INVESTIGATION OF DYNAMIC MECHANISM OF THE TOPRIM & 5Y-CAP DOMAINS OF A TYPE IIA TOPOISOMERASE

by

Süleyman ÜÇÜNCÜOĞLU

June   2009

INVESTIGATION OF DYNAMIC MECHANISM OF THE TOPRIM & 5Y-CAP DOMAINS OF A TYPE IIA TOPOISOMERASE

by

Süleyman ÜÇÜNCÜOĞLU

A thesis submitted to

the Graduate Institute of Sciences and Engineering

of

Fatih University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Physics

June 2009
Istanbul, Turkey

APPROVAL PAGE

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

<u>Prof. Dr. Mustafa KUMRU</u>
Head of Department

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

<u>Assist. Prof. Dr. Levent SARI</u>
Supervisor

Examining Committee Members

Assist. Prof. Dr. Levent SARI   …………………………………………..

Assist. Prof. Dr. Sedat ÇOŞGUN   …………………………………………..

Assist. Prof. Dr. Sadık GÜNER   …………………………………………..

It is approved that this thesis has been written in compliance with the formatting rules laid down by the Graduate Institute of Sciences and Engineering.

<u>Assoc. Prof. Dr. Nurullah ARSLAN</u>

Director

June 2009

INVESTIGATION OF DYNAMIC MECHANISM OF THE TOPRIM & 5Y-CAP DOMAINS OF A TYPE IIA TOPOISOMERASE

Süleyman ÜÇÜNCÜOĞLU

M. S. Thesis - Physics
June 2009

Supervisor: Assist. Prof. Dr. Levent SARI

ABSTRACT

DNA Topoisomerases are wondrous proteins that plays vital role in cellular activities. The double helical structure of DNA lies behind primal requirement for the activity of topoisomerases. Different types of Topoisomerases solve the topological problems that occur during the transcription and replication of DNA. Topoisomerases included in Type II cut two strands of one DNA double helix, other intact DNA strand is passed through it, and then relegate the scissile strand back.

In this study, we have investigated the large scale domain motions of the enzyme, employing a high level biased MD approach. Initial atomic coordinates were obtained from the crystal structure of E.coli DNA GyraseA (pdb code 1ab4). The main goal of the study is to get a stable dimer form of the protein that has not been obtained experimentally, and dynamically simulate the opening and closing motions of the gate in the upper and lower parts of the protein.

We have also tested the stability of the BIOMT structure in the solution. By simulating the mechanism in real time, we have found out that the V-shape structure of the protein, where only lower gate is closed, is found to be stable. This observation supports the current proposed mechanism of supercoiling relaxation.

Also, in all different simulation set-ups, we have observed that the closure of the lower gate is more favorable that the closure of the upper gate. The completely closed form of the protein is found to be not stable. It is also important to note that our simulations show us the most mobile secondary structures to be; $\beta1$, $\beta17$, $\beta20$, $\alpha14$, $\alpha14^1$, and $\alpha15$.

TİP IIA TOPOİZOMERAZLARININ TOPRIM VE 5Y-CAP KISIMLARININ DİNAMİK MEKANİZMALARININ ARAŞTIRILMASI

Süleyman ÜÇÜNCÜOĞLU

Yüksek Lisans Tezi - Fizik
Haziran 2009

Tez Yöneticisi: Yrd.Doç. Dr. Levent SARI

ÖZ

Topoizomerazlar hücresel aktivitelerde hayati rol oynayan harikulade proteinlerdir. DNA'nın çift sarmal yapısı, topoizomerazların aktivitelerinin gerekliliğini ortaya koyan yegane sebeptir. Değişik tipteki topoizomerazlar, DNA bölünmesi ve açılması sırasında ortaya çıkan topolojik problemleri çözerler. Tip IIA grubuna dahil olan topoizomerazlar, DNA'nın çift sarmalının her ikisini birden keser, bu oluşan kesikten diğer DNA sarmalını geçirir ve ardından kesik DNA'yı yeniden birleştirir.

Bu çalışmamızda, enzimin büyük çaptaki hareketlerini, yüksek seviye uyarlamalı moleküler dinamik yaklaşımı kullanarak araştırılmıştır. E.coli DNA Cayraz A proteinindeki atomların başlangıç koordinatları, daha önce kristalize edilmiş ve literatüre aktarılmış olan yapıdan okunmuştur. (Protein veritabanı kodu 1ab4). Bu çalışmadaki temel amaç, hala deneysel olarak stabil şekilde elde edilememiş olan proteinin karşılıklı zincir yapısını, proteindeki alt ve üst kapıları dinamik bir şekilde açmak ve kapatmak suretiyle elde etmek.

Bu çalışmamızda aynı zamanda, literatürde proteinin karşılıklı yapısı olarak öngörülen BIOMT yapının doğruluğunu test ettik. Mekanizmayı, gerçek zamanlı olarak yaptığımız simülasyonlarda, V-şeklindeki protein yapısının, üstün

açık bırakılıp sadece alt kısmın kapandığı durumlarda, daha kararlı olduğunu gözlemledik. Bu sonuç, önerilen süper sarmalın gevşeme mekanizmasının doğruluğunu desteklemektedir.

Ayrıca, değişik simülasyon düzeneklerinde, alt kısımdaki kapak açılmasının, üst kısmın açılmasına nazaran daha tercih edilebilir olduğunu gözlemledik. Öte yandan, proteinin hem alt kısım hem de üst kısmının aynı anda kapanmasının kararlı bir sonuç vermediğini gözlemledik. Bütün simülasyon düzeneklerinde, bazı ikincil yapıların, β1, β17, β20, α14, α14[1], ve α15 daha hareketli olduğunu gözlemledik.

Anahtar Kelimeler: DNA topolojisi, süpersarmal, Moleküler Dinamik Simülasyonlar, HQBMD.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOL/ABBREVIATION

DNA   :   Deoxyribonucleic acid

ccDNA   :   Closed circular DNA

topo   :   Topoisomerase

E-coli   :   Escherichia coli

MD   :   Molecular Dynamics

HQBMD   :   High Quality Biased Molecular Dynamics

Lk   :   Linking number

Lk_0 :   Linking number in relaxed state

Tw   :   Twist

Wr   :   Writhe

ΔLk :   Linking difference

σ :   Supercoiling density or specific linking difference

T-Segment: Transported Fragment

G-Segment: Gate Fragment

$\overset{0}{A}$:   Angstrom

NMR   :   Nuclear Magnetic Resonance

k_b :   Force constant for bond interaction energy

k_θ :   Force constant for bond angle potential energy

k_Φ :   Force constant for dihedral angle potential energy

k_ω :   Force constant for bond improper potential energy

k_u :   Force constant for Urey-Bradley potential energy

ε_0 :   Depth of the minimum energy for Lennard-Jones potential

ε :   Dielectric constant

α_{hqbmd} :   Force constant for HQBM

ρ	:	Reaction coordinate
V	:	Volume
T	:	Temperature
N	:	Number of particles
CHARMM:		Chemistry at HARvard Macromolecular Mechanics
VMD	:	Visual Molecular Dynamics
LSF	:	Load Sharing Facility
CMU	:	Cluster Management Utility
PSF	:	Protein Structure File
RTF	:	Residue Topology File
PARAM	:	Parameter File
PSF	:	Protein Structure File
CRD	:	Coordinate File
SD	:	Steepest Descent
ABNR	:	Adapted Basis Newton-Raphson
Mg	:	Magnesium
RMSD	:	Root mean square deviations

CHAPTER I

INTRODUCTION

1.1 DNA MOLECULE

Deoxyribonucleic acid (DNA) is a molecule containing the genetic code, which is used in a vast amount of cellular activity. In a way DNA is the fingerprint of an organism from where all the specific information about its form and function could be deduced. The chemical composition of the DNA is comprised of two long polymers termed as nucleotides. The nucleotides form ester bonds through sugar and phosphate groups to build the backbone of DNA. In living organisms, DNA does not usually exist as a single molecule, but instead as a tightly associated pair of nucleotides [1]. This two long DNA strands usually twist together in shape of a double helix with a width of about 22 Å and a length of 3.3 Å [2]. The two nucleotides unite each other via sugar or phosphate bridges between their purine or pyrimidine bases. The purine bases are adenine and guanine and the pyrimidines are thymine and cytosine. Adenine links to thymine with two hydrogen bonds whereas cytosine links to guanine with three hydrogen bonds.

1.2 DNA TOPOLOGY AND SUPERCOILING

After the discovery of helical DNA structure by Watson and Crick by the years of 60s [1], researches focused on the topology of this miraculous material. The scientists gain a new insight to DNA topology by discovery of a new form of DNA in viruses [3]. The study indicates two fractioned DNA group, whose molecular weight are the same and components belong to double-stranded DNA. However first group have higher

sedimentation and resistant to penetration. They came to result that first group composed of "circular base-paired duplex molecules without chain ends" and second group is just linear type of same DNA molecule [4]. The name "closed" chosen because there is not any break and opening at the ends of helix. Figure 1.1 represents the structure of closed-circular DNA.
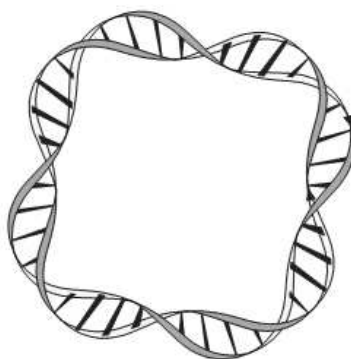
Figure 1.1 Closed-circular DNA structure [4]

The physical properties of linear and the closed-circular DNA are different from each other. DNA could be knotted in form of closed-circular DNA that is the most noticeable distinction from linear DNA form [5].

We can define topology term as spatial properties that stays unchanged under successive deformations. From the definition, topology remains constant during the geometry alterations, which is significant for DNA structure and function. Applying the ribbon theory, Fuller implement investigated the topological properties of closed circular DNA [5]. The theory states that, in addition to topological attribute of a ribbon, there are major quantities that play significant roles to define the topological state of a molecule such as linking number (Lk), the twist (Tw) and writhe (Wr) of ribbon. Tw and Wr are two differential geometric properties depending on geometry; on the other hand, Lk is an explicit topological constant, which does not rely on geometry. There is a mathematical connection between Lk, Tw and Wr, which can be formulated by the following equation:

$$Lk = Tw + Wr \qquad\qquad (1.1)$$

This equation is quite fascinating because as we told earlier Lk does not change with change of geometry but it is also the sum of two variables that they alter with change of geometry. Simply, twist is the number of turns of double helix and writhe is the number of helix crossovers. Lk number in for different shapes and geometries are given in the figure below.

(a) *Lk* = 1

(b) *Lk* = 6
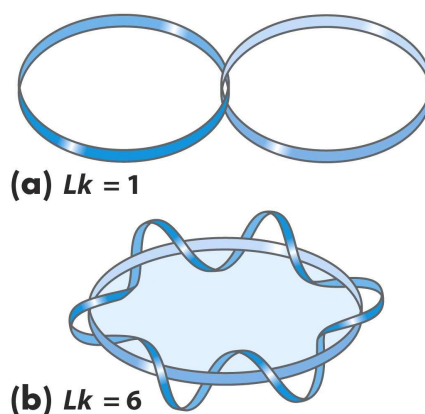
Figure 1.2 Examples of Lk number in different geometries [6].

Scientist has already known the structure of DNA, which is double helix, but what is not known is the natural Lk numbers in not supercoiled the DNA. Then, they have investigated the Lk0 and it is approximately the number of base pairs (N) divided by the periodicity of the helix, nearly 10.5 bp per turn [7].

$$Lk0 \approx N/10.5 \tag{1.2}$$

DNA has to reduce its size to a more compact form in order to fit into the nucleus. Therefore, DNA helices round over themselves to lover the size, which is called as supercoiling.

The length of human DNA is nearly 2 m long and the size of a eukaryotic cell is usually between 5-100 μm [8]. Nucleus is only a small portion of cell that contains DNA molecules. It can be seen clearly that DNA must be packed in order to fit into the nucleus. Supercoiling is a feature of DNA that tremendously decreases the size of DNA

molecule. On the other hand, supercoiled DNA in such a squeezed form has to allow reading of genetic code during cellular activity involving the strand separation such as transcription and replication.

To illustrate the concept of supercoiling, we can imagine a phone cord (see figure 1.3). The wire that coiled regularly represents Lk0 that is natural linking number. If we coil that cord on itself, it becomes a "supercoiled" phone cord. Same as the phone cord exemplary, if DNA twists on its helical axis, we called that DNA supercoiled. We can say that DNA is relaxed if we cannot see any contortion of DNA.

Figure 1.3 DNA supercoiling with phone cord illustration [6]

Twisting direction of DNA helixes indicates the type of supercoiling. If helical axis under winds each other, the result will be negative supercoiling. DNA negative supercoils are common in cells. Rarely, DNA produce positive supercoils also which is over winding of helixes. Both negative and positive supercoiled DNA can be seen in figure 1.6.

Supercoiling density is the change of topology from relaxed position and it is shown as σ. In order to determine the topological state of supercoiled DNA, supercoiling density is useful entity since obtaining its value with experimental methods is effortless and the length of DNA has no association with supercoiling density. Supercoiling density is also known as super helical density and given by the formula:

$$\sigma = \frac{\Delta Lk}{Lk_0}$$

(1.3)

Supercoiling density is negative in general. Despite the fact that DNA helix has positive linking number, it is in under wound state in the cell so its linking number is positive [9].

The only one way to change linking number is cutting one strand of DNA and passing the other strand around the cleaved strand. This relaxation is achieved during the normal cellular activity during transcription or replication. The particular enzymes that relax the supercoils and alter the DNA topology during those activities are called topoisomerase. There are different types of topoisomerase, which can be mainly categorized with respect to the number of strands broken. The detailed information about types of topoisomerase is given in next section [10].

1.3 TOPOISOMERASE ACTIVITIES AND CELLULAR FUNCTIONS

Topoisomerases are enzymes that solve topological problems stemming from the helical structure of DNA such as knotting supercoiling and catenation, which occur during replication, transcription, recombination and chromosome segregation [11]. Topoisomerases increase or decrease the linking number to alter the topological state of DNA without making any change on its primary structure. These enzymes accomplish this duty by temporarily breaking one or two strands of DNA and transferring the intact strand from that persistent gap, and then combining previous broken strands together again [12].

DNA supercoiling can be eliminated by topoisomerase with revolving and relaxing. While the transcription or recombination processes continuing, topoisomerase remove the twisting force on the helical axes of DNA. Therefore, one of most important properties of DNA molecules can be altered or removed by these amazing enzymes. They also, change knotting or catenation by crossing DNA strands one to another.

Topoisomerases cut strand of DNA by the following method. Firstly, tyrosyl oxygen of topoisomerase makes transient covalent bond with phosphorus of broken DNA strand. During this process, topoisomerase also breaks DNA phosphodiester bond [13]. Topoisomerases are able to change topology of DNA while this transient covalent bond exits. After the topological problems have been solved, enzymes rejoin the broken strand. Process now flows in opposite direction. The phosphorus atom of the phosphotyrosine bond is now connected by hydroxyl oxygen atom, this breaks the covalent bond, and DNA backbone is renovated [14]. Renovating DNA process does not need any external energy because phosphotyrosine binding energy is utilized in this reaction. Therefore, process of cutting and rejoining does not have any energy requirement factor [15].
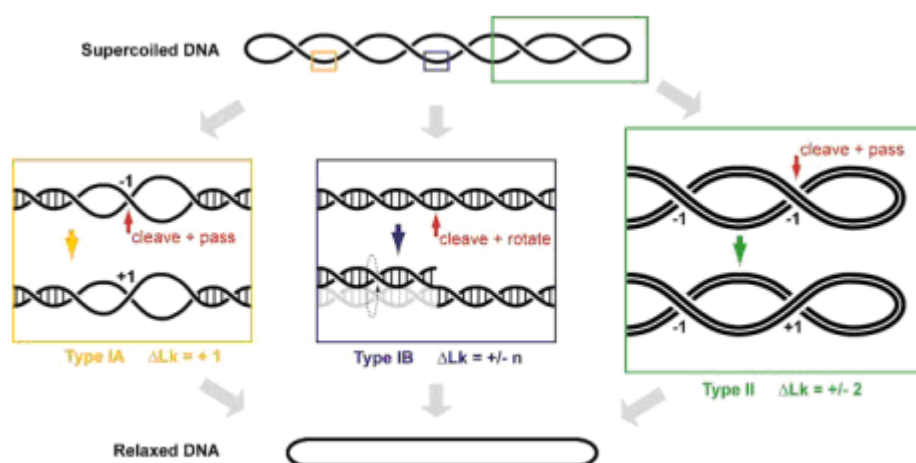
Figure 1.4 The activities of topoisomerases [16]

There two types of topoisomerases; type I and type II that are classified by their function, amino acid sequence and structure. Type I enzymes are monomeric and effect the topology of DNA by transiently cleaving one DNA strand whereas Type II enzymes are dimeric and introduce double strand breaks in a DNA duplex. Figure 1.4 represents the basic differentiation of the topoisomerases with respect to the number of strands cleaved and the number of steps that the Lk can take. Type I enzymes further subdivided based on the mechanistic and structural differences. Type I enzymes further

classified as IA and IB that have different structures and nick the DNA strand through 3' and 5' respectively. Various topoisomerase types and some subfamily members of them are given in the table 1.1.

