Berman, et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**: p. 235-242.

70. S. A. Ahmed, J. Protein Chem, 2001. **20**: p. 221-231.

71. F.J. Lebeda and M.A. Olson, Proteins: Structure, Function, and Genetics, 1994. **20**: p. 293-300.

72. Li Li, et al., *Probing the Mechanistic Role of Glutamate Residue in the Zinc-Binding Motif of Type A Botulinum Neurotoxin Light Chain.* Biochemistry, 2000. **39**(9): p. 2399-2405.

73. G Schiavo, et al., *Botulinum neurotoxins are zinc proteins.* J. Biol. Chem, 1992. **267**(33): p. 23479-23483.

74. N. M Hooper, *Families of zinc metalloproteases.* FEBS Lett., 1994. **354**(1): p. 1-6.

75. B. L. Vallee and D. S Auld, *Zinc coordination, function, and structure of zinc enzymes and other proteins.* Biochemistry, 1990. **29**: p. 5647-5659.

76. Subramaniam Eswaramoorthy, et al., *Role of Metals in the Biological Activity of Clostridium botulinum Neurotoxins.* Biochemistry, 2004. **43**(8): p. 2209-2216.

77. CP PONTING, *Evidence for PDZ domains in bacteria, yeast, and plants.* Protein Science, 1997. **6**(2): p. 464-468.

78. Sean D. Speesea and Vivian Budnik, *Wnts: up-and-coming at the synapse.* Trends in Neurosciences, 2007. **30**(6): p. 268-275.

79. C Nüsslein-Volhard and E. Wieschaus, *Mutations affecting segment number and polarity in Drosophila.* Nature, 1980. **287**(5785): p. 795-801.

80. R Nusse, et al., *Mode of proviral activation of a putative mammary oncogene (int-1) on mouse chromosome 15.* Nature, 1984. **307**(5947): p. 131-136.

81. P. Polakis, *Wnt signaling and cancer.* Genes. Dev, 2000. **14**: p. 1837-1851.

82. R.T. Moon, et al., *The promise and perils of Wnt signaling through β-catenin.* Science, 2002. **296**: p. 1644-1646.

83. A. Wodarz and R. Nusse, *Mechanisms of Wnt signaling in development.* Annu. Rev. Cell Dev. Biol., 1998. **14**: p. 59-88.

84. H.C. Wong, et al., *Stuctural basis of the recognition of the dishevelled DEP domain in the Wnt signaling pathway.* Nat. Struc. Biol., 2000. **7**: p. 1178-1184.

85. Jufang Shan, et al., *Identification of a Specific Inhibitor of the Dishevelled PDZ Domain.* Biochemistry, 2005. **44**(47): p. 15495-15503.

86. M. J. Frisch, et al., *Gaussian 03*. 2004: Gaussian, Inc., Wallingford CT.