Table 1.1 Some example for topoisomerase types [14]

Enzyme	Type	Source
Bacterial topoisomerase I	IA	Bacteria (e.g. e-coli)
Eukaryotic topoisomerase I	IB	Eukaryotes
Vaccinia virus topoisomerase I	IB	Vaccinia
Reverse gyrase	IA	Thermophilic Archaea
DNA gyrase	IIA	Bacteria (e.g. e-coli)
Eukaryotic topoisomerase II	IIA	Eukaryotes
Topoisomerase VI	IIB	Archaea

As a summary Type I enzymes are monomeric, do not require additional energy or metal cation and alters the Lk number with one-step. They link to one strand of the DNA and the other intact strand rotates around the cleaved one with torsional stress. The classification of the type I enzymes are based on structural differences in addition with the residue type that they are connected in the backbone.
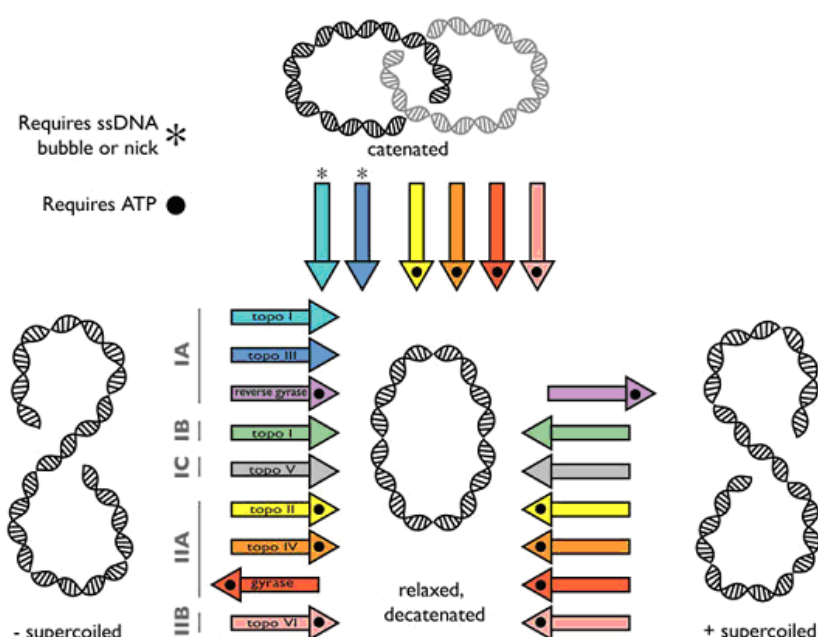
Figure 1.5 A summary of the topoisomerase types and their functions [16].

Type II enzymes, on the other hand, are dimeric, consume ATP during their action and change the Lk in two steps. They cut both DNA strands and it is believed that both separated and treated in different parts of the enzyme. In figure 1.5, the topoisomerase types are shown with the arrows where the representatives colors as defined in the figure. The ATP requirement is shown with dot on the arrows.

1.4 E.COLI DNA GYRASE A SUBUNIT

In this study, the enzyme of interest is the *Escherichia coli* (e-coli) DNA Gyrase A that is a member of type II topoisomerases. Topoisomerase type II is an extensive enzyme family having members in both eukaryal and bacterial organisms [17]. The eukaryotic enzymes have a dimeric subunit structure while the prokaryotic enzymes are A_2B_2 type tetramers. Bacterial DNA Gyrase is a prokaryotic enzyme having two subunits B and A, referring to the amino and carboxyl terminals of the eukaryotic enzyme respectively [18]. The structural and mechanistic information on the e-coli DNA Gyrase is included in the following sub-sections.

1.4.1 Structure

Although the subunit organization may differ, the two-fold symmetry does not change among the type II enzyme family. The complete crystal structure of an active topo II enzyme is not available yet; however combining the biochemical information with the current structural resolution could give a rough picture of the two-fold symmetrical fragments [16].

Type II topos consist of catalytic components attached each other by flexible hinge regions [19]. The combination of these components forms at least three interacting subunit interfaces. The subunits introduce two cavities or gates to the enzyme as in figure 1.6. The N-terminal gate (a.k.a. ATPase region or N gate) involves two elements; ATP binding fold and transducer domain jointly linked to N gate [19]. The transducer domain is thought to transmit the activity of the ATPase region to the

region where catalytic reactions between DNA and the enzyme occurs, which is termed as DNA gate. This DNA gate includes TOPRIM or CAP domain where the metal cation binds and ATPase region connects to the C-terminal domain. The tower or shoulder is the region where the third protein interface is formed that is believed to hold the holoenzyme together [16]. Figure 1.6 shows the domains on the crystal structure for e-coli enzyme.
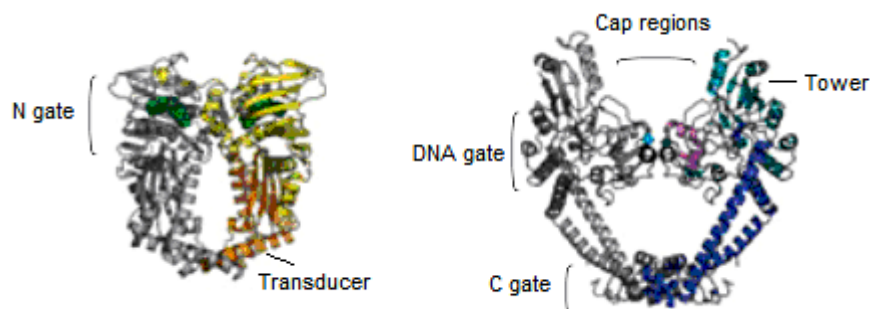
Figure 1.6 Structures of different fragments of E. coli DNA Gyrase [16]

The grey portions of the figure 1.6 are not colored intentionally due to the symmetry of the enzyme. The remaining colors are consisted with the figure 1.7 where the same domains are shown with respect to residues. The image on the left side of the figure 1.6 is the A segment of the DNA Gyrase.
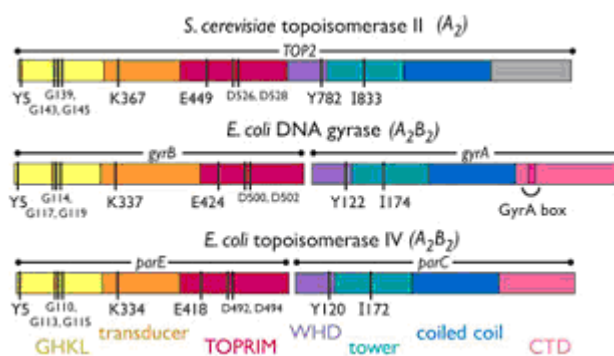
Figure 1.7 Primary domain structures of type IIA topoisomerases [16].

1.4.2 Mechanism

As mentioned before, type II topoisomerases cleave the two strand of the DNA duplex by covalently attaching the enzyme subunits to the 5' end of it. It is believed that the enzyme requires Magnesium (Mg) ion during the catalytic action of the type II topos [15]. In addition, some particular units of the enzyme hydrolyze ATP for the enzyme mechanism [17]. Figure 1.8 represents a possible mechanism of the enzyme.
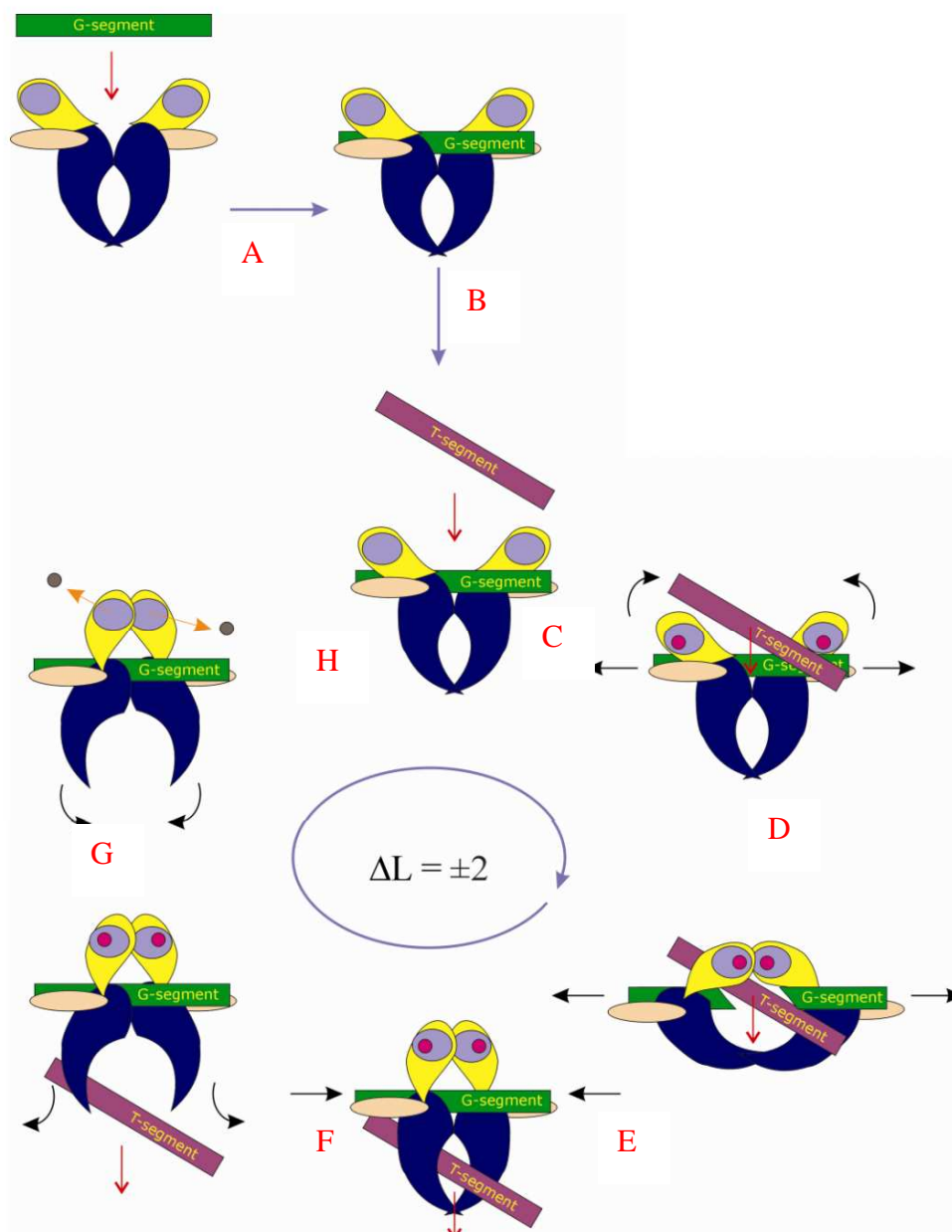
Figure 1.8 Schematic representation of a suggested mechanism for the type II topos [21]

The topoisomerase II enzymes are like a machine that grabs the double strand of DNA. Transported fragment (which is termed as T segment) is braced by the ATPase region of DNA gyrase. The gate fragment (which is termed as G segment) cut with tyrosine residue with employing phosphodiester bond. Then, T segment passed through over the gap of G segment. After passing of T segment, broken G segment reunion. All of that mechanism resulted with relaxation of DNA supercoils and linking number changed ± 2 [22]. In figure 1.8 the gray and pink dots are the unbound and bound ATP respectively. The green bar is the G-segment and the purple bar is the T-segment of the DNA. In this figure Gyrase A is the dark-blue and light orange parts of the enzyme picture.

1.4.3 The unique function of DNA Gyrase

DNA Gyrase is the only topoisomerase type that is able to introduce negative supercoils to DNA duplex. The additional domains of these enzymes increase the control over the DNA molecule and bias the activity towards the change in the topological state of it in order to maintain compaction and unwinding of the chromosomes [24].

The auxiliary domain of DNA Gyrase enzyme introducing negative supercoil is the DNA-binding region in the C-terminal domain of A subunit [15]. The supercoiling reaction requires the DNA wrapped around the C-terminal domain with a negative writhe number and a spatial relationship between T and G segments of DNA in the binding/cleavage region [15]. The enzyme uses the two DNA segments to form a positive crossover, which later on causes the negative supercoil [16].

1.5 TOPOISOMERASES AS DRUG TARGETS

Topoisomerases are the essential for the cellular growth hence these enzymes are the targets for the cytotoxic drugs which stabilize the enzyme-DNA covalent complexes [18]. These clinically approved drugs are effective as either an antimicrobials or anticancer chemotherapeutics [15]. Table 1.2 lists some examples of topoisomerase inhibitors reported so far.

Table 1.2 Inhibitors of different topoisomerases and their area of use [13]

Inhibitor	Target Enzyme	Therapeutic Value
Quinolones (e.g. ciprofloxacin)	DNA gyrase and topoisomerase IV	Effective antibacterial agents
Coumarins (e.g. novobiocin)	DNA gyrase and topoisomerase IV	Antibiotics, but not widely used
Camptothecin (e.g. topotecan)	Human topoisomerase I	Anticancer drug
Amsacrine (mAMSA)	Human topoisomerase II	Anticancer drug
Epipodophyllotoxins (e.g. teniposide)	Human topoisomerase II	Anticancer drug

The mechanisms of the most topoisomerase inhibitors have still been investigated. The Quinolones, one of the drugs against the action of the DNA Gyrase A, disturbs the DNA cleavage and reunion process and reports indicate that the enzyme sometimes could be trapped in the 5' covalent linkage [25]. The eukaryotic inhibitors are believed to act the same way, stabilizing the enzyme-DNA attachment. On the other hand, the Coumarins prevent the action of the enzyme by blocking the ATP hydrolization [13].

CHAPTER II

THEORETICAL APPROACH

2.1 FORCE FIELD AND MOLECULAR DYNAMICS SIMULATIONS

Computer simulations are of great utility in explaining dynamical and structural properties of the molecular systems. The information provided from the simulations may be used as an accuracy test between theory and application or may involve the details that the experimental methods fail to present. An effective tool of computer simulations for almost 30 years is molecular dynamics (MD) which contribute the understanding of the internal motions of the systems as a function of time [26] and determining equilibrium properties.

The history of the MD simulations ways back to the study of Alder and Wainwright with sphere balls [27]. About 20 years later the first study with a continuous potential on realistic system of liquid water was published [28]. The studies on the applied potential MD simulations are extensively in use since they first appeared in a study on a biological macromolecule [29]. The studies are mostly focused on the refinement of the experimental data, energetic calculations, structural interpretations and the internal motions of the biological molecules and their complexes. With the rapid advances in high performance computing and the scientific contributions, the simulations evolved to be more powerful tools and produce more meaningful results in years. Today, the progress is still continuing by the numerous studies in this field.

Molecular dynamics is basically the science of describing the time-depending behavior of a system of particles with the ideas of the classical physics. Although, the classical mechanics is not suitable exactly for the molecular systems which have a

quantum nature, quantum mechanical calculations is not feasible for this systems due to the large number of atoms [30]. Thus the calculations are based of the empirical potential function which does not count the quantum mechanical effects.

In a MD simulation the classical equations of motion of the many body system is calculated numerically under the effect of a boundary condition. The boundary condition should be appropriate for the geometry of the molecule which is formed from the experimental data [31]. The classical approach starts with the force-potential relation:

$$m_i \bullet \ddot{r}_i = f_i \quad \text{and} \quad f_i = -\frac{\partial}{\partial_{r_i}} U_{total} \tag{2.1}$$

where the potential energy is defined as the sum of the interaction energies of all atoms and molecules of a system of N particles. The coordinates $\vec{r}_i$ are the atomic positions of atoms. So we need to make a definition for the potential energy for the further calculations. The general treatment for the energy calculations is to separate the energy into two parts:

$$U_{total} = U_{non-bonded} + U_{bonded} \tag{2.2}$$

where the bonded energy term describes the interaction among atoms involved in covalent bonds and the non-bonded energy is the energy arising from the non-bonded interactions. The bonded energy terms are computationally easier to handle than the non-bonded terms because of being more clear and limited. The terms of the energies and the appropriate energy graphics of the terms is given in the Table 2.1.

The bonded potential consists of the interaction energies of atoms having different covalent connections:

$$U_{bonded} = U_{bonds} + U_{angle} + U_{dihedral} + U_{improper} + U_{Urey-Bradley} \tag{2.3}$$

Bond potential, U_{bonds} describes the simple harmonic motion of the two atoms connected by a covalent bond and defined as:

$$U_{bonds} = \sum_{bonds} k_b (b - b_0)^2 \tag{2.4}$$

where b is the bond length at any time, b_0 is the bond length when the system is in equilibrium and k_b is the force constant. The values of k_b and b_0 is given by the quantum mechanical *ab initio* calculations or experimentally derived from X-ray, NMR, IR, microwave, Raman spectroscopy.
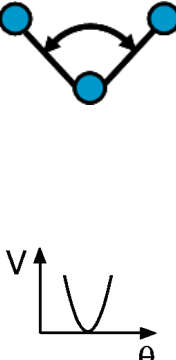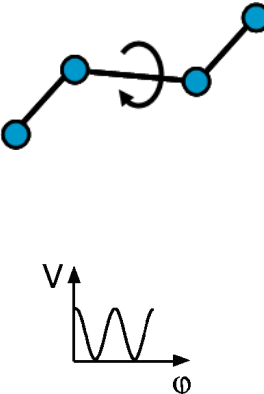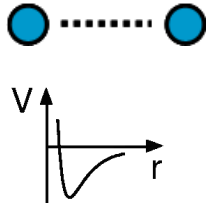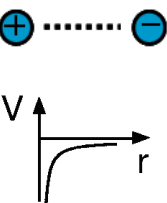
Bonded Interactions			
Bond	Angle	Dihedral	Improper

Non-Bonded Interactions	
Van der Waals	Electrostatic

Figure 2.1 Empirical potential energy terms with the proper graphical representations

Interaction of the two sequential bond forms an angle, U_{angle} contributing the bonded potential again with a harmonic function:

$$U_{angles} = \sum_{angles} k_{\theta}(\theta - \theta_0)^2 \qquad (2.5)$$

where k_{θ} is the force constant, θ_0 is the equilibrium value of the angle between the bonds and θ is the angle at any instance. The values of k_{θ} and θ_0 are again specific to the atoms and should be ready for the simulations.

The four atoms connected by three covalent bonds exhibits a periodic motion around the middle bond. The $U_{dihedral}$ term in the equation (2.3) refers to this torsional motion by the formula below:

$$U_{dihedral} = \sum_{dihedrals} k_{\phi}[1 + \cos(n\phi + \delta)] \qquad (2.6)$$

where k_{ϕ} is the force constant, ϕ is the dihedral angle, δ is the phase of the motion and n is the coefficient of the symmetry around the axis of the motion.

The last term, $U_{improper}$ in the bonded potential involves the harmonic motion of two planar conformations with the formula:

$$U_{improper} = k_w(w - w_0)^2 \qquad (2.7)$$

where k_w is the force constant and the w is the angle between two planes.

The second part of the total energy is the non-bonded potentials representing the two weak interactions between two separate atoms. The first one is the electrostatic interaction formulized by the Coulomb's equation:

$$U_{coul.} = \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \qquad (2.8)$$

where the $q_{i,j}$ is the charges of the two atoms and ε is the dielectric constant.

The second term of the non-bonded potential is the Van der Waals interaction between two atoms. A pair of distant and separate neutral atoms attracts each other by forming a dipole. While as the distance gets smaller between them, the electron clouds of the atoms overlap and in this case they repel each. One of the best approximations to explain this behavior of the non-bonded neutral atoms is given by the Lennard-Jones potential:

$$U_{r_{ij}}^{LJ} = \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad (2.9)$$

where σ_{ij} is the diameter of the atoms, r_{ij} is the distance between them and ε_{ij} is the depth of the potential well. The exponential nature of the attractive force and the surrounding N-body interactions are neglected in this model to facilitate the calculations.

All these equations together with the force constant and the other parameters used to model the behavior of different atoms or bonds and called as force field [32]. The force field parameters are both gathered from experimental data and quantum-chemical *ab-initio* calculations.

The next step after defining the energy is to derive the equations of motion of the system of N particles using Newton's second law. For the ith particle the basic equation is:

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \qquad (2.10)$$

where F_i is the total internal force acting on the ith particle and r_i is the spatial positions of the particles as a function of time. We also have a potential force relation stated as:

$$F_{i(r)} = -\nabla_i U_{(r)}$$ (2.11)

Newton's equation of motion can relate the derivative of the potential energy to the change in the positions with respect to time. The equations (2.10) and (2.11) give:

$$\frac{-dU_{(r)}}{dr_i} = m_i \frac{d^2 r}{dt^2}$$ (2.12)

The derivation of the equation of motion seems to be trivial using the above equation for a particle. Whereas the system subjected to molecular dynamics consist of many particles so the calculation gets difficult and time-consuming. Hence we require a numeric algorithm to simplify the computing. In this study we employed the Leap-frog algorithm.

The basic idea to make the numerical analysis is to divide the integration into many small stages δt. We determine the acceleration from force and distribute initial values for velocity and positions with respect to the temperature using Boltzmann statistics. From the equilibrium kinetic energy, the standard deviation for the initial velocities is [33]:

$$\left\langle v_i^2 \right\rangle = \frac{k_B T}{m_i}$$ (2.13)

where k_B is the Boltzmann constant, T is the temperature and m_i is the mass of the ith particle. Then, finally we will use these values to find the velocities and positions at time $t + \delta t$. Verlet algorithm [34] employs these calculation steps on the other hand Leap-frog algorithm makes an adjustment on this basic idea by calculating the velocities and positions at different time steps.

For a simpler calculation we need an approximation combining the positions, velocities and accelerations which is provided by Taylor expansion of 3rd order:

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + \ddot{r}(t)\delta t^2 \tag{2.14}$$

The Verlet algorithm conducts the calculations using the positions at time t, and accelerations at time $t - \delta t$ to find the positions at $t + \delta t$. The velocities are given by the below formula:

$$\dot{r}_i(t) = \frac{r_i(t + \delta t) - r_i(t - \delta t)}{2\delta t} \tag{2.15}$$

The velocity in the Leap-frog algorithm is used at half time-step [35]:

$$\dot{r}_i(t + \frac{\delta t}{2}) = \dot{r}_i(t - \frac{\delta t}{2}) + \ddot{r}(t)\delta t \tag{2.16}$$

whereas the positions are calculated at each time-step:

$$r(t + \delta t) = r(t) + v(t + \frac{\delta t}{2})\delta t \tag{2.17}$$

Starting from the acceleration at current step, Leap-frog algorithm gives the velocities at half time step and the position at the next time step using these velocities. So the velocities leap over the positions and then positions leap over the velocity [35]. Decreasing the time step and providing explicit calculation of velocities Leap-frog algorithm is known to be the most accurate and stable numeric method to overcome the massive calculations.

2.2 HALF QUADRATIC BIASED MD (HQBM)

The dynamic simulation of a conformational change in the molecule is performed by the method termed as high quality biased molecular dynamics (HQBM).

The movement of the specific segment of the molecule without altering the macroscopic state is achieved by the small time dependent external perturbations hence the constant of the motion remains constant. The conservation of the macroscopic quantities could be examined by employing an extended Lagrangian. The HQBM method is run by defining two main elements [36]: a reference (or target) structure which implies the desired conformation and a reaction coordinate which is the distance between the atoms of the initial and the aimed conformation.

The molecular potential energy is supplemented by the external perturbation applied through the reaction coordinate which is introduced as:

$$\rho(t) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq 1}^{N} (r_{ij}(t) - r_{ij}^{R})^2 \qquad (2.18)$$

where N is the number of particles, R marks the final coordinates and r_{ij} is the distance between the ith and the jth atom [37].

The half quadratic external perturbation depending on this reaction coordinate has the form [38]:

$$W(r,t) = \begin{cases} \dfrac{\alpha_{hqbm}}{2}(\rho - \rho_0)^2 & if \quad \rho(t) < \rho_0 \\ 0 & if \quad \rho(t) \geq \rho_0 \end{cases} \qquad (2.19)$$

where the ρ_0 and the $\rho(t)$ is the initial reaction coordinates at $t = 0$ and at any time. The perturbation does not act on the dynamics of the system when the reaction coordinate undergoes a spontaneous increase from time step t to $t + \Delta t$. Instead, the change in the reaction coordinate is adjusted to continue the dynamics. The case when $\rho(t)$ is smaller than ρ_0, the harmonic force keeps the reaction coordinate from a significant decline. The term α_{hqbm} is a user defined value restricting the backward fluctuation of the reaction coordinates [39]. Since α_{hqbm} is an external effect to the

system, the smaller α_{hqbm} gives more accurate results. The α_{hqbm} value should be set similar to the free system fluctuations to prevent artificial results.

The resulting force of the perturbation energy is given by:

$$F(t) = \frac{d}{dr}\left[W(r,t)\right] \qquad (2.20)$$

which yields the equation below.

$$F(t) = \alpha_{hqbm}\left[\rho(t) - \rho(t + \Delta t)\right]\ \left[\rho(t)\right]^{1/2} \qquad (2.21)$$

As mentioned in Chapter 1, stable dimeric structure of DNA Gyrase A subunit has not been found yet. Therefore, we generated "missing" chain with symmetric matrix operation. The proposed three dimensional structures, which are called BIOMT structure, are modeled by applying HQBM method on both two chain of protein. Also, suggested mechanisms that involves gate opening and closings have been accomplished by applying HQBM method on relevant segment of enzyme.

The perturbation does not affect the macroscopic state of the system when the free energy surface of the system is diffusive, i.e. the temperature remains the same during the time dependent motion. However when the system has to deal with a free energy barrier during the conformational transitions, an amount of potential energy could be transformed into kinetic energy [39]. Hence, the simulations should be run at constant temperature to avoid such transitions.

2.3 CONSTANT TEMPERATURE/PRESSURE MD (CPT MD)

The core of the molecular dynamics is the Newton's second law which involves only the energy conservation during the simulation. In order to keep the temperature constant the system should be imposed to a heat bath. The system conserving the temperature (T), the number of molecules (N) and the volume (V) is called as the

canonical ensemble (NVT). Under NVT conditions Boltzmann probability distribution function is valid which gives the initial velocities as in the equation (2.13).

In this study the Berendsen [40] thermostat method is applied. Berendsen method is based of Andersen method [41] with a decreased drastic change to the system. The particles in contact with the heat bath collide to the thermal bath particles and their velocities change. Andersen method simply resets the velocities of these particles and assigns new random velocities according the Boltzmann distribution causing a drastic error as a result. On the other hand Berendsen method deals with this thermal energy exchange between the system and the thermal bath by gradually assigning velocities to the interacting particles with a factor depending on the time step and the temperature:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T}{T_{ins}} - 1 \right) \right]^{1/2} \qquad (2.22)$$

where Δt is the time step, τ_T is the time-constant, T_{ins} is the instantaneous and T is the desired temperature. This factor provides a gentle adjustment to keep the temperature at the desired value [42].

2.4 COMPUTATIONAL TOOLS

2.4.1 Software

We have used the CHARMM (Chemistry at HARvard Macromolecular Mechanics) program for our molecular dynamic simulations and analysis. CHARMM has been developed be Karplus et.al. about 30 years ago and its cumulative progress are still continuing [43]. The program includes comprehensive analysis facilities for structure comparison, energy evaluation as well as dynamic calculations and model-building.

We have used c34a1 version of CHARMM. There are lots of installation options according to the hardware and system that software used. We have preferred to install software with "install.com gnu xxlarge M" options. We have utilized Visual Molecular Dynamics (VMD) in order to display and analyze the molecular assemblies. Version of vmd is 1.8.6. GNUPLOT program is used for the data plotting. We also have run a variety of LINUX shell scripts to modify and arrange the data sets.

2.4.2 Hardware

Our simulations employed by using clustered supercomputer, which we named "ulubatli". It has a head node and nine compute nodes. Compute nodes connects each other with high-speed infiband network, which has 20 Gb/sec bandwidth. Seven of compute nodes are HP DL 140 G3 model and other two nodes are HP DL 160 G5 model. Each of HP DL 140 G3 node has two Intel Xeon 5365 cpus and 12 GB DDR2 memory. HP DL 160 G5 nodes have two Intel Xeon 5462 cpus and 16 GB DDR2 memory. We have 72 cpus and 120 GB memory for computational facilities. Our head node Ulubatli is a HP DL 380 G5 server which has a fiber connected disk unit of HP MSA 30. Totally, ulubatli has 4.2 TB disk but some of them used as redundant with raid configuration. We run Load Sharing Facility (LSF) for the job management and scheduling and for cluster management, we used Cluster Management Utility (CMU) which is a HP branded software. The Scientific & Technological Research Council of Turkey (TUBITAK) funded our supercomputer Ulubatli. Documentation for support is 107T209 as the grant number.

2.5 DATA STRUCTURE FILES

CHARMM program stores the chemical arrangement and composition, atomic properties, internal coordinates and force-energy parameters of the molecules in the data structures files. The topology and parameter files contain information about a particular class of molecules. Later, when these particular molecules combine to a specific molecule the protein structure and coordinate files are written.

2.5.1 Residue Topology File (RTF)

Residue topology file (RTF) specifies the partial charge of each atom in the molecule, hydrogen-bond acceptors and donors, internal coordinates, bond, angle, dihedral and improper organization of atoms of the molecular building blocks such as proteins or nucleic acids. CHARMM needs these data to characterize a specific molecule and perform its energetic calculations and structural manipulations [44].

2.5.2 Parameter File (PARAM)

The parameter file is associated with the RTF file and it contains all required information for the calculation of the empirical potential. The potential energy terms defined in section 2.1 including the equilibrium bond distances, bond stretching angles, angle bending and dihedral angle and the force constants and parameters are listed for the particular atom types and water molecules form which the specific molecules are constructed [44].

2.5.3 Protein Structure File (PSF)

CHARMM reads the required information regarding to the compositions and connectivity of the molecules of interest. The bond, angle, dihedral and improper formation is specified and the molecular entities are grouped as segments. The molecular topology is defined using the information in the PSF file. The parameters on the other hand are read from the PARAM file [44].

2.5.4 Coordinate File (CRD)

The CRD file contains Cartesian coordinates of all atoms in the specific molecule which is obtained by the NMR or X-ray crystal structures. The hydrogen atoms and some missing coordinates could be generated by proper facilities of CHARMM.

2.6 SYSTEM PREPARATION

2.6.1 PSF Generation

Before employing a simulation with CHARMM, we have initialized some parameters and files. First, we have generated the protein structure file. Then, we set up the initial coordinate from crystal structure and generated missing hydrogen atoms in pdb file. The name of the initial input that we used is build.inp. Needed commands for psf creation and hydrogen generation are contained in build.inp file. CHARMM can read crd and pdb files to initiate the coordinates. We used PDB formatted data to produce the protein structure [45]. The setup coordinates of the E.Coli GyraseA were gathered from the Protein Databank (entry code 1AB4) (Figure 2.2) [45].
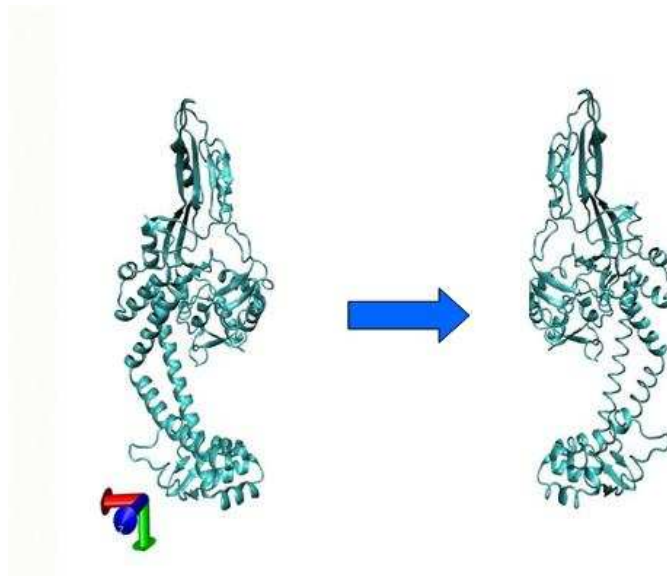
Figure 2.2 E.Coli GyraseA obtained from the Protein Databank (entry code 1AB4) [45]

A typical CHARMM input file begins with loading topology file (RTF) and parameter (PAR) file. These files is used for loading the geometric and topological structure of amino acids, nucleic acids, water and all other molecules that can be found in cell. Parameter file consist of force-field constants of molecules. In our simulations, we used mod_top_22.rtf for topology file and mod_par_22.prm for parameter file. In our build.inp file after commanding parameter and topology files, we have added amino

acid residues with the order in the pdb file. After a successful psf generation, we used HBUILD command to add hydrogen atoms that can not be found in crystal structure. After creation of su.psf for our system, we have red the formatted pdb file for placing our atoms as in the structure. Then, we have added crystal water molecules, which appears in pdb file. After creation of su.psf and su.crd files, we have generated missing chain with help of symmetric matrix operations and added them to our su.psf and su.crd file. After all, we have saved all output files to hard drive.

2.6.2 Preparation for Dynamic Calculations

There are several procedures that have to be done before employing any dynamical simulation. In preparation, we perform particular steps to make our system ready for dynamical run. First step is minimization which is the releasing local stretch occurred during crystallization. Second step is the putting system into a water box. Third system is heating system from 0 K to 300K. Final step is the equilibration of the total water-protein system.

2.6.2.1 Energy Minimization

We employed minimization with using *in_minimize.inp* input file and we calculated energies of system. As every CHARMM input file, we first read parameter and topology files, then we read our protein structure with *su.psf* and *su.crd* files that we have created missing chain with 180 degree rotation in previous input. We have not calculated coulombs potential and Van Deer Waals potential further than 14 Armstrong with using cutoff keyword. During minimization process, we used two algorithms, firstly, steepest descent (SD) algorithm for a sharp and quick minimization (Figure 2.3) and secondly, adopted based Newton-Raphson (ABNR) to perform more accurate but relatively slow minimization (Figure 2.4). After minimization performed, we have written new minimized coordinates of atoms to the disk as *su_mini.crd*. We have obtained total potential energy during minimization with CHARMM which includes bonded and non-bonded potential energy terms as explained in Figure 2.1 [44].
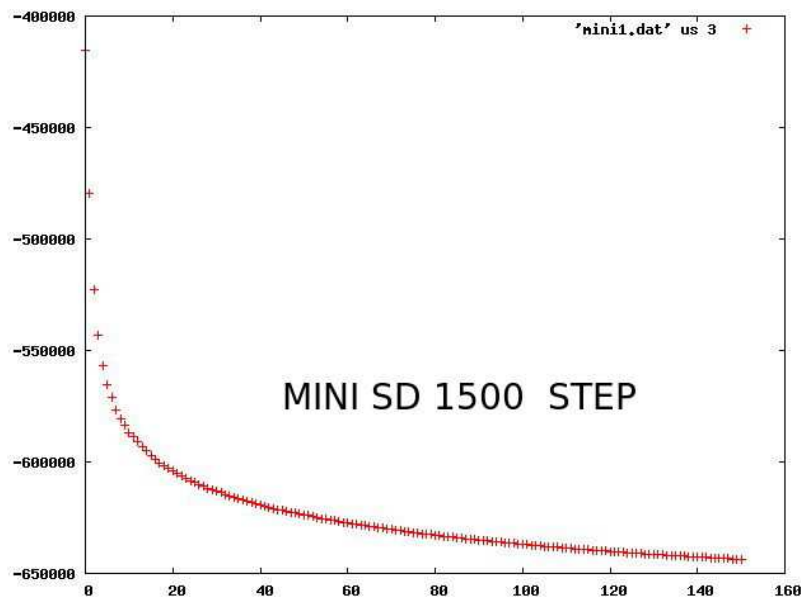
Figure 2.3 The Steepest Descent (SD) graph of 1500 steps of minimization
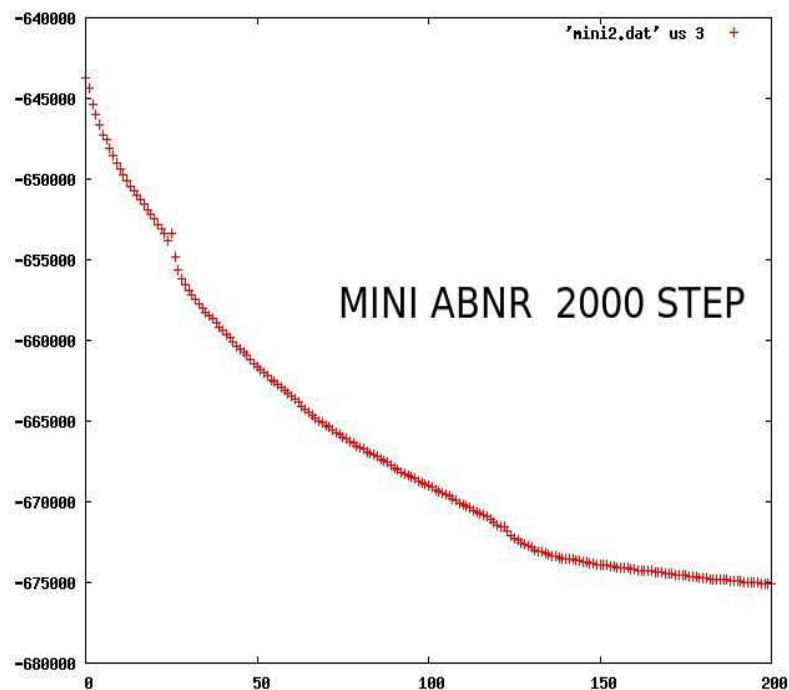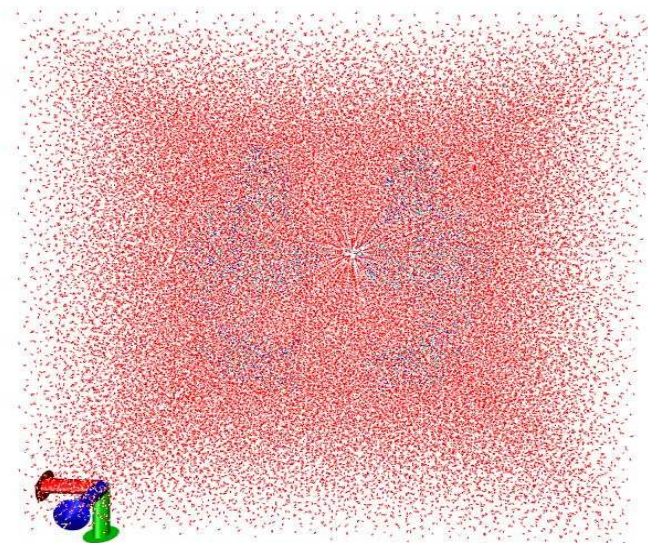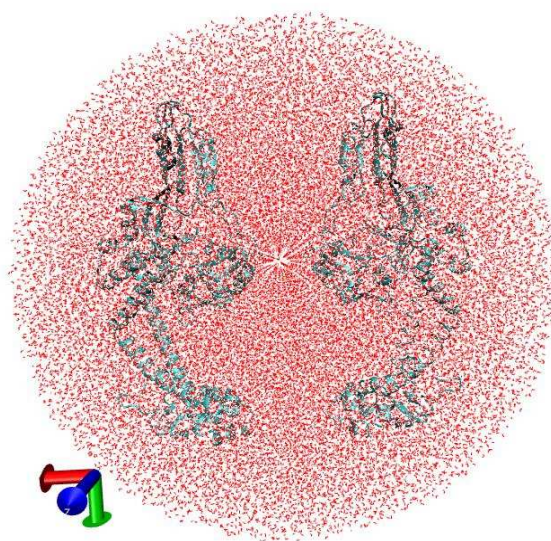

Figure 2.4 The Adopted Based Newton- Raphson (ABNR) graph for 2000 steps of minimization

a)

b)

Figure 2.5 The generated water cube (a) around the protein, (b) cube water box turned into spherical volume water box with size of 74 Armstrong

To simulate the cell environment for protein dynamics, we have generated a water box around the protein that covers all protein atoms. (Figure 2.5a) After this step, we cut our cube as sphere to reduce the atom number and to apply boundary potential to prevent our water-protein system from separation. (Figure 2.5b) Then, we have written new psf and crd files to the disk. As we examine the charge of the system, it was -24 qe and in order to neutralize the system we have added 12 Mg atoms the place with highest potential energy. (Figure 2.6)
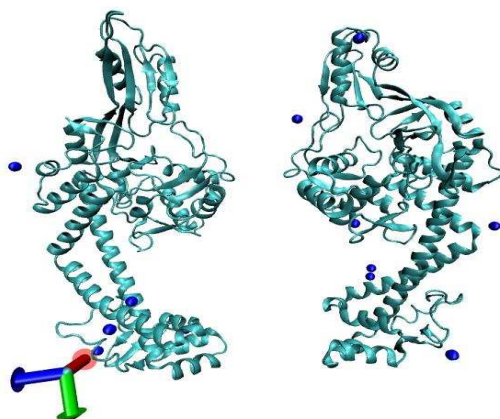
Figure 2.6 Magnesium (Mg^{+2}) atoms as seen with blue dots have added in the system

2.6.2.2 Heating the protein

Previous steps, we have generated protein, water and ion system. However, our atoms do not have any action because they have not any temperature terms yet. To make our atoms mobile and perform more realistic calculations, we have increase the temperature of the system up to 300K with 100000 step equilibration dynamics. As seen on the Figure 2.7 our potential energy increases steadily with the increase of thermal energy. We have saved our psf and crd file to the disk. In addition to that, we have saved restart file which contains the information about motion of the atoms that we will use for further dynamic simulations in future.
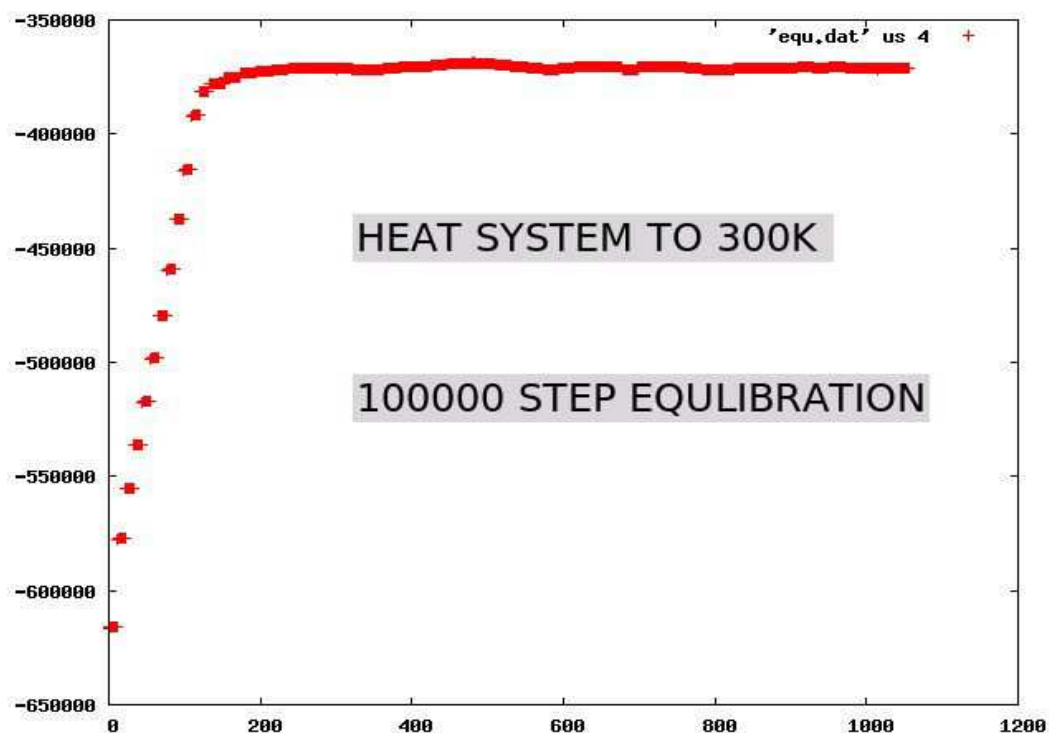
Figure 2.7 The system was heated from 0 K to 300 K with 100000 steps.

2.6.3 Equilibration

After heating system to the 300K, we have to stabilize our system before starting simulations. To achieve this, we first read our psf and crd files and then we read restart file to continue from the previous dynamic step. In equilibration we fixed our temperature as 300K and we controlled the temperature of the system during equilibration. Our equilibration dynamics consist of two parts, in first part we have employed 100000 steps under constant temperature and in the second part we used shake option at 100000 steps which takes hydrogen atoms motion as constant in order to increase our step size as 2 fs. (Figure 2.8)

a)

b)
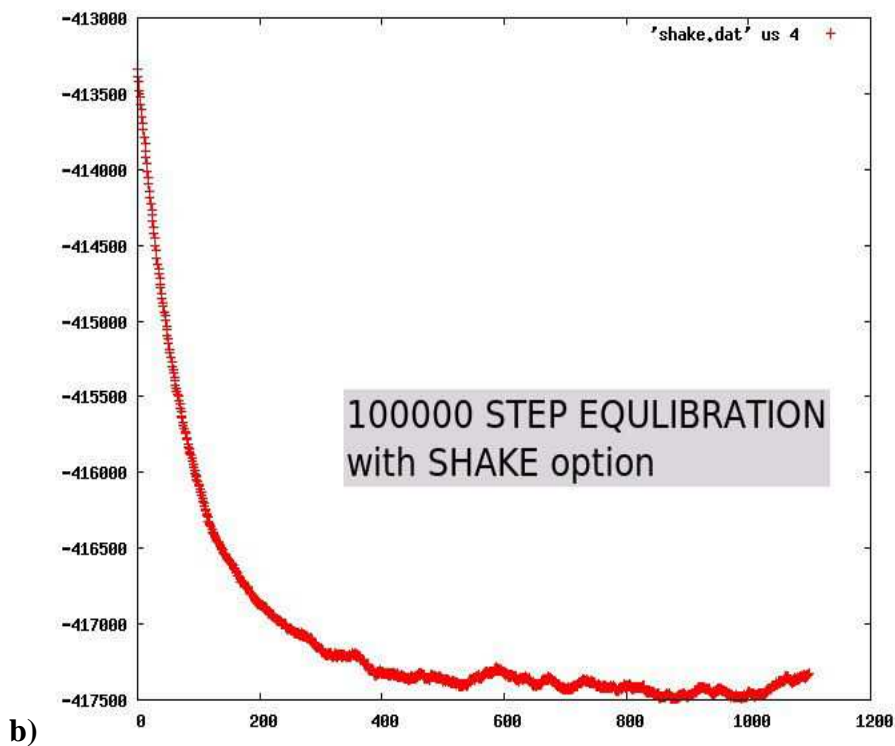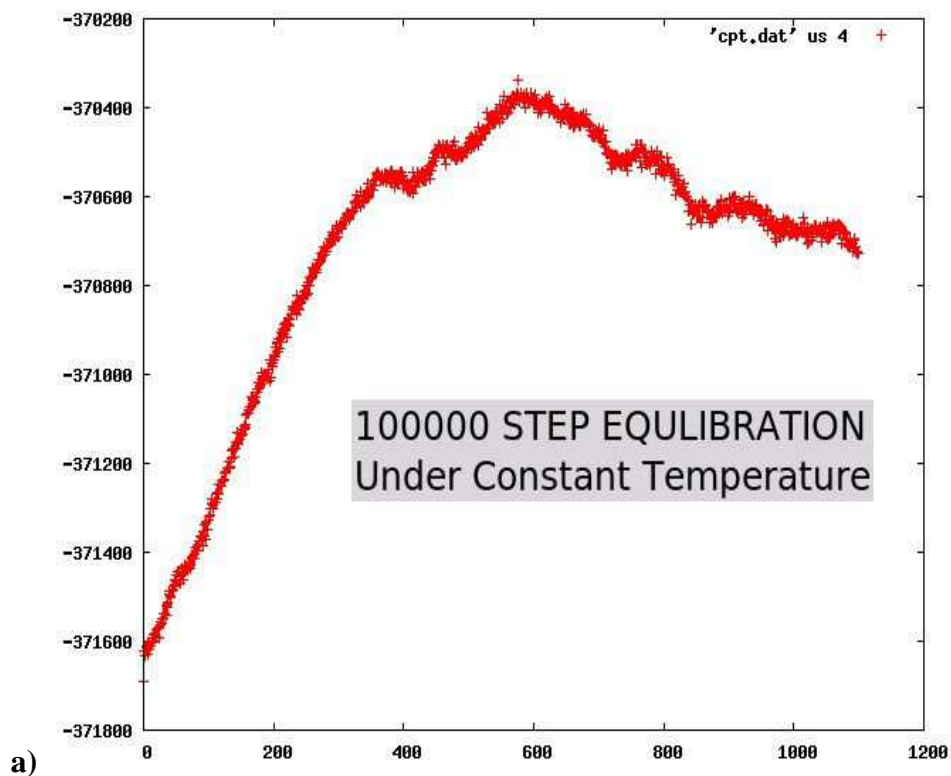
Figure 2.8 Fluctuation of total potential energy as a function of time (a), as seen in (b) at the end of equilibration we have seen stabilization of energy

2.6.4 Preparing System for Biased Molecular Dynamics

Our aim is to investigate of dynamic mechanism of closing and opening of lower and upper part of enzyme and to find the most suitable dimmer interface of dimmer structure of protein. We used HQBM approach to carry out the simulations. In all simulation we used shake keyword to achieve 2 fs step size.

In order to find a suitable dimmer interface, we have put dummy atoms at both A chain and B chain and push the chains to each other. We have placed dummy atoms in the place of desired place of CA, backbone carbon atoms of amino acid residues. For different simulations we have used different number of dummy atoms that is placed suitable part of the proteins. The details of simulations will be described in the next section.

CHAPTER 3

RESULTS AND DISCUSSIONS

To investigate the mechanism of DNAGyraseA, we first started to establish the dimer form of the protein as it is not available in literature. To do this, we have started from the monomer (as present in pdb code 1ab4.pdb), and form the second part of the dimer, as explained above in the section of theoretical methods. Then, we form 8 different systems and apply the biased MD method (see the theory part) to bring the two half of the dimer together. Our simulations are canonical (NVT) and lasted about 1 nanosecond for each one.

In the first set up (close up), we have closed the upper part of the enzyme while keeping the lower gate free. In the second set up (close down), we close lower gate domain, while keeping the upper part free. In third simulation, we have done a simulation on the final configuration of the first simulation. We closed lower gate after a successful closing of upper part. At forth setup, we have change the order of simulations, which we have closed upper part after a successful closing of lower gate in the second setup. At fifth setup, we have closed both upper and lower domain at the same time. Sixth simulation has a significant role to understand the dimer interface of the enzyme. At this setup, we forced the system to go the proposed BIOMT structure in literature. We have applied potential only the lower gate of enzyme and modeled structure is the only 14 Kda part of lower gate which crystal structure has already found. At seventh setup, we have closed upper part starting from the BIOMT structure that has accomplished at sixth setup. Finally, at eighth setup, we have opened upper gate starting from the BIOMT structure.
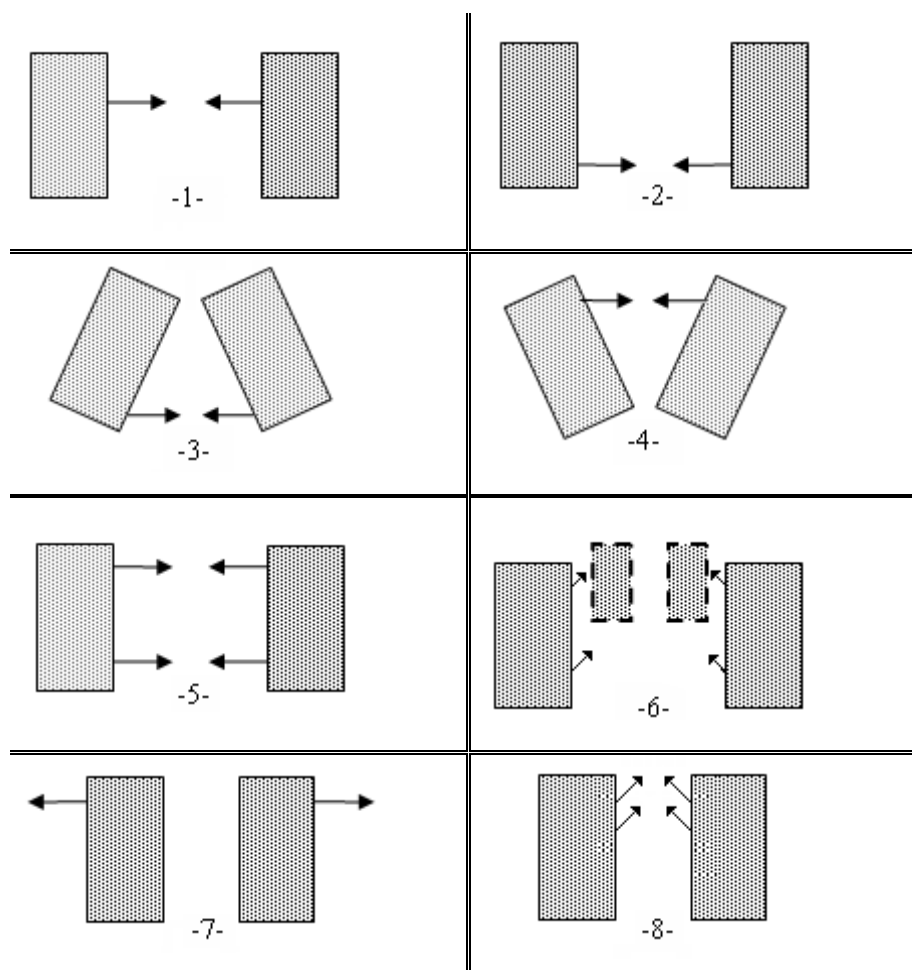
Figure 3.1 Molecular simulations that employed during HQBM dynamics

As the post-simulation analyzes, we have focused on two different aspects: structural and energetic analyses. In the structural analyses, we have looked at the, root mean square deviations (rmsd) of each secondary structure, as we superimposed the final geometry on top of the initial structures. This is done in two different theoretical approach; a rigid rmsd calculation where we used just the initial and final coordinates, and secondly an 'oriented' rmsd calculation where we minimize the rmsd by orienting the two structure. The first kind rigid rmsd will give us rigid movement of each secondary structure, while the second one will show us amount of deformation in each secondary structure. The second structural analyzes are based on calculation of the distances between each amino acid pairs. Here, we plot the changes in these distances, as we can get information about the regions of the protein which is relatively more mobile. Lastly, we have quantified amount of rigid rotations of the protein during the

gate closures. This is important to note, as such rotations may be crucial in regards to sign of supercoils that the DNA has.

In regards to energetic analyses, we have performed calculations of interaction energies between the two chains of the enzyme. Here, we wanted to see whether this interaction energy is increasing or decreasing as the two half of the protein comes together. This analyzes can give us some clues as to whether the dimer form is stable in solution or not. Secondly, the total potential energy surfaces are drawn for each simulation, to investigate possible potential barriers in the reaction coordinate. Lastly, we have looked at the amount of forces that we apply to bring the needed conformational changes for simulations.

First of all, it is better to look at the rmsd plots, and figure out any possible un-physical deformations. In an equilibrium dynamics, under room temperature, local rmsd values of a secondary structure of up to 3 to 4 Angstroms are acceptable.
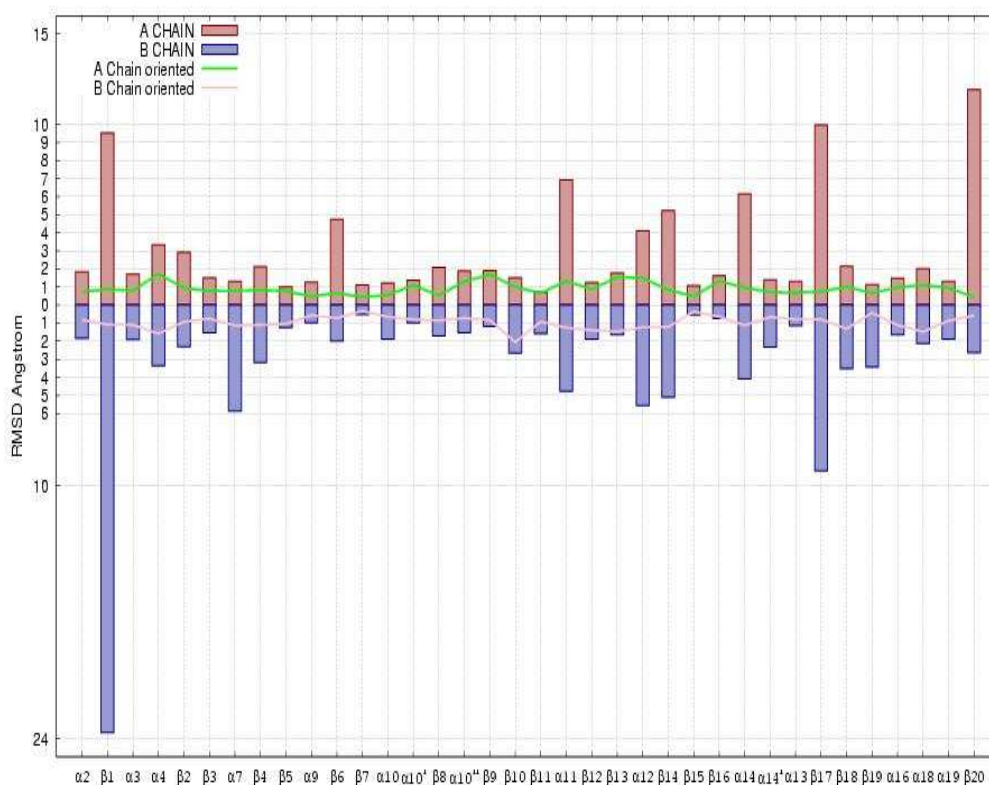

Figure 3.2 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the closing is at the upper part of enzyme

As seen in figure 3.2, the oriented rmsd values for this system are quite logical as being up to 1.8 Angstrom. Therefore, the secondary structures here are very well conserved, as it should be. However, the rigid rmsd values are seen to be up to 24 A. Secondary structures β1, β17, and β20 are observed to be very mobile, as β1 being the most mobile subunit in this system. It has about 23 Angstrom rigid rmsd value for B chain, while that of A chain is only about 9 angstrom. The other important values are 10 and 9 angstroms for β17, 12 and 3 angstroms for β20, respectively for A and B chains. One very important point here is that the rigid rmsd values of β1 are not the same for A and B chains. This asymmetry is interesting as the two half of the protein are identical in amino acid sequence.

Figure 3.3 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the closing is at the lower part of enzyme

In figure 3.3, maximum oriented rmsd values for this system are logical as being up to 2.5 Angstrom. So, the secondary structures here are very well preserved. On the other hand, the rigid rmsd values are seen to be up to 12 Angstrom. Secondary

structures β1, α14, α14¹, α15 and β20 are observed to be very mobile. Especially, secondary structures around α14 and α15 are the most moving subunit in this system. It has nearly 10 Angstrom rigid rmsd value for B chain, while that of A chain has higher value about 15 Angstrom. Other prominent values are 12 Angstroms for α15 B chain and 14 Angstroms for A chain. Activity of α15 during closing down lower part is meaningful because this residue is very close to the segment that applied force. For α14, α14¹, both residues are very flexible to move and their movements are about 10 Angstroms for both A and B chains. Residues of α11 and α16 are moving nearly 5 Angstroms with both sides of chains. We can see symmetric behavior in β1, α14, α14¹, α15 and β20 which means that both chain closes to each other when lower part comes together with the same segment of amino acid residues.
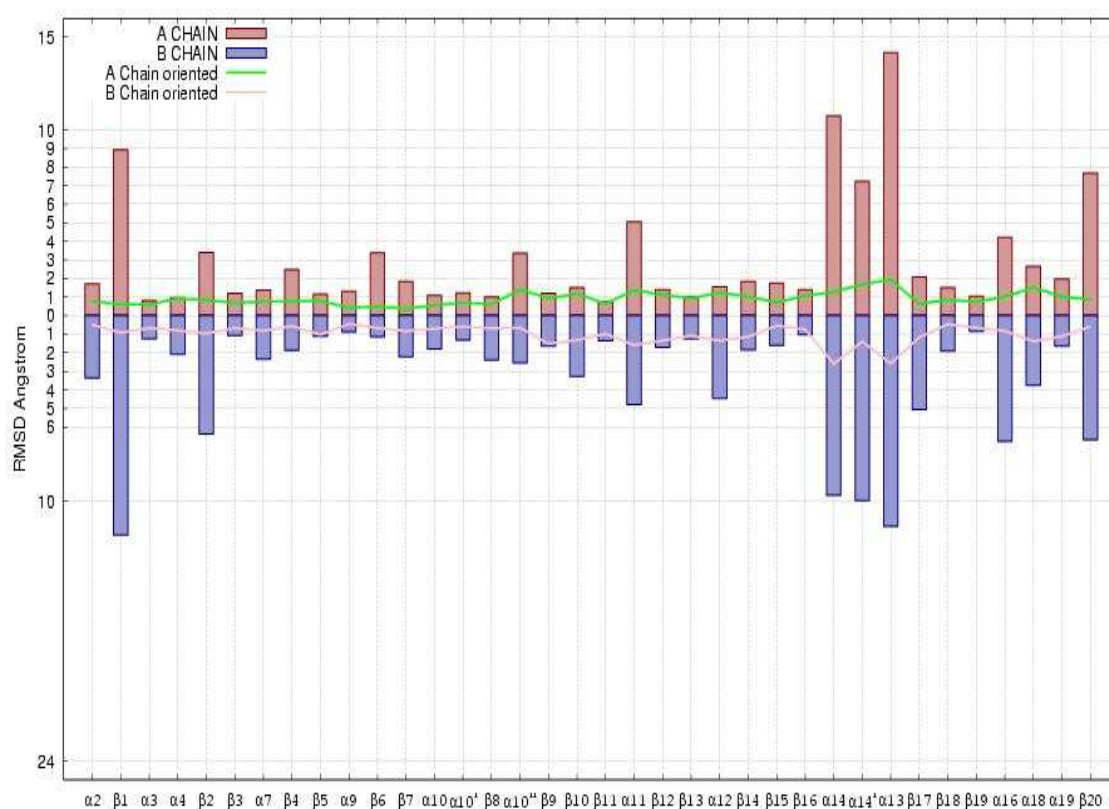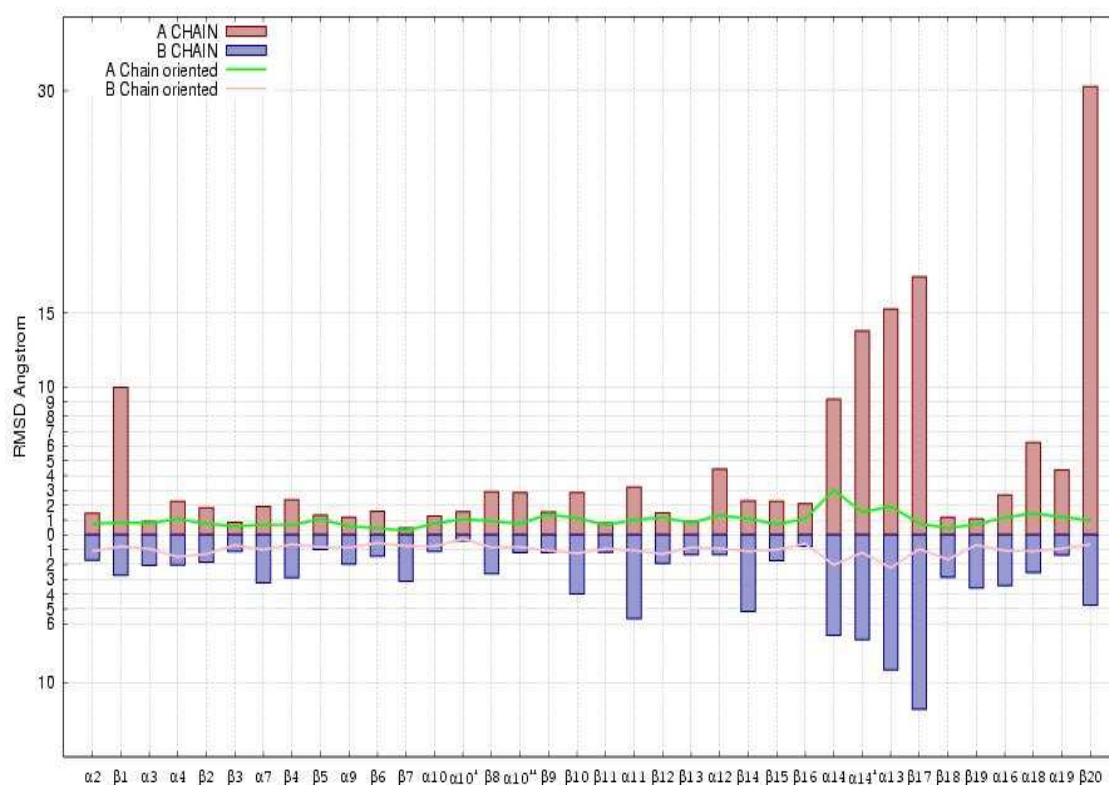

Figure 3.4 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the closing the lower part starting from closed upper part structure which is defined as setup 1

In figure 3.4, if we analyze the total change in rmsd value for different secondary structures, we can state that, maximum change the oriented rmsd values for

this system are up to 3.0 Angstrom which corresponds to the α14 residue. Change in oriented rsmd values nearly 3.0 Angstroms are well enough to say that secondary structures in this simulation are conserved. Nevertheless, rigid rmsd values can alter up to 30 Angstrom. Secondary structures β1, α14, α14¹, α15, β17 and β20 are observed to be very movable as β20 is the most mobile subunit in this system. Rigid rmsd value for A chain is about 30 Angstrom and for the other chain, B chain is moving with around 13 Angstrom. Another significant remark for this setup is β1 residue has change of 10 Angstrom in A chain and nearly 3 Angstrom change in B chain. β1 residue has least action in this setup with respect to other simulations. Secondary structures α14, α14¹, α15, β17 change about 9, 14, 15, 17 Angstrom respectively in A chain and 7, 8, 9, 13 for B chain respectively. On the other hand, residue β20 is the most mobile during this simulation with change above 30 Angstrom in A chain and 5 Angstrom in B chain. This secondary structure has always large degree of mobility in other systems but not as high as observed is this setup. If we take the movement of β20 residue as exception, change in rmsd values mostly symmetric.

In figure 3.5, some residues of enzyme have high mobility than other residues. β1, α14, β17 and β20 have rmsd values around 10, 13, 28, 9 Angstrom for A chain and 16, 4, 11, 14 Angstrom for B chain. β17 has the most mobile secondary structure with total 28 Angstrom rigid change in A chain and 11 Angstrom rigid change in B chain. Movements of secondary structures are mainly asymmetric and that is interesting because both chain A and B have the same structure. Oriented rmsd values for secondary structure are plausible which nearly 3.0 Angstroms is. Hence, we can infer that secondary structures in this setup conserved during simulation.
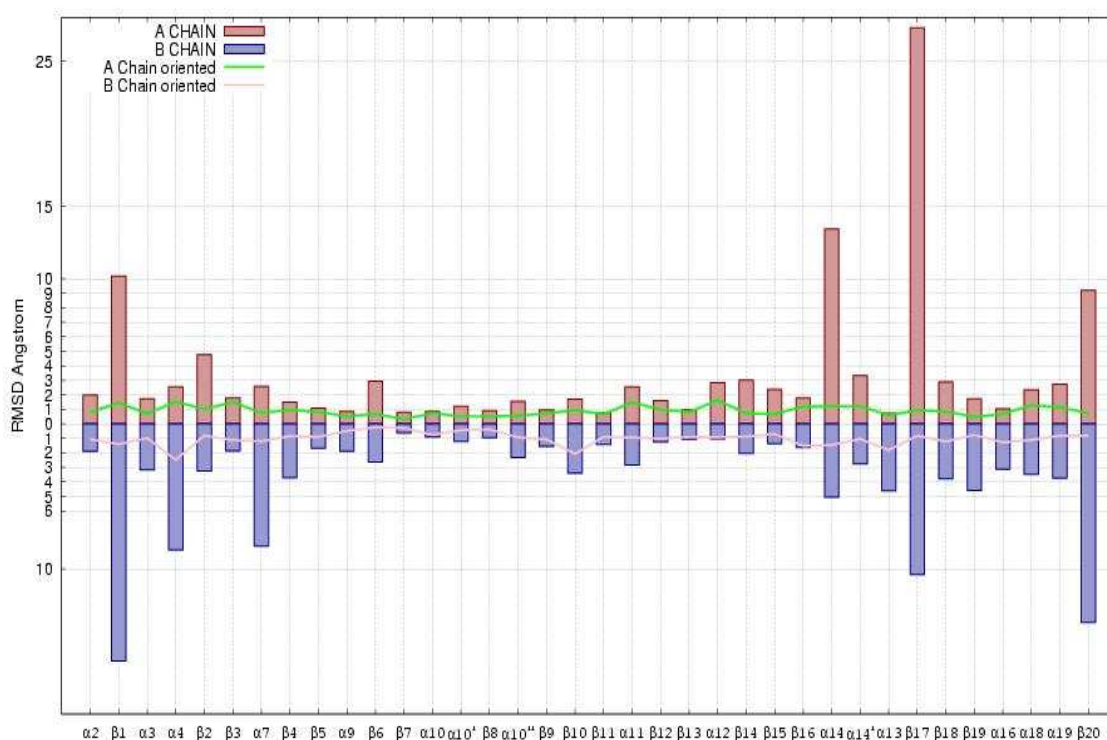
Figure 3.5 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the closing upper part of enzyme started from closed lower part structure as explained in the second setup

In figure 3.6, maximum oriented rmsd values for this system are logical as being up to 4.0 Angstrom. So, the secondary structures here are preserved. However, the rigid rmsd values are seen to be up to 27 Angstrom. Secondary structures β1, α14, α14¹, α15 and β20 are observed to be very mobile. Especially, secondary structures around β1 and β20 the most moving subunit in this system. It has nearly 15 Angstrom rigid rmsd value for B chain, while that of A chain has higher value about 15 Angstrom. Other prominent values are 12 Angstroms for α15 B chain and 14 Angstroms for A chain. Activity of α15 during closing down lower part is meaningful because this residue is very close to the segment that applied force. For α14, α14¹, both residues are very bustling and their movements are about 10 Angstroms for both A and B chains. Residues of α11 and α16 are moving nearly 5 Angstroms with both sides of chains. We can see symmetric behavior in β1, α14, α14¹, α15 and β20 which means that both chain closes to each other when lower part comes together with same segment of amino acid residues.
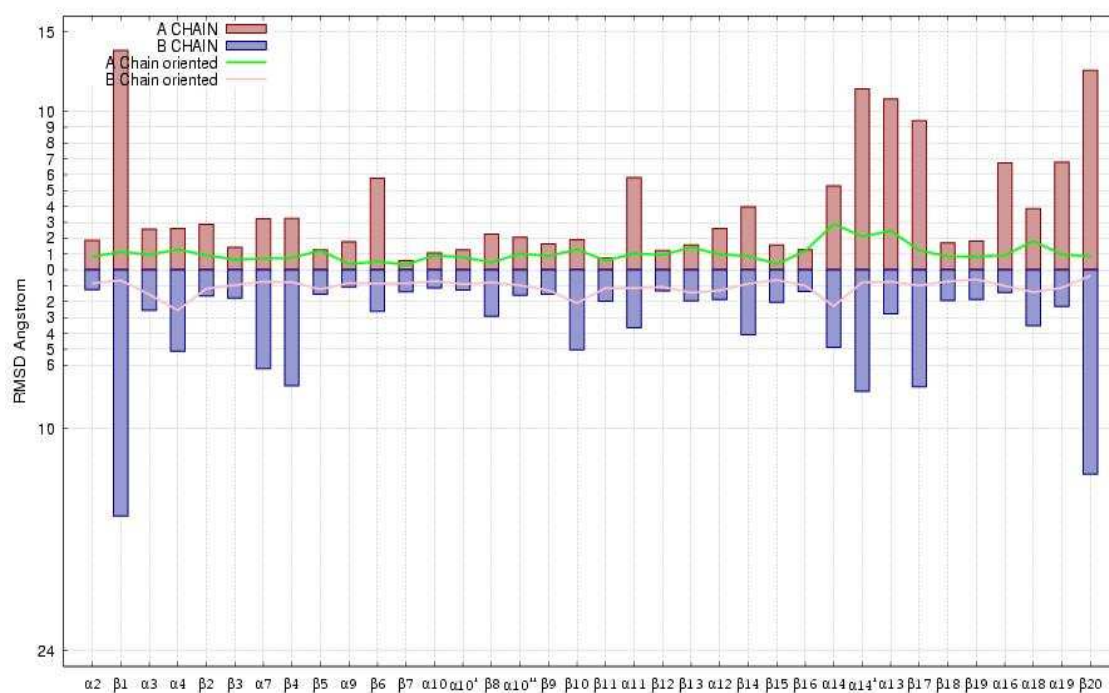
Figure 3.6 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as closing enzyme both upper and lower gate with the same time
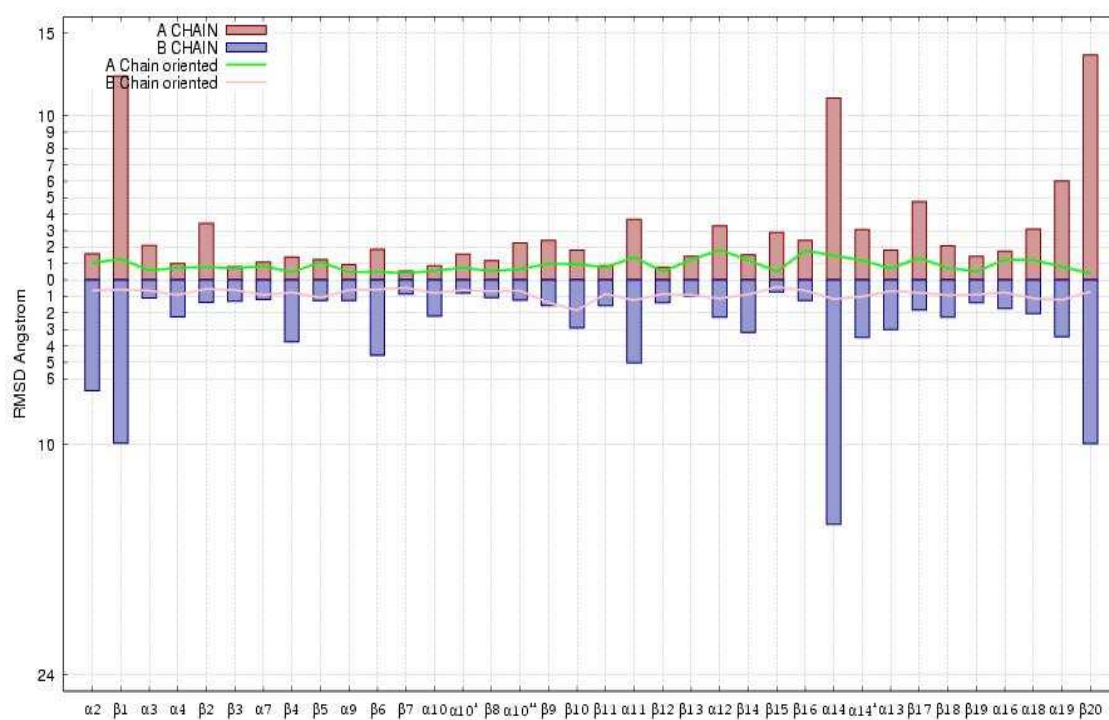
Figure 3.7 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as pushing to the system to go to BIOMT structure with closing the lower part of enzyme

In figure 3.7, the oriented rmsd values for this system are quite logical as being up to 3.0 Angstrom. Therefore, the secondary structures here are very well conserved, as it should be. However, the rigid rmsd values are seen to be up to 15 A. Secondary structures β1, α14, and β20 are observed to be very mobile, as α14 being the most mobile subunit in this system. It has about 15 Angstrom rigid rmsd value for B chain, while that of A chain is only about 11 angstrom. The other important values are 10 and 9 Angstroms for β1, 13 and 10 angstroms for β20, respectively for A and B chains. One very important point here is that the rigid rmsd values of β1 are not the same for A and B chains. If we take the β1 as exception, generally secondary structures movements are shown in symmetric.
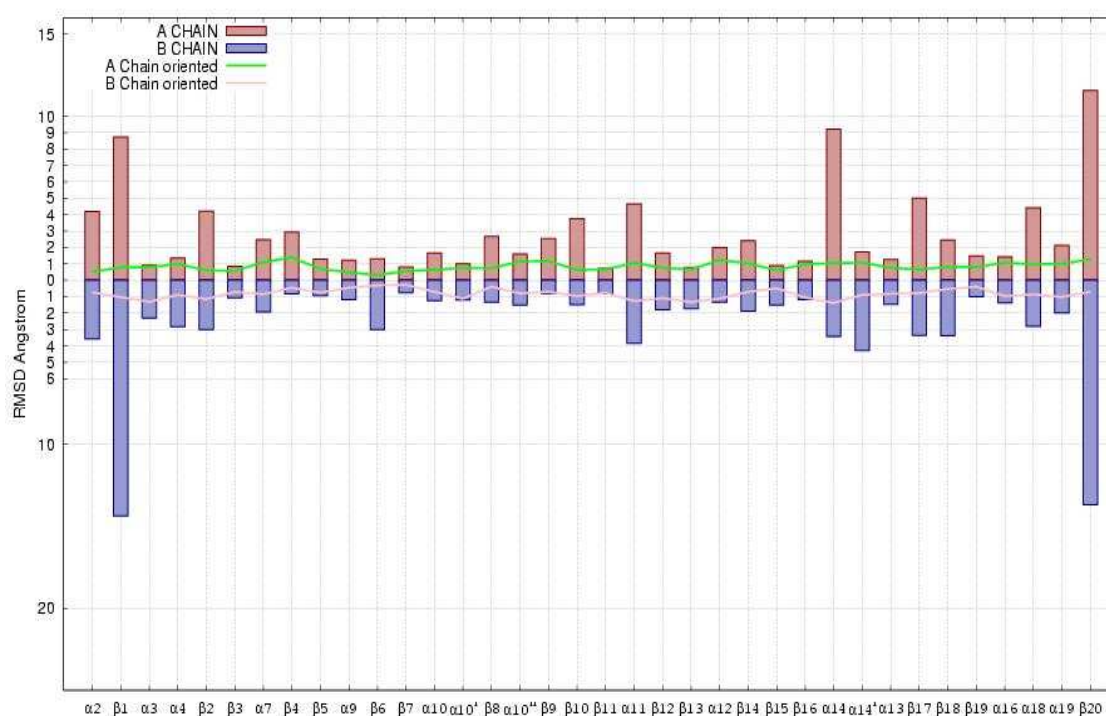

Figure 3.8 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as opening the upper part of the system starting from the proposed BIOMT structure with accomplished with sixth set-up

In figure 3.8, if we analyze the total change in rmsd value for different secondary structures, we can state that, maximum change the oriented rmsd values for this system are up to 2.0 Angstrom which corresponds to the α14 residue. Displacements are well enough to say that secondary structures in this simulation are conserved. Nevertheless, rigid rmsd values can alter up to 15 Angstrom. Secondary structures β1, α14, and β20 are observed to be very movable which has rigid rmsd value for A chain is about 15 Angstrom and for the other chain, B chain is moving with around 13 Angstrom. Another significant remark for this setup is β1 residue has change of 9 Angstrom in A chain and nearly 15 Angstrom change in B chain. Rmsd values change with symmetrically in both chains.

.

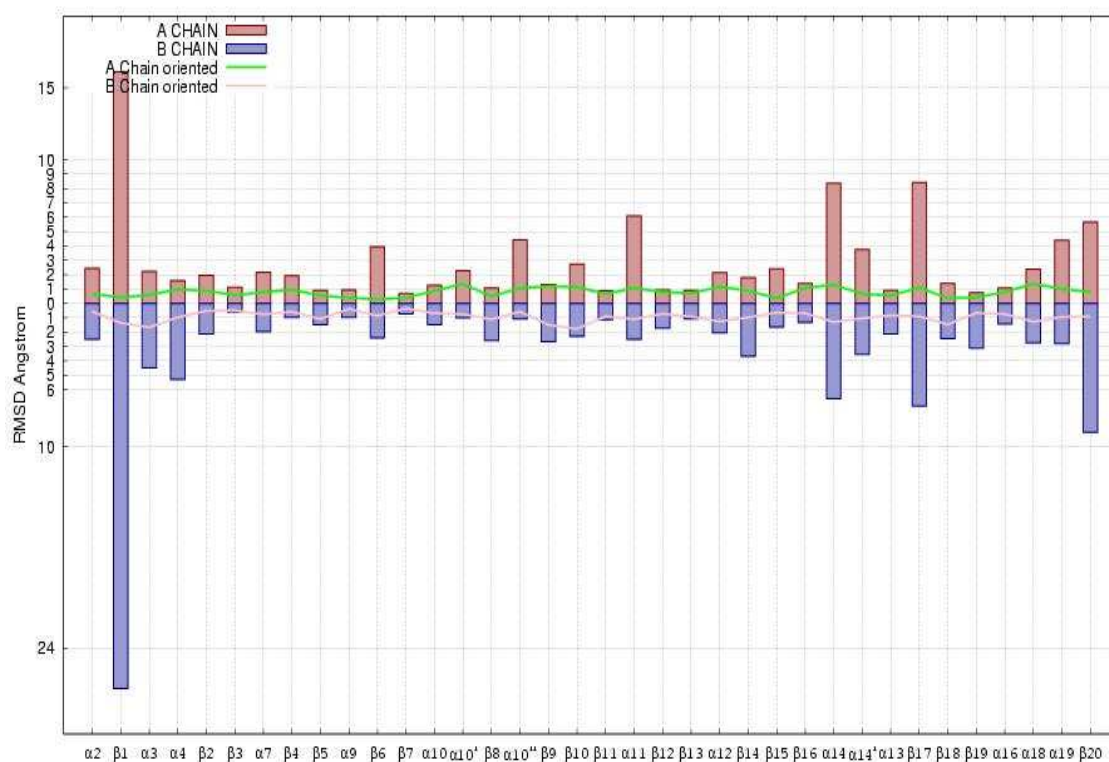**Figure 3.9** Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as closing the lower part of the system starting from the proposed BIOMT structure with accomplished with sixth set-up

In figure 3.9, the rigid rmsd values are seen to be up to 26 A. Secondary structures β1, α14, β17, and β20 are observed to be very mobile, as β1 being the most

mobile subunit in this system. It has about 26 Angstrom rigid rmsd value for B chain, while that of A chain is only about 16 angstrom. The other important values are 9 and 8 angstroms for β17, 5 and 7 angstroms for β20, respectively for A and B chains. One very important point here is that the rigid rmsd values of β1 are not the same for A and B chains. The oriented rmsd values for this system are quite logical as being up to 2.1 Angstrom. So, the secondary structures here are very well conserved. If β1 is neglected, we can say that there is a symmetric behavior in change of rigid rmsd values.
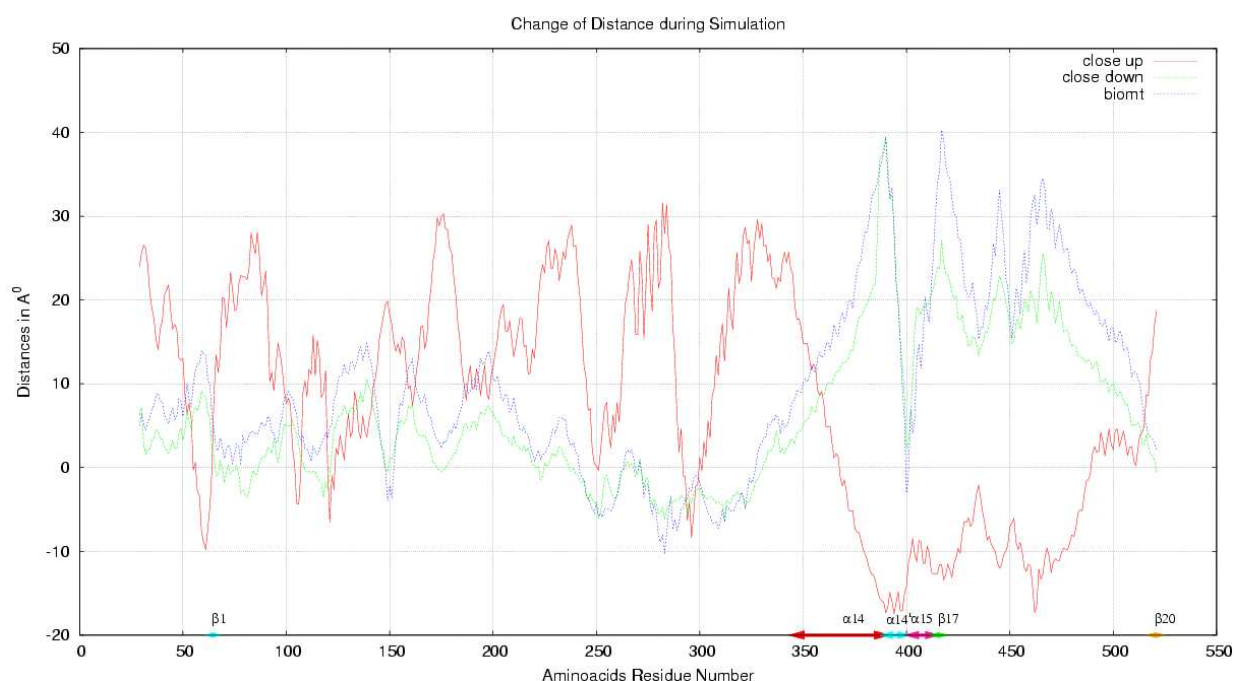
Figure 3.10 Change of distance of amino acid residues for closing upper part, closing lower part and closing lower part for BIOMT structure

Another structural analysis we have is the calculations of the amino acid separations between the same residues on each side. Here, we categorized the systems into two groups; one group is composed of the closure of upper side, closure of the lower gate, and closure to the BIOMT structure. The other group has the same systems, except that the closure motions have some histories from the first group simulations. The figure 3.10 shows the results for the first group analyses. Here, we see that the closing upper gate of the enzyme results in opening of the lower gate (red line in the figure). However, when we close the lower gate, the upper parts of the protein get closer to each other. This difference is nicely in support of our earlier observation that the V-Shaped protein is somewhat stable.
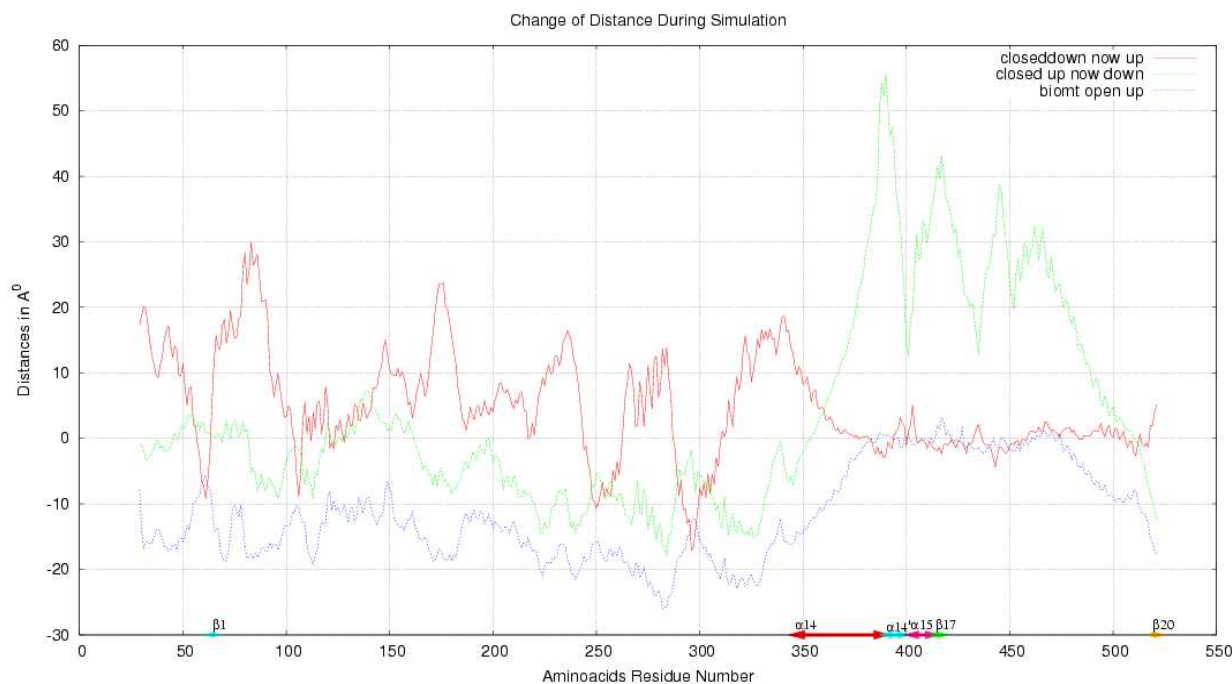
Figure 3.11 Change of distance of amino acid residues for closing upper part starting from a closed lower gate, closing lower part starting from a closed upper part and opening upper part from BIOMT structure

In the second group, we have observed that when we want to close the lower gate of the upper-gate-closed systems, the upper gate wants to get separated again. This observation is nicely in agreement with the step H of the overall mechanism suggested by Burger et.al. [18]. Because when the lower gate is closing after it leaves the T-segment, the upper gate needs to open to let the G-segment DNA to out the protein. However, when we do the same thing for the upper gate, the lower gate is keeping its closed form. This observation supports the process E in the overall mechanism. Another interesting result we have is that when we open the upper gate of the closed BIOMT structure, the lower gate again keeps itself closed. This behavior of the protein is very important because, as proposed by Burger et.al. [18], the D step of the mechanism exactly requires this flexibility of the enzyme, where the opening of the upper side does not affect the closed lower gate.

In regards to energetic analyses, we have done amount of forces applied for each simulations and potential energy surfaces followed. Here, we also did the same distinction, whether or not the simulations have some histories.
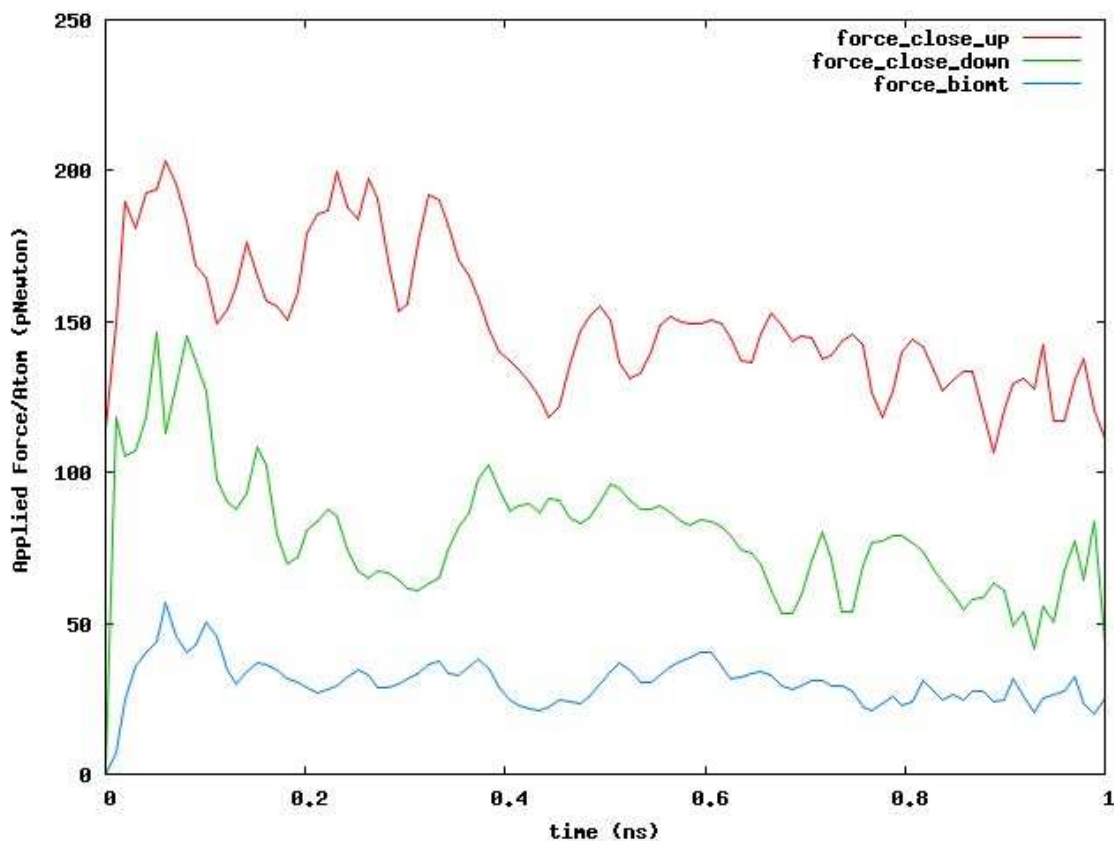
Figure 3.12 Applied forces during the simulation of closing down part, closing upper part and closing system for BIOMT structure

In the first graph we have, figure 3.12, forces needed to close the upper gate is higher than that of closing the lower part, all the time during simulations. This is very well in agreement with the above discussion that the closure of the lower gate is much more feasible. However, closing the lower gate towards the BIOMT structure requires much less forces than the others. This says that the BIOMT structure is quite stable. In first setup, average force per atom is about 150 pico Newton. This setup has 200 pico Newton peek value. In the second setup, average force per atoms for closing lower part of enzyme is less than closing upper part that is around 90 pico Newton. Applied average force per atom to push the for BIOMT structure is about 40 pico Newton and this is the least value comparing to the other setups. Therefore, enzyme prefers to be in BIOMT structure than other closing systems.

Figure 3.13 Applied forces during the simulation of closing upper part starting from a closed lower gate, closing lower part starting from a closed upper part and opening upper part from BIOMT structure

In figure 3.13, we have seen same pattern as in figure 3.11 as closing upper part requires more force than closing lower gate of enzyme. Starting from a lower gates closed structure of enzyme, average required force per atom in order to close upper part is around 175 pico Newton which is the highest value with respect to the other closing and opening mechanisms. Average force per atom for closing lower part started from an upper closed part structure is about 100 pico Newton. Lastly, applied average force per atom beginning with a BIOMT structure and opening upper part to allow the T-segment pass through the gate is around 30 pico Newton and this value is the least if we make a comparison with other closing systems that described above.

After analyzing the applied force, we now begin to analyze interaction energies between A chain and B chain. Interaction energy is sum of electrostatic energy and Van Der Walls energy.
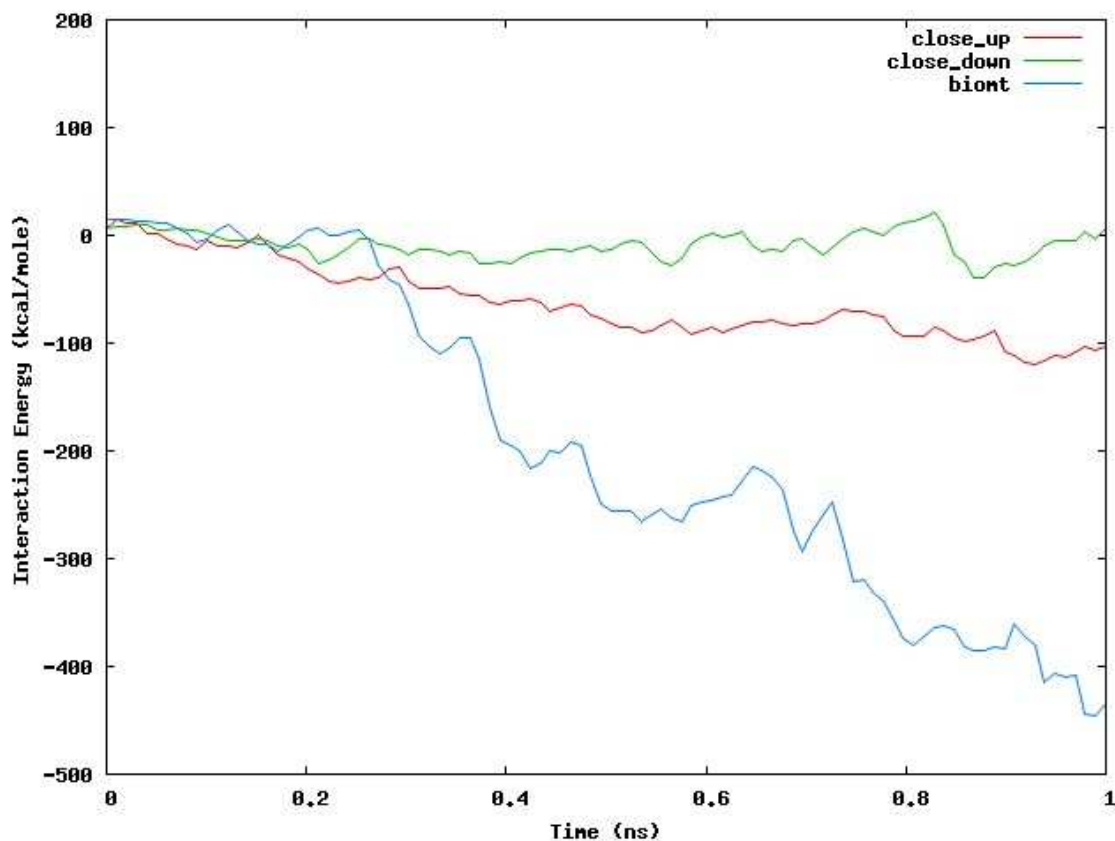
Figure 3.14 Interaction energies between A chain and B chain during the simulation of closing lower part, closing upper part and closing system for BIOMT structure

In figure 3.14, we can say that there was no any important interaction between chains of protein while the upper part of enzyme is closed. However, during the closure of the lower gates, chains interacts each other with about 70-80 kcal/mole energy. The most interesting section is the closure of lower gates for pushing system to go the BIOMT structure. We see that interaction energy is decreasing steadily up to -400 kcal/mole. This shows us, the more chains get closer, the more they attract each other. Another significant result that we come is to BIOMT structure is the most stable structure.

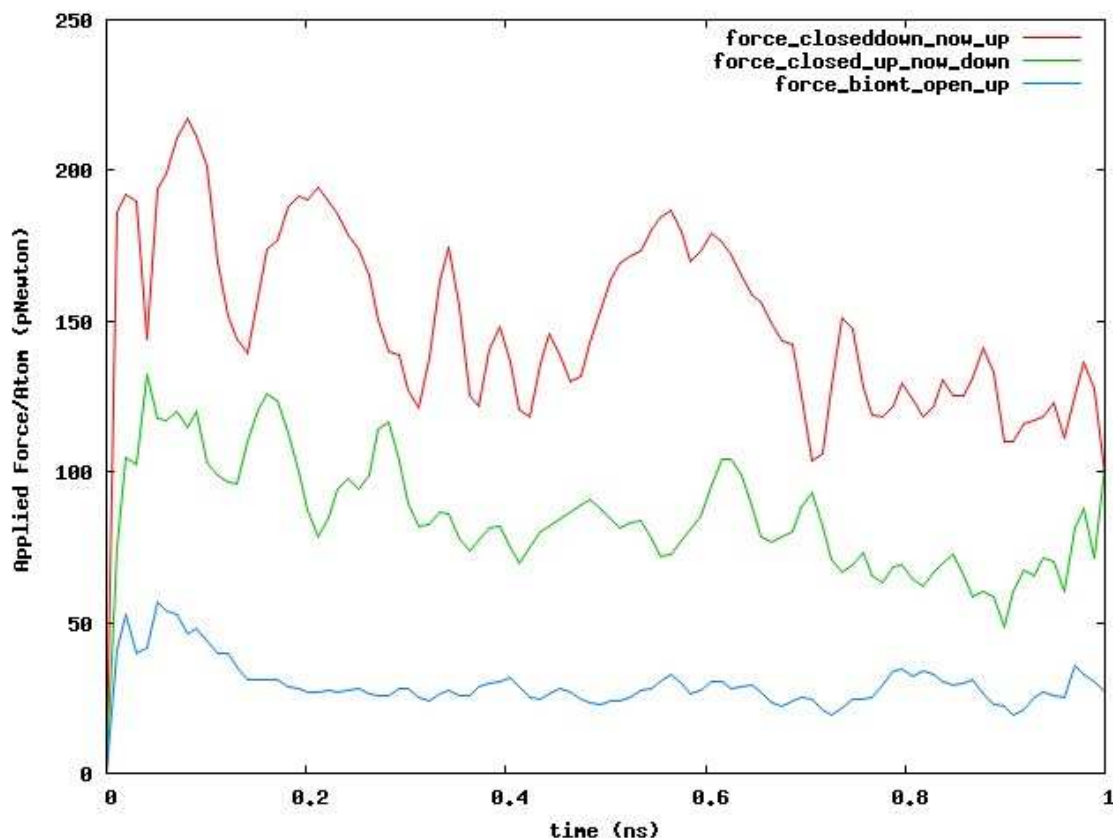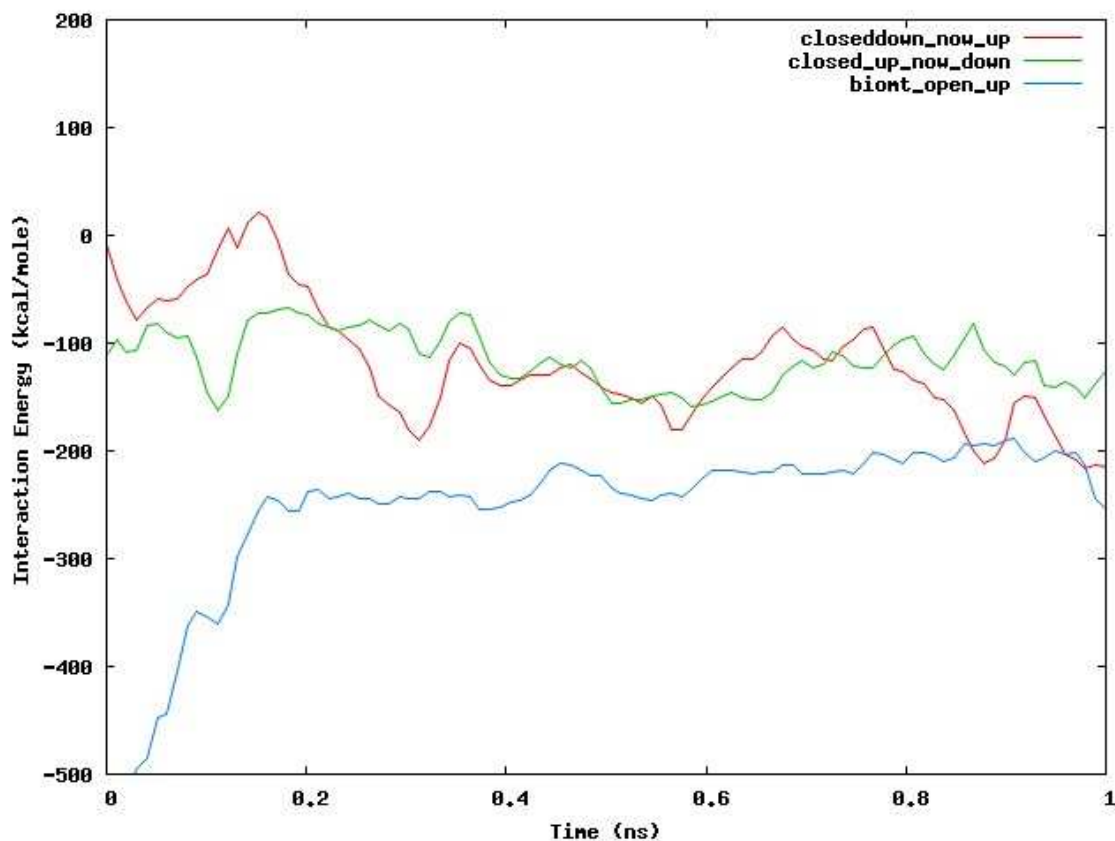**Figure 3.15** Interaction energies between A chain and B chain during the simulation of closing upper part starting from a closed lower gate, closing lower part starting from a closed upper part and opening upper part from BIOMT structure

In figure 3.15, there was not any distinction of whether the closing the lower gate followed by closing upper gate or closing upper gate followed by closing lower gate in terms of interaction energies. Average interaction energy for the both the closing upper or lower part with closed other part is about -100 kcal/mole. In addition to that, interaction energy of opening upper section starting from BIOMT structure is increasing rapidly from 0 ns to 0.2 ns time interval. After that, it becomes more or less stable with the value of – 200 kcal/mole.

Figure 3.16 Total potential energies of enzyme during the simulation of closing upper part starting from a closed lower gate, closing lower part starting from a closed upper part and opening upper part from BIOMT structure

In figure 3.16, total potential energy for the third system in the graph which is closing lower gates to push the enzyme in BIOMT structure has the lowest value. In addition to that, both closing lower and upper gates approximately have the same potential energy surface. The scale of change in total potential energy is relatively narrow and it is about 400 kcal/mole. By analyzing potential energy surfaces, it is evident that, DNA GyraseA enzyme prefers to be neither lower gate closed form nor upper gate closed form but BIOMT structure.

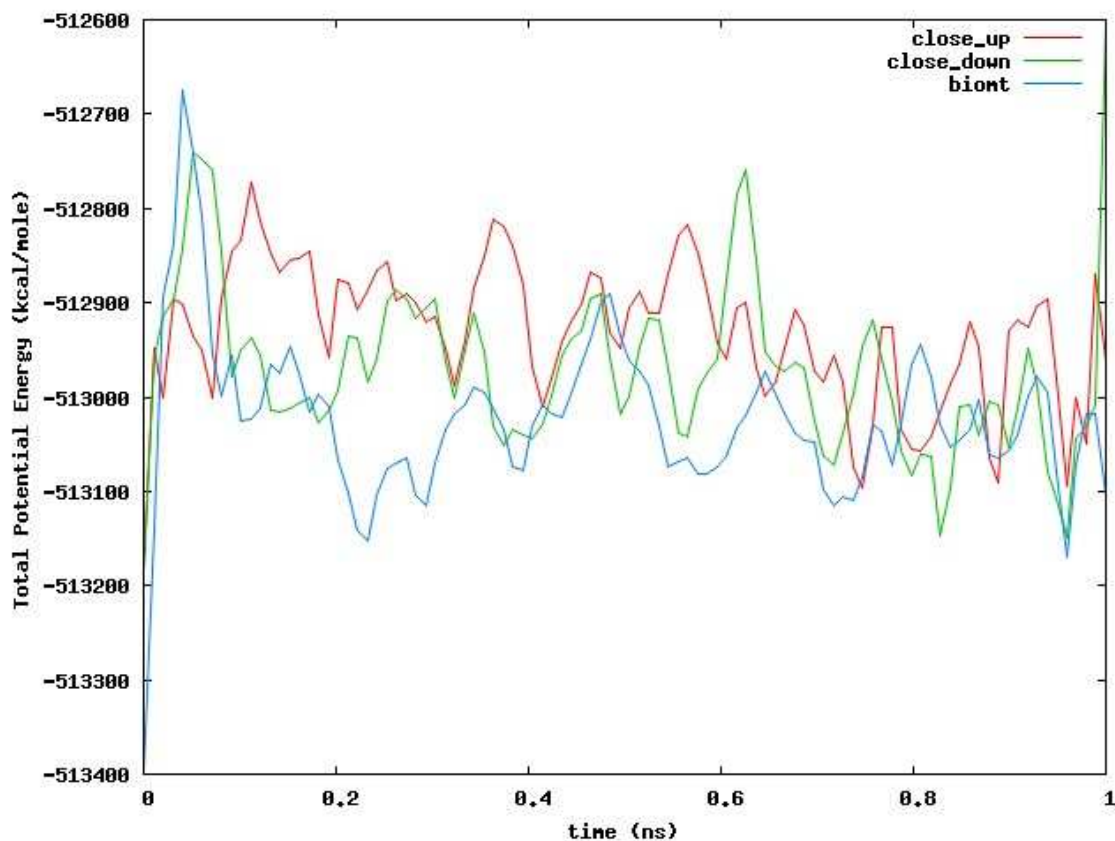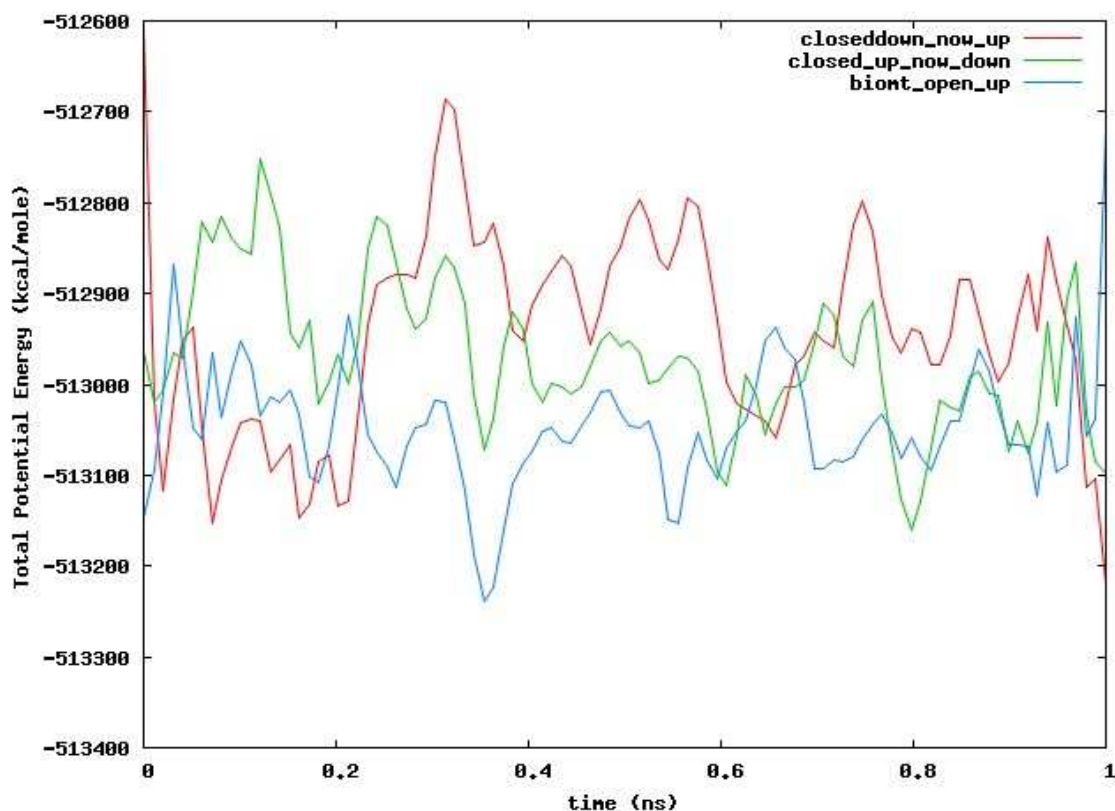**Figure 3.17** Total potential energies of enzyme during the simulation of closing upper part starting from a closed lower gate, closing lower part starting from a closed upper part and opening upper part from BIOMT structure

In figure 3.17, we can see the total potential energy for following systems. First, we take the previously lower gate closed enzyme and try to close upper gates. Secondly, we start from the upper gate have already closed system and try to close lower gates. In the third system, we start from BIOMT structure and try to open upper gates pave the way for allowing transition of T-segment of DNA. As seen by the graph, potential energy for the third system is the lowest with respect to other set-ups. Therefore, enzyme can easily make the opening motion for the BIOMT system. It is hard to say about other two system but we can infer that closing the upper gate of previously lower gate closed protein follows highest potential energy surface. This is consistent with our previous observations supporting the idea that closing upper gate is not relatively easily.

Table 3.1 Rotations of A Chain, B Chain, and overall protein for the different types of closing and openings that described in figure 3.1 with α_{hqbmd} 10 kcal/mol/A^4. To remove the total rigid rotation of the system, rotations are obtained by superimposing final geometry onto initial geometry

	A CHAIN	*B CHAIN*	*PROTEIN*
SIMULATION 1	30.80	20.75	7.65
SIMULATION 2	38.17	30.37	37.22
SIMULATION 3	45.47	29.44	37.76
SIMULATION 4	36.47	47.89	57.02
SIMULATION 5	28.46	25.13	33.82
SIMULATION 6	76.99	97.58	80.23
SIMULATION 7	13.59	22.78	13.05
SIMULATION 8	33.79	29.99	41.29

We also make an analysis with the rotation A chain, B chain and total protein. It is interesting that, in simulation 1 system, rotation for A chain is 30.80 degree and rotation of B chain is 20.75 but total rotation of protein 7.65. Simulation 6 has the biggest value of rotation with almost 80 degree for A chain, 97.58 degree for B chain and 80.23 for the total amount of rotation that the enzyme has reached in order to go the BIOMT structure. Hence, we can say that before the enzyme start to process on DNA, it makes some appreciable amount of rotation in order to capture the segment of DNA, which is another evidence that supports the suggested mechanism in figure 1.8.[18]

CHAPTER 4

CONCLUSIONS

In this study, basically we have investigated structural and energetic behaviors of a type IIA topoisomerase; DNAGyraseA. As we do not have a dimmer structure of the protein, first we tried to find most favorable dimer structure, by closing the two monomers by employing a biased molecular dynamic simulation (HQBMD) on eight different set-ups. Based on the results and discussion given in the previous section, we can deduce the following main conclusions out this study.

a-) The most important result we have is that the BIOMT structure is stable relative to any other new dimmer form we got. We have observed that the interaction energy between the two chains is becoming attractive when we close the two parts towards the BIOMT structure. Also, potential energy surface is observed to be quite low with respect of other systems.

b-) The closure of the lower gate of the enzyme is observed to be more feasible than the closure of the upper gate. Here, we have found that the upper gate closures are always follow relatively high potential energy paths, and also amount of force needed to bring the two monomers are notably higher than those for systems.

c-) The *simultaneous closure*s of the upper and lower gates are found to be not favorable energetically. The reason for this is that this motion follows the highest potential energy surface, and also amount of force applied is relatively higher than the others.

d-) The most mobile secondary structures are found to be $\beta 1$, $\beta 17$, $\beta 20$, $\alpha 14$, $\alpha 14^{1}$, and $\alpha 15$. These subunits show considerable amount of displacements during the

closures. In almost all systems, we have observed that these rigid movements are very symmetric as we expect. Interestingly, this symmetry is conserved for both upper and lower gate movements.
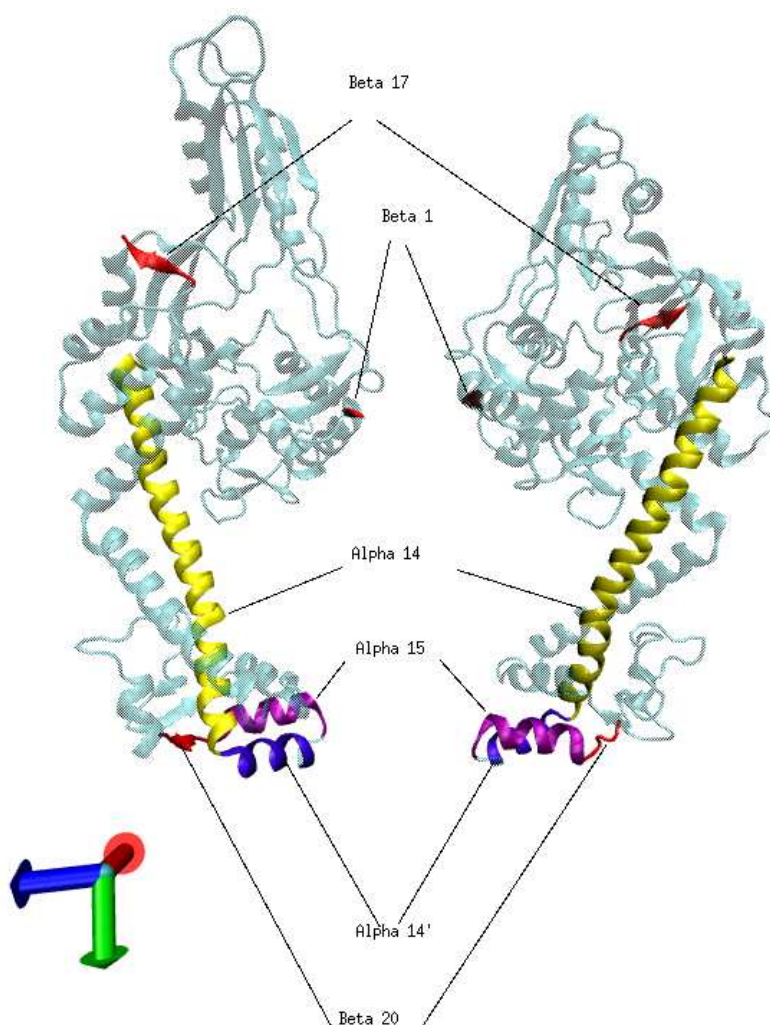
Figure 4.1 Most mobile secondary structures during simulations

e-) The dimmer structure where the top gate is open, and the lower gate is closed is observed to be favorable. This is nicely in agreement with the current proposed mechanism where the protein opens up from the top gate to capture G-segment DNA. We have found that the completely closed protein (both from top and bottom parts) is not favorable than the V-shaped structure (where only the bottom part is closed).

f-) We have found out that the dimmer form obtained by a 180 degree rotation around the inter-monomer axis is much more favorable than the dimmer structure obtained by a reflection through the plane passing from the inter-monomer interface.

g-) Our simulations very nicely support the overall relaxation mechanism proposed by Berger et. al.[18] Because, we have not carried out all proposed steps, our simulations supports only steps D, E, and H in figure 1.8. These are as follows: if the enzyme opens the upper gate from the BIOMT structure, to allow the passage of T-segment over cut G- segment, lower gates remains closed during this process which is illustrated as D in figure 1.8. In E step, when we try to close upper gate, lower gate opens itself like the teeterboard mechanism. This mechanism pave the way for letting T-segment released from lower gate. In H step, we see the reverse of the previous action, after releasing T-segment if we want to close lower gates, again upper gates spontaneously open to release the held G- segment.

h-) We have found that the opening of the upper gate of the enzyme requires very small amount of forces, compared to the forces needed to bring closures. This tells us that the upper gate is very loose, and easily opens up to catch the DNA.

REFERENCES

[1]   Watson J.D. and Crick F.H.C., *Nature* , Vol.171, pp.737–738, 1953

[2]   Mandelkern M et.al. *J Mol Biol*, Vol.152-1, pp.153–61, 1981

[3]   Dulbecco, R. and Vogt,M., *Proc. Natl. Acad. Sci. USA*, Vol. 50, pp.236–243, 1963

[4]   Weil, R. and Vinograd J., *Proc. Natl. Acad. Sci*. USA Vol. 50, pp. 730–738, 1963

[5]   F. B.Fuller, *Proc. Natl. Acad. Sci. U. S. A.* 1971, 68, 815

[6]   L. M. Fisher et al., *Phil. Trans. R. Soc. Lond. B* 1992, 336, 83

[7]   Depew,R.E.and Wang, J. C. (1975). *Proc. Natl. Acad. Sci USA*, Vol.72, pp. 4275–4279, 1975

[8]   Steven R. et.al. *Cell Biology: A Short Course,* John Wiley and Sons, 2004

[9]   Jeon, J. and Sung, W., *Biophysical Journal*, Vol: 95-8, pp. 3600-3605, 2008

[10]   D.L.Nelson, M. M. Cox, *Principles of Biochemistry,* Freemann, New York, 2000.

[11]   Wang, J.C, *Annu. Rev. Biochem*, Vol. 65, pp. 635-682, 1996

[12]   L. Srewart, M. R. Redinbo, X. Qiu, W. G. J. Hol, J. J. Champoux, *Science* 1998, 279, 1534

[13]   A.D. Bates, A. Maxwell, *DNA Topology*, Oxford University Press

[14]   J. C. Wang, *Nature Reviews,* Vol. 3, pp. 430, 2002

[15]   J.J. Champoux, *Annu. Rev. Biochem*, Vol. 70, pp. 369–413, 2001

[16]   Kevin D. Corbett and James M. Berger, *Annu. Rev. Biophys. Biomol. Struct*., Vol. 118, pp. 33-95,2004

[17]   Gadelle, D., et.al., *BioEssays,* Vol.25, pp. 232–242, 2003

[18]   Berger J.M. et.al., *Nature,* Vol. 379, pp. 225-231, 1996

[19]   Schoeffler A.J. and Berger J.M., *Quarterly Reviews of Biophysics*, Vol. 41/1, pp. 41–101, 2008

[20]   Roca J and Wang JC., *Cell* , Vol. 77, pp. 609–16, 1994

[21]   Laponogov I, et.al, *Nat Struct Mol Biol*. 2009, Epub ahead of print

[22]   Bates B.A., *Current Biology* ,Vol:16, pp. 6, 2006

[23]   Yamaguchi, H. et.al., *Chem. Letters,* Vol. 29, pp. 384-385

[24]   Schoeffler A.J. and Berger J.M., *Biochem Soc Trans*., Vol.33-6, pp.1465-70, 2005

[25]   Heddle J.G. et.al., *N.N. Nucleic Acids,* Vol. 19, pp.1249-64, 2000

[26]   Karplus, M. and McCammon J.A., *Nature Structural Biology* , Vol. 9, pp. 646 – 652 2002

[27]   Alder, B. J. and Wainwright, T. E. *J. Chem. Phys*., Vol.27, pp. 1208, 1957

[28]   Stillinger, F. H. and Rahman, A. *J. Chem. Phys.* , Vol. 60, pp. 1545, 1974

[29]   McCammon, et.al., *Nature (Lond.),* Vol 267, pp. 585, 1977

[30]   Morra, G., *Role of electrostatics explored with molecular dynamics simulations for protein stability and folding*, Ph.D. Thesis, Universitat Berlin, 2005.

[31]   Tuckerman, M. E., Martyna, G.J., *J. Phys. Chem. B* Vol. 104, pp.159-178, 2000

[32]   MacKerell et el. *J. Phys. Chem. B.*, Vol. 102, pp.3586, 1998

[33]   McQuarrie, D.A., *Statistical Mechanics*, University Science Books, 2000.

[34]   L. Verlet, Phys. Rev., Vol.159, pp. 98, 1967

[35]   Hinchliffe, A., *Molecular modelling for beginners*, John Wiley, New York, 2003.

[36]   Martin Karplus, *The CHARMM Development Project*, http://www.charm.org

[37]   E. Paci and M. Karplus*, J. Mol. Biol*., Vol. 288, pp. 441-459, 1999

[38]   M. Marchi, P. Ballone *J. Chem. Phys*., Vol. 110/8, pp 3697, 1999

[39]   Karplus, M et.al*., J. Mol. Biol.,* Vol.306, pp. 329–347, 2001

[40]   Andersen H.C., *J. Chem. Phys*., Vol. 72, pp. 2384–2393, 1980

[41]   Berendsen H.J.C, et.al., *J. Chem. Phys,.* Vol. 81, pp. 3684–3690, 1984

[42]   Hu, Y. and Sinnott, S.B., *Journal of Computational Physics*, Vol. 200/1, pp.251-266, 2004

[43]   M. Karplus et.al., *J. Comp. Chem.,* pp. 4- 187, 1983

[44]   Stote, R. et.al., *"Molecular Dynamic Simulations and CHARMM tutorial"*, http://vit-embnet.unil.ch/MD_tutorial/

[45]   J.H.M. Cabral et.al., *Nature,* Vol 388, pp. 903- 906, 1